SCHOOL-LEVEL MATTHEW EFFECTS IN READING:

AN ANALYSIS OF WITHIN-YEAR BENCHMARK DATA

By

MELISSA A. HARMAN

A DISSERTATION

Presented to the Department of Special Education and Clinical Sciences
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2022

DISSERTATION APPROVAL PAGE

Student: Melissa A. Harman

Title: School-Level Matthew Effects in Reading: An Analysis of Within-Year Benchmark Data

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Dr. Hank Fien   Co-Chairperson
Dr. Gina Biancarosa   Co-Chairperson
Dr. Joseph Nese   Core Member
Dr. Cengiz Zopluoglu   Institutional Representative

and

Dr. Andrew Karduna   Interim Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2022

DISSERTATION ABSTRACT

Melissa A. Harman

Doctor of Philosophy

Department of Special Education and Clinical Sciences

June 2022

Title: School-Level Matthew Effects in Reading: An Analysis of Within-Year Benchmark Data

The Matthew Effect is an educational theory that children who begin their academic careers with lower reading scores do not catch up to their peers, instead falling farther and farther behind as their schooling progresses. This study includes a hierarchical multiple-regression analysis of within-year benchmark data for the presence of Matthew Effects in literacy. School mean and median school reading scores for three DIBELS measures: oral reading fluency, word reading fluency, and nonsense word fluency were analyzed. The sample included schools in the DIBELS data system which completed universal screening during the 2018-2019 school year for their second-grade students. The analysis of school scores compared schools with above average proportions of historically under-resourced student populations to better understand systematic inequalities in education, specifically including groups identified by the Department of Education as experiencing achievement gaps. Findings were not statistically significant, meaning schools that served above average proportions of racially underrepresented students or students living in poverty did not have significant differences in beginning of year DIBELS scores nor beginning to end of year change in scores. Limitations and future directions are addressed.

CURRICULUM VITAE

NAME OF AUTHOR: Melissa A. Harman

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon (APA accredited, NASP approved), Eugene, Oregon
September 2017 – June 2022 (Expected)
PhD., School Psychology

The State University of New York at Buffalo State, Buffalo, New York
September 2013 – April 2016
M.S., Interdisciplinary Studies (International Education)

University of Michigan, Ann Arbor, Michigan
September 2008- April 2011
B.A., Social Science and B.A., Political Science

DEGREES AWARDED:

Doctor of Philosophy, School Psychology, Expected 2022, University of Oregon
Master of Science, Interdisciplinary Studies, 2016, SUNY Buffalo State
Bachelor of Arts, Political Science and Social Theory & Practice, 2011, University
of Michigan

AREAS OF SPECIAL INTEREST:

Child & Family Therapy
Universal Screening and Progress Monitoring
Multi-Tiered Systems of Support
Identification and Intervention for Students with Learning Disabilities
Supporting Culturally and Linguistically Diverse Students
Dismantling the Patriarchy and All Systems of Oppression
The Intersection Between Children's Behavior, Academics, and Mental Health

PROFESSIONAL EXPERIENCE:

Pre-Doctoral Intern and Clinical Fellow , Judge Baker Children's Center and
Harvard Medical School, July 2021- July 2022

Extern, Comprehensive Diagnostic Assessment Clinic (CDAC), University of
Oregon, July 2020- June 2021

Practicum Therapist, Child and Family Center (CFC), University of Oregon, September 2019- June 2021

Graduate Teaching Fellow, College of Education, University of Oregon September 2018- June 2021

School-Based Practicum Student, 4J Eugene School District, September 2018- June 2019

GRANTS, AWARDS, AND HONORS:

Multi- Tiered Systems of Support Early Career Scholar, Institute of Educational Science's MTSS Research Network Leadership Team, 2020

The Wes Becker Scholarship, University of Oregon, 2020

The Education Studies Department Award for Graduate Employee Excellence, University of Oregon, 2019 and 2020

The Dynamic Measurement Group Award, University of Oregon, 2017, 2018, 2019, 2020

First-Year Fellowship, Graduate School, University of Oregon, 2017- 2021

Teacher of the Year, Colegio Americano de Torreon, 2016

ACKNOWLEDGMENTS

I am incredibly grateful for the guidance and mentorship of my co-chairs Dr. Hank Fien and Dr. Gina Biancarosa. They have nurtured my love of literacy and guided my understanding of the science of reading. The completion of this dissertation would not have been possible without the feedback and participation of my committee members Dr. Joe Nese and Dr. Cengiz Zopluoglu. I wish to thank my cohort mates Nichole Freiboth, Jillian Hamilton, Antonella Onofrietti, and Kaitlyn Roy for their continual encouragement and friendship. A special thank you to Katie Tennant Beenen for introducing me to the field of school psychology. I owe a deep sense of gratitude to my parents, Teresa and Richard Harman, and sisters, Katie and Amy Harman, for their constant and unwavering support. Lastly, I wish to acknowledge all the strong women in my family who have dedicated their lives to the service of others and found careers in education, healthcare, and social work.

I dedicate this study to my former students and future clients.

TABLE OF CONTENTS

LIST OF FIGURES

Figure   Page

Figure   Page

LIST OF TABLES

CHAPTER I
INTRODUCTION

Reading has long been identified as an essential academic skill for students. Although gains have been made in reading research, especially surrounding identification of reading disabilities and intervention, research over the past twenty years has indicated the persistence of populations of children within our public-school system that continue to struggle with low reading abilities. Various studies have indicated that 1 in 5 students in the school system struggle with significant problems learning to read during the elementary school years (Lyon, 1995; Moats, 1999). These reading difficulties are correlated with school-based academic problems in later grades and general adverse life outcomes such as high school graduation rates and future participation in the economic system (Fiester, 2013; Kame'enui et al., 2000).

Although guidelines from the National Reading Panel (2000) indicated a clear way forward with teaching reading and intervening with struggling students, recent nationwide data indicates that by the 4th grade, over 60% of students persist at reading below grade level proficiency (National Center for Educational Statistics, 2019). The 2019 National Assessment of Educational Progress (NAEP) showed both 4th and 8th-grade students had lower reading scores on average nationwide than in 2017, meaning that a smaller percentage of students scored at proficiency. Only one state, Mississippi, showed an overall increase in their statewide reading data compared to 2017 (National Center for Educational Statistics, 2019).

Special attention must be placed on students who struggle with developing early reading skills as these students often continue to struggle throughout their elementary

school years (Butler et al., 1985; Snow et al., 1998). A 2010 report by the Annie E. Casey Foundation showed that only 25% of students who do not read on grade level by the end of second grade can catch up to their peers by the end of fifth. Those who struggle to read by the end of second grade have a 75% chance of not reading on grade level by the end of their elementary school experience (Fiester, 2010). This pattern has persisted as average NAEP reading scores have increased between 1992 and 2019 for children at the 90th percentile, but those at the 10th percentile showed no such increase in scores (National Center for Educational Statistics, 2019).

Pfost et al. (2014)'s meta-analysis provides an explanation of three potential patterns in school achievement data: Matthew Effects, a Compensatory Model, and a Stable Achievement Gap. Matthew Effects necessitate statistically significant lower fall scores and lower rates of growth for schools based on concentration of poverty or proportion of historically underserved students. A Compensatory Model indicates schools with lower initial scores having greater growth over time than their more privileged comparison schools. This higher rate of growth would allow the closure of the achievement gap, as the difference in mean scores would decrease across the three time points. A Stable Achievement gap would show significant differences in fall scores but similar rates of growth for all groups, resulting in stability over time.

Research into the Matthew Effect has shown that our public education system is often unable to ameliorate achievement gaps between lower-scoring readers and their not at-risk peers. The Matthew Effect is based on the idea that 'the rich get richer and the poor get poorer.' Those who enter the educational system with lower reading levels

tend to fall further and further behind peers over time (Francis et al., 1996; Stanovich, 2000; Walberg & Tsai, 1983). Previous studies have expanded our understanding of the Matthew Effect to include analyses of demographic variables, showing that students from historically disadvantaged racial or socioeconomic groups are often at greater risk for Matthew Effects (Logan & Petscher, 2010; Morgan et al., 2008; Rumberger & Palardy 2005). However, aggregate analysis of school scores is a neglected area of Matthew Effect research. Prior research has analyzed the Matthew Effect through student-level analysis, centering on a systems-level social issue instead of within-child. However, some international research from the Organization for Economic Co-operation and Development (2010) and Thomson & De Bortoli (2010) has shown the importance of school-level factors in influencing student learning trajectories. Through looking at the Matthew Effect in the aggregate, on a school comparison level, we can instead analyze the Matthew Effect as a social issue reflecting widespread and historical systemic disadvantages.

Study Goals

This study analyzes school-level, within-year growth variability of curriculum-based reading measures for evidence of the Matthew Effect. An analysis of school scores compares schools that serve proportionally above average populations of historically under-resourced students as identified by the U.S. Department of Education as experiencing achievement gaps in reading (e.g., Black, Hispanic, Native American / Alaskan Native students, and those living in poverty).

Analyzing the relationship between the school-level proportion of students identified by the U.S. Department of Education as experiencing achievement gaps in reading and schools' mean and median reading scores allows for a systems-level examination of achievement inequities. That is, districts and states often examine mean performance of students as an indicator of the extent to which schools are facilitating learning. Use of within-year benchmark curriculum-based reading scores mirrors how schools and districts often track data regarding student progress in reading. Given the widespread use of DIBELS, this study may provide a useful window into potential disparities in how the public education system serves our most vulnerable children based on the concentration of those children in schools. Evidence of statistically significant differential mean and median reading scores initially and increasingly divergent scores over time based on the proportion of under-resourced students served would indicate the presence of a Matthew Effect. In essence, if schools serving fewer under-resourced students demonstrate better gains in reading, it would indicate a system where schools serving above average proportions of at-risk students cannot catch up to those serving below average proportions of those students.

Statement of Purpose

The following study will analyze within-year trajectories in second grade literacy to examine school-level Matthew Effects. Previous research into the Matthew Effect has centered on individual students as the level of analysis, asking if individual students who struggle with reading are able to catch up to their peers (Bast & Reitsma, 1998; Cain & Oakhill, 2011; Francis et al., 1996; Juel, 1988; Stanovich, 2000; Walberg & Tsai, 1983).

Yet, there is increased interest in analyzing literacy, achievement gaps, and the Matthew Effect on a school-level of comparison, instead focusing on average literacy scores for schools (Hanushek & Rivkin 2006; Holmes-Smith, 2006; Palardy 2008; Stiefel et al., 2007; Thomson & De Bortoli, 2010).

Analysis of school average scores allow us to analyze the impact of the Matthew Effect with a systems-wide lens. Rather than centering research on the achievement gap as a within-student concern, this study focuses on school scores as reflection of historical inequities in education. Within the U.S. public school system, certain racial and socioeconomic groups have historically been excluded from the educational benefits provided to others. Focusing on schools that provide services to these groups (Black, Hispanic, and socioeconomically deprived students) provides insight into a system that should strive towards effectively ameliorating disadvantages that students face. Second grade is the focus of the study due to prior research showing the importance of early literacy skills on students' long-term trajectories (Lyon, 1995; Moats, 1999). Furthermore, cross-sectional research from Silberglitt and Hintze (2007) has shown greater differences in within-year growth rates of second grade students in comparison to their older peers.

CHAPTER II

LITERATURE REVIEW

The Matthew Effect

Due to our knowledge of the importance of reading skills in the early grades, it becomes essential to understand how student skill levels progress through their early academic careers. Longitudinal data has shown a troubling trend where readers who struggle in early elementary school continued to struggle with reading throughout their elementary school years. Students who began their early academic careers behind their peers in skill level never seemed to catch up and instead fall farther and farther behind (Francis et al., 1996; Stanovich, 2000). This pattern is called the Matthew Effect for the adage 'The rich get richer and the poor get poorer' and is thought to be rooted in access to education and instruction early reading skills provide to students (Stanovich, 1986). This effect is also called the fan spread pattern in reading where students, although they start at different but somewhat similar skill levels, progress in skill differently leading to widening gaps between the highest and lowest skilled students much in the way the parts of a fan spread from their source (Stanovich, 1986). The most at-risk groups for reading failure are those students who enter Kindergarten with lower literacy skills than their classmates. Fiester (2010)'s analysis of longitudinal data found that 88% of students who failed to complete high school were struggling readers in the third grade.

Matthew Effects Across Years

Various studies have shown that students who begin their educational careers behind their peers in literacy achievement never seem to catch up and rather fall farther

and farther behind (Francis et al., 1996; Stanovich, 2000; Walberg & Tsai, 1983). One of the earliest studies on the Matthew Effect analyzed reading data. Juel (1988) analyzed the longitudinal data from the first to the fourth grade. The 54 students were all from the same Austin, Texas elementary school which enrolled a high proportion of minority and socioeconomically disadvantaged students. Juel (1988) found "the probability that a child would remain a poor reader at the end of fourth grade if the child was a poor reader at the end of first grade was .88" (pg. 437). He attributed the lack of improvement in reading scores to students' low decoding abilities at the beginning of the academic careers, which prevented them from accessing text at both school and home.

Some studies have tied the Matthew Effect across years to demographic risk variables. In one such study, Morgan et al. (2008), analyzed data from the Early Childhood Longitudinal Study (ECLS) from the Institute of Educational Sciences on the basis of risk variables rather than beginning reading level. They analyzed data from 10,587 students who entered Kindergarten during the 1998 school year, including literacy scores from Fall of Kindergarten to Spring of 3rd grade. Rather than creating comparison groups by original literacy score, Morgan et al. (2008), analyzed demographic risk, including students' gender, racial background, and family socioeconomic status based on students who have historically been identified as having learning disabilities at higher rates than more privileged peers. Morgan et al. (2008) calculated z-scores for their demographic groups, quantifying the Matthew Effect as a significant change in z-scores from Kindergarten to 3rd grade. Students they had

identified as being demographically at risk did have an average decrease in z-scores, meaning those students did become poorer readers with an increasingly large gap between their literacy scores and the population mean.

Multi-level analysis showed that male students, Hispanic and Black students, and those from lower socioeconomic groups did have lower beginning Kindergarten reading scores than their more advantaged peers. Additionally, growth rates were higher in female students, White and Asian students, and those from higher socioeconomic backgrounds. Interestingly, the authors additionally found evidence for compounding intersectionality within the sample, meaning students who were male, Black, and from a lower socioeconomic group had the lowest growth rate, leading to a Matthew Effect in their literacy scores. Morgan (2008) found the discrepancy between the reading scores of White students and Black students grew from Kindergarten to third grades; "black males in the lowest SES quintile lagged further behind their peers in reading growth. Between the beginning of Kindergarten and the end of third grade their average reading Z-score declined from -0.66 to -1.12" (pg. 9).

However, Morgan et al. (2008) did not find that more privileged students had accumulating benefits shown in their reading scores, as their z-scores remained mostly constant between Kindergarten and the 3rd grade. These results led the authors to conclude that students with demographic risk did become 'poorer' readers over the years and did not catch up to their peers although the more privileged peers did not necessarily show significant cumulative advantages. Although Morgan et al. (2008) viewed their study's results as being only partially conclusive of a Matthew Effect, their

students with higher demographic risk did show a widening achievement gap- the hallmark of the Matthew Effect.

In contrast, Northrop (2017) likewise analyzed data from the same 1998-1999 Kindergarten Early Childhood Longitudinal Study (ECLS) to look at cumulative disadvantages in reading achievement using both initial reading achievement and demographic variables such as race and socioeconomic status. Building on Morgan et al. (2008), Northrop (2017) expanded the comparison timeline, looking at the progression of reading scores between Kindergarten and 8th grade. Whereas Morgan et al. (2008) had primarily analyzed demographic risk, Northrop (2017) analyzed the same sample instead looking for demographic differences of Kindergarten struggling readers who were or were not able to catch up to their peers by the end of 8th grade. Within the sample, there was evidence for a Matthew Effect; students who began Kindergarten with lower reading achievement skills had slower reading growth than peers with stronger reading achievement. Most importantly, Northrop (2017) found significant race and socioeconomic differences for low-reading ability Kindergarten students who were able to catch up to their peers by the end of middle school as opposed to those who remained in the lowest reading ability groups. Northup further identified that 54% of children who struggled in reading during Kindergarten were able to catch up to their peers by the end of middle school; however, students who experienced these compensatory effects "are more likely to be white, come from a higher SES household, attend schools with students from higher SES households, spend more time reading at home…" (pg. 396). Northup identified not only children's individual demographic

information but also school-level characteristics. Northup concluded that students experiencing compensatory effects were more likely to attend schools that served high SES students. Matthew Effects were present for those with lower socioeconomic statuses, students of color, and those who had less general access to literacy-rich environments (Northrop, 2017). Morgan et al. (2008) and Northrop (2017) both centered analysis of the Matthew Effect on long-term outcomes for individual students, showing that there are widening achievement gaps between students from more and less privileged demographic groups. However, individual schools and districts often use within-year data to track student progress and make curriculum decisions.

Matthew Effects Within-Year

Across-year longitudinal studies have showed some evidence of Matthew Effects for at-risk and not-at risk readers (Francis et al., 1996; Stanovich, 2000; Walberg & Tsai, 1983); within-year longitudinal studies provide additional evidence. Schools often use within-year analyses of curriculum-based measurement growth to track student progress and make decisions concerning students' education (Fuchs & Fuchs, 2011). Within-year CBM reading scores allow a closer look at the progression of Matthew Effects within a single school year. Schools often make curriculum, instruction, and intervention decisions based on school-wide or individual benchmark or progress monitoring data, tracking within-year CBM scores (Fuchs & Fuchs, 2011). While across-year CBM data allows us to see systems-wide progressions, it is also of limited utility for educators in the field. Data from a single school year has more practical relevance for instruction and intervention, allowing for a more immediate educational response.

Considering that immediate intervention is of the utmost importance for student skill progression in the early elementary years, it is essential to look at a more targeted analysis of within-year Matthew Effects.

Figure 1
Within-Year CBM Growth Rates
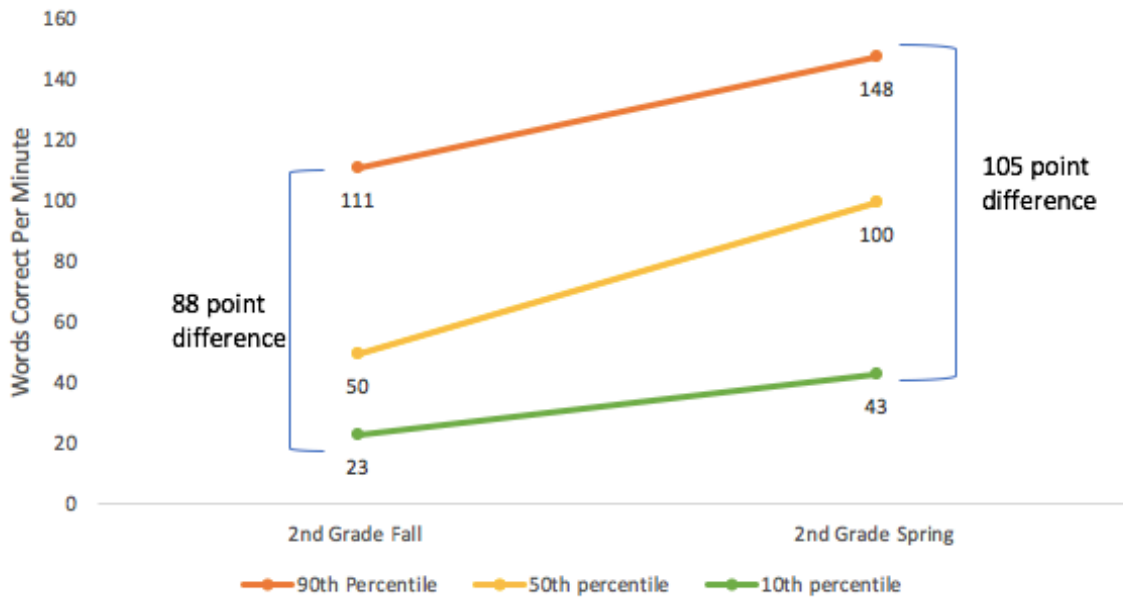
Hasbrouck and Tindal's (2017) seminal report on oral reading fluency norms overviewed 30 years of research that students' oral reading fluency growth rates differed by their beginning of the year reading achievement score. As shown in Figure 1 , second-grade students beginning the school year at the 10th percentile only showed a 20-point increase in words correct per minute by spring. In comparison, those at the 90th percentile showed a 37-point increase while those at the 50th percentile showed a 50-point increase by the spring timepoint. This leads to an increase from a beginning of second-grade difference of 88 words correct per minute between the 10th percentile and the 90th to a difference of 105 words correct per minute by spring. When comparing

the 10th and 50th percentiles, there is a fall gap of 27 words correct per minute which grows to a gap of 57 words per minute by spring (Hasbrouck & Tindal, 2017). Although Hasbrouck and Tindal's (2017) report focused on percentile benchmark analysis of oral reading fluency scores for the purposes of progress monitoring, their data shows the potential for Matthew Effects in within-year growth.

Hasbrouck and Tindal (2017)'s data analyzes percentile ranks overall, not the progression of individual students. However, further research on rank-order stability suggests that their findings do indeed relate to within-year Matthew Effects. Bast and Reitsma (1998)'s analysis of Dutch longitudinal reading data showed the existence of rank-order stability within reading scores. In their 3-year longitudinal study of 235 Kindergarten students in the Netherlands, students showed strong rank-order stability, meaning those who entered Kindergarten with lower skills not only grew slower than their peers but remained similarly ranked three years later. This rank-order stability paired with slower growth rates amongst lower-achieving students led to growing performance gaps, giving evidence of a Matthew Effect. Bast and Reitsma (1998)'s analysis of longitudinal rank-order stability suggests that Hasbrouck and Tindal (2017)'s analysis of oral reading fluency achievement by percentile group likely does show evidence of the Matthew Effect, as students comprising each percentile group are likely to remain stable over time. Yet how does this compare to within-year growth rates specifically for students in the early grades?

Silberglitt and Hintze (2007) studied within-year growth rates on curriculum-based measures across three time points for elementary school students in Grades 2-6

in five Minnesota school districts. In their examination of 7,544 students using hierarchical linear modeling, they found differences in rate of growth (i.e., slope between the Fall, Winter, and Spring time points) based on beginning level of performance. Using district collected Aimsweb reading data, Silberglitt and Hintze (2007) grouped students based on their fall benchmark scores into ten percentile groups for each grade of their cross-sectional analysis. Silberglitt and Hintze (2007) represented their results in differences in parameter estimates between their 50th-59th percentile reference groups and other decile groups based on weekly growth rates.

Silberglitt and Hintze (2007) did not find that students who initially scored into the highest decile groups had stronger rates of growth than their comparison peers at the 50th-59th percentile, rather finding the highest percentile scoring students had lower than average rates of growth. Therefore, Silberglitt and Hintze (2007)'s data did not show a fan spread, although it did show a widening gap between students. These results were most pertinent in that they show significantly slower growth rates for the lowest ability students. Results also showed stronger differences in growth rates between percentile groups in grades 2-3 than in grades 4-6, suggesting that these earlier grades are where Matthew Effects are more prevalent. Silberglitt and Hintze (2007)'s results indicate the opportunity for further analysis of within-year Matthew Effects, concentrating on the earlier elementary school grades.

However, their results do show some evidence of a within-year Matthew Effect. Students scoring in the lowest percentile groups show significantly lower rates of growth than not at-risk peers. The authors found, "slopes of growth were significantly

lower for both the bottom and top deciles at most grade levels. In addition, a greater

number of bottom deciles were significantly lower in the early grades" (pg. 75). Pairing

with results from Bast and Reitsma (1998), Silberglitt and Hintze (2007) suggest that

Hasbrouck and Tindal's (2017) oral reading fluency norms do suggest evidence of a

Matthew Effect, as it is likely individual students for each decile group remained

constant throughout their academic careers and that the lowest decile groups showed

significantly lower growth rates than their average, 50th percentile ranking peers.

Figure 2
Silberglitt and Hintze (2007)'s Growth Curves by Decile at Grade 2 (From Silberglitt and Hintze, 2007, pg. 78)
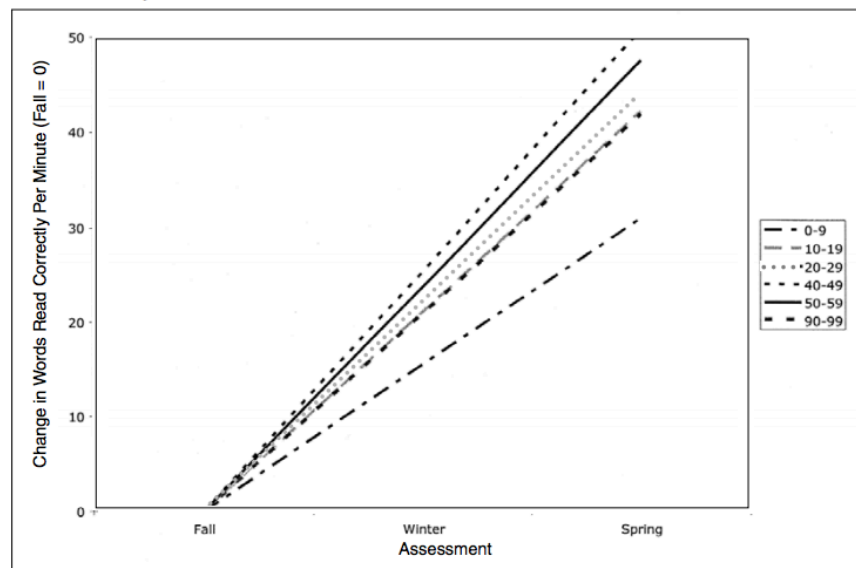
Achievement Inequities

Although many students show early reading difficulties, there is a persistent

achievement difference between students of varying demographic backgrounds

*(*Chatterji, 2006*).* The National Assessment of Educational Progress (NAEP) explains that

"achievement gaps occur when one group of students (e.x. students grouped by

race/ethnicity, gender) outperforms another group and the difference in average scores for the two groups is statistically significant" (National Center for Educational Statistics, 2018).

NAEP's 2019 4th grade literacy scores identified Black, Hispanic, and Native American / Alaskan Native students as having statistically significantly lower scores than their White peers. See Table 1 for the NAEP Data Explorer generated report of 2019 4th grade literacy scores by race/ethnic group then compared score differences for statistical significance. All of the scores were shown to be significantly different ($p < .05$) from White students.

Table 1
2019 Average Scale Scores in Literacy by Race

Race/Ethnicity	Average Literacy Scale Score 2019 NAEP
White	230
Black	204
Hispanic	209
Asian / Pacific Islander	237
American Indian / Alaska Native	204
Two or More Races	226

Note. Data from the NAEP Data Explorer, National Center for Educational Statistics

The work of education researcher Dr. Gloria Ladson-Billings (2006) identifies these achievement gaps as being representative of historically accumulated inequities due to an education debt. Ladson-Billings explains that existing achievement gaps are instead reflections of historical inequities amongst groups whom systemic inequalities in

the education system have victimized. Ladson-Billings cites Randall Robinson (2000) in making her argument that these achievement gaps are mere reflections of a societal debt owed to communities historically excluded from equal opportunity:

> "No nation can enslave a race of people for hundreds of years, set them free bedraggled and penniless, pit them, without assistance in a hostile environment, against privileged victimizers, and then reasonably expect the gap between the heirs of the two groups to narrow. Lines, begun parallel and left alone, can never touch" (Robinson as cited in Ladson-Billings, 2006, pg. 8).

Amongst Indigenous students, the public education system originally existed as religious mission and boarding schools that forced Christian conversion and used Indigenous students as forced labor. Ladson-Billings references an 1864 U.S. Congressional decision that legally prohibited Native American students from accessing education in their indigenous languages. For Black students, due to our nation's history of slavery, Black individuals were originally denied the right to learn to read and later excluded from whites only schools through segregation. These educational inequities were enshrined in law until the Civil Rights movement. Nevertheless, persistent inequities are reflected in the U.S. public education system with significant differences in achievement and school funding, as majority White schools spend more per pupil than schools that serve students of color (Ladson-Billings, 2006). Through a historical and social-justice lens, it is clear that the achievement gap is a reflection of societal inequity and opportunity.

Hispanic, Latino/a, and Latinx Students

It is important to first note, that although NAEP refers to this group as Hispanic, this term encompasses students who also identify as Latino/a and Latinx. Hispanics accounted for half of all population growth in United States between 2000 and 2016.

Now as the largest ethnic or racial minority group in the United States, Hispanic students account for 25% of elementary school enrollment (Bauman, 2017). Relatively, the proportion of English Language Learning (ELL) students in U.S. public schools is growing, with ELL students comprising 9.6% of the total school population or almost 5 million students by 2016. For these students, Spanish is the most common home language, representing 76.6% of all ELLs and 7.7% of all enrolled K-12 students in public education (National Center for Educational Statistics, 2019). Hispanic students are increasingly enrolled in schools where diverse students comprise at least 75% of the student body. In the Fall of 2017, 60% of Hispanic students were enrolled in such schools, and increase from 56% in the Fall of 2000 (NCES, 2020a).

Over the past fifteen years, the U.S. public education system has shown a persistent achievement gap between the educational outcomes of White and Hispanic students. This achievement gap is evident when students enroll in Kindergarten and persists throughout their elementary school years (Chatterji, 2006). In 2009, the National Assessment for Educational Progress showed a 25-point achievement gap between assessment results for 4th-grade students and a 24-point achievement gap for 8th-grade students (Hemphill et al., 2011). However, there are signs that the achievement gap between white and Hispanic students in literacy is narrowing. Between 2003 and 2009, NAEP reported that the overall achievement gap in literacy was narrowing between white and Hispanic students, although three states, California, Connecticut, and Rhode Island, had widening achievement gaps (National Center for Educational Statistics, 2011).

African-American and Black Students

     African-American and Black students have a long history of experiencing systematic inequality in the U.S. education system. In 2017, there were 7.7 million Black children enrolled in U.S. public schools, comprising 15% of the overall student population (NCES, 2020a). The proportion of Black students in public education decreasing, having shrunk from 17% of the overall student population in 2000 to 15% in 2017 (NCES, 2020a). Similar to Hispanic students, Black students are generally concentrated in schools where diverse students make up at least 75% of the student population. In the Fall of 2017, 58% of Black students were enrolled in schools with majority diverse students, an increase from 51% in the Fall of 2000 (NCES, 2020a). In the 2019 NAEP reading report card, Black students scored on average 204 points, in comparison to an average of 230 points for white students, leaving a 26-point discrepancy at the 4th grade level (NCES, 2019b). At the 8th grade level, the contrast is similar with a 28-point achievement gap between white students who scored 272 and Black students who scored 244 (NCES, 2019b).

     While the Hispanic-White achievement gap is decreasing overall, the Black-White achievement gap has had no such gains. Between 1992 and 2007, the Black-White reading achievement gap only narrowed in three states: Delaware, Florida, and New Jersey (NCES, 2009). The NAEP (2009) showed no significant progress in closing the gap between 1980 and 2004, although overall reading scores for both Black and white students increased (NCES, 2009).

Indigenous Students

Although the majority of research in achievement inequities centers on Black or Hispanic students, those of Native American and Alaskan Native descent are also identified by the 2019 NAEP reading report card as having statistically significantly lower scores in comparison to White peers. With scores comparable to Black students, Native American and Alaskan Native students scored on average 204 points, leaving a 26-point discrepancy in reading achievement at the 4th grade level in comparison to their white peers (NAEP, 2019). Indigenous communities in the United States have historically had complicated relationships with the public education system, as over the last century, the U.S. government has used public education as a weapon to colonize and assimilate indigenous populations (Faircloth, 2009). A 2008 NCES report, Status and Trends in the Education of American Indians and Alaska Natives, Indigenous students only represented 1% of the overall K-12 student population with 644,000 students overall from 560 federally recognized tribes. Of these students, 27% lived in poverty, a rate two times higher than the general U.S. population (NCES, 2008).

Socioeconomically Deprived Students

Students from economically deprived backgrounds are especially at risk for low reading scores. On the 2009 National Assessment of Educational Progress, 83% of 4th-grade children from low-income families failed to reach grade-level proficiency in reading, with 49% of low-income 4th-grade children failing to gain even basic literacy skills (National Center for Educational Statistics, 2009). In the 2019 NAEP reading report card, there were achievement inequities between students who were eligible for free

and reduced-price lunch and those who were not eligible. Fourth-grade students living in poverty, as defined by NSLP eligibility, scored 207 points in comparison to their non-impoverished peers, who scored 235 points leading to a difference of 28 points. At the 8th grade level, those living in poverty scored 250 points while those not living in poverty scored 275 points, leading to a difference of 25 points (National Center for Educational Statistics, 2019).

Analysis of data from 12,261 students in the Department of Education's Early Childhood Longitudinal Study has shown that when dividing Kindergarten students into low, average, and high literacy ability groups, 33% of the students in the low literacy level group were from families with a low socioeconomic status where only 4% of the students in the high literacy level group were from facing poverty. By the end of the three-year longitudinal study, students in the low literacy level could not match their high achieving Kindergarten classmates' skills until the third grade (Foster & Miller, 2007). Although income differences are found in most countries, a longitudinal cross-national analysis of twenty countries in the Organization for Economic Co-operation and Development showed the United States ranked as having one of the highest income-based achievement gaps in education (Chmielewski & Reardon, 2016).

Buckingham et al. (2013) reviewed links between socioeconomic disadvantage and literacy achievement throughout English-speaking countries. They describe that socioeconomic status is a composite variable, compiled by several different factors that independently correlate with children's literacy achievement, including household income, parental occupation, and parental education level (Buckingham et al., 2013).

They explain, "gaps in children's literacy abilities are evident when children begin school, with children from low socioeconomic backgrounds tending to demonstrate lower proficiency in the two main aspects of emergent literacy— phonological awareness and vocabulary / oral competency"… and "that low SES students are more likely to remain poor readers if they begin school as poor readers" (Buckingham et al., 2013, pg. 4). Buckingham et al. (2013)'s results linking socioeconomic disadvantage with lower reading achievement suggests the importance of including these students in analyses of achievement gaps.

In their 2005 study, Rumberger & Palardy analyzed data from the 1988 National Education Longitudinal Survey to study achievement and the effects of racial and socioeconomic segregation in U.S. public schools. Rumberger & Palardy (2005) focused their research on questions surrounding the importance of school desegregation programs on academic achievement, using data from 14,217 students across 913 schools. The authors were especially interested in the idea of compositional effects or rather that student outcomes are not only related to their own individual demographic characteristics but also to their school's overall or aggregate characteristics. Rumberger & Palardy (2005) found "the average socioeconomic level of students' schools had as much impact on their achievement growth as their own socioeconomic status, net of other background factors" (pg. 1999). They hypothesized school SES indirectly affected achievement growth due to higher percentages of students reporting feeling unsafe, teachers' lower expectations of students in high SES areas, students completing less homework, and a lower average number of college prep courses. Their results echoed

the seminal 1966 Coleman to Congress that "composition of the student body is more highly related to achievement, independent of the student's own social background, than is any school factor" (pg. 325). Rumberger & Palardy (2005)'s study promoted SES integration in schools as a way to lessen achievement gaps as they found school socioeconomic composition to be "as large, and sometimes much larger than the effect of student SES on achievement growth" and that "what appears to matter most is the socioeconomic, not the racial composition of schools" (pg. 2020).

Schools as the Level of Comparison

Literacy is often analyzed using a student level of analysis, that is looking at each individual student as a discrete data point, there is increased interest in looking at schools as the unit of analysis (Hanushek & Rivkin 2006; Holmes-Smith, 2006; Palardy, 2008; Stiefel et al., 2007; Thomson & De Bortoli, 2010). These studies have focused on between-school factors and comparisons of literacy acquisition to study systemic barriers and influences on student growth.

The average socioeconomic level of the school has been shown to have a stronger relationship with student literacy achievement than individual students' socioeconomic status (Holmes-Smith, 2006; OECD, 2010; Thomson & De Bortoli, 2010). The Organization for Economic Co-operation and Development (OECD) 's Program for International Student Assessment (PISA) compares academic outcomes for 15-year olds every three years. PISA's 2010 report, Overcoming Social Background, analyzed results from the 2009 report showing on average, 42% of variance in reading scores could be attributed to between-school differences rather than within-school differences. Adding

22

another layer of analysis, the OECD identified that over 70% of the variance of literacy scores between schools were explained by socioeconomic status in the United Kingdom, the United States, and New Zealand (OECD, 2010). According to the OECD, "In the majority of the OECD countries, the relationship between the average economic, social and cultural status of students in a school and their performance is steeper than the relationship between the individual student's socioeconomic background and their performance in the same school" (OECD, 2010, pg. 92). The OECD's study on students' social backgrounds is a compelling look at the influences of a school's overall socioeconomic status at a single time-point in students' educational careers,

Logan and Petscher (2009) analyzed reading scores for 175,857 who participated in Florida's Reading First program. They completed an analysis that separated schools into clusters based on concentrations of historically at-risk groups within the school in comparison to their sample mean, then analyzed CBM growth over time for each school cluster. Their comparison variables for risk included low-risk schools, average risk schools, those with a higher proportion of students receiving free and reduced-price lunch and minority students, as well as schools who not only higher a higher proportion of students receiving free and reduced-price lunch and minority students but also English Language Learners.

Logan and Petscher (2009) found significant differences in within-year ORF growth rates based on a school's risk level, with the Low-Risk group a slope of +4.01 WCPM/month and the Language-Risk group having a slope of +3.08 WCPM/month during the first grade. Although all four clusters of schools had increasing literacy scores,

the gaps between clusters grew. Looking more closely at the differences between the highest-scoring group (low-risk) and schools in the lowest scoring group (language risk), differences in scores approximately doubled between September and April. While the Logan and Petscher findings are compelling, further investigation of Matthew Effects using schools as the unit of comparison is needed.

Dynamic Indicators of Basic Literacy Skills

The Dynamic Indicators of Basic Literacy Skills (DIBELS) are curriculum-based measurements designed as a general screening assessment for early reading ability and provide valuable insight into students' reading levels. As curriculum-based measures, DIBELS assess basic literacy skills students will later need to become fluent and skilled readers. The DIBELS assessment is meant to measure student progress over time in acquiring skills identified by the National Reading Panel such as fluency, phonological awareness, alphabetic principal, oral language, and comprehension (DIBELS, 2019).

The Dynamic Indicators of Basic Literacy Skills- 8th edition (DIBELS-8) is used in schools to provide universal screening for literacy in the elementary school grades. DIBELS-8 includes a set of five short individually administered subtests that measure students' skills in the fundamental areas of reading: letter naming fluency (LNF), Phonemic Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Word Reading Fluency (WRF), Oral Reading Fluency (ORF), and Maze. Most measures take only 60 seconds to administer, with only Maze taking longer at 3 minutes (DIBELS[1], 2019).

DIBELS-8 has strong alternate form reliability (.90+) for WRF, and ORF, and good reliability (.80+) for NWF. Concurrent validity was calculated as a comparison between

DIBELS-8 and DIBELS Next; for the second-grade subtests validity scores were (.8) in the Fall, (.62) in the Winter, and (.74) in the Spring. Predictive validity compared DIBELS-8 scores with the Iowa Assessment Total Reading score and were (.39) in the Winter and (.5) in the Spring (DIBELS, 2018).

Oral Reading Fluency

Reading fluency has long been used in research as an indicator of students' true reading abilities. Students who are fluent with reading are those who can read with both speed and accuracy. In its most essential form, fluency is the ability to read correctly but also quickly and is known to be a critical component of literacy. A report by the National Research Council explains, ""Adequate progress in learning to read English (or, any alphabetic language) beyond the initial level depends on sufficient practice in reading to achieve fluency with different texts" (Snow, Burns, and Giffin, 1998). Oral Reading Fluency is a measurement indicator of a students' speed and accuracy in text read out loud. Oral reading fluency in the early elementary grades has been shown to be a strong indicator of later reading ability. Students who are able to decode text quickly are able to focus instead on higher level comprehension skills rather than spending time decoding words inefficiently. This is due to the reading process being defined mainly as two inherent cognitive tasks; first the decoding of specific words then the students' construction of meaning from the text. Students who possess low reading fluency skills are not able to inherently comprehend meaning from text as they spend valuable cognitive resources on decoding (Report of the National Reading Panel, 2000).

As measurements of student's oral reading accuracy and rate, DIBELS Oral Reading Fluency probes asks students to read out loud from a grade-level text for one minute. These passages are written by professional authors and include both narrative and informational text. Oral Reading Fluency texts are standardized in length with first and second-grade passages being between 150-200 words and third grade passages being between 175-225 words. Passages are aligned with Flesch-Kincaid grade levels with first grade probes being 1.5-2.0, second-grade probes being 2.5-3.0, and third grade probes being 3.5-4.0. Internal and external teams of reviews assessed passages for theme, background knowledge requirements, and content acceptability to ensure appropriate grade level (DIBELS 8- Administration Guidelines, 2018-2019).

When utilizing DIBELS Oral Reading Fluency probes, an assessor, usually the classroom teacher, gives a standardized instruction, "please read this out loud. If you get stuck, I will tell you the word, so you can keep reading. When I say 'Stop' I may ask you to tell me about what you read, so do your best reading. Start here. Ready? Begin." (DIBELS 8-Administration Guidelines, 2018-2019, pg. 57). The assessor follows along in the passage as the student reads aloud, with the assessor marking every incorrectly read word with a slash (/). If the student hesitates for more than 3 seconds on a word, the assessor provides the word and marks it as incorrect. If a student does not read any words correctly within the first line of the story, the subtest is discontinued, and the student receives a score of 0. Each student receives two ORF scores, the total number of words read correctly and an accuracy percentage (DIBELS 8-Administration Guidelines, 2018-2019).

DIBELS-8's ORF has strong concurrent alternate form reliability for second grade, with the median reliability being .93–.97. Median DIBELS delayed alternate form reliability, correlating benchmark forms, was .86–.92 for ORF's words read correctly in the second grade. ORF test-retest reliability between Fall and Spring was .83–.91 in the second grade. Mean second grade intercept reliability for ORF is .93–.99 for words read correctly while the slope reliability is .45–.99 (DIBELS-8 Technical Manuel, 2020)

Since its creation, DIBELS Oral Reading Fluency (ORF) has been used in numerous studies to predict students' likelihood of future reading difficulties. Numerous studies have demonstrated strong predictive abilities with correlational between students' oral reading fluency scores using DIBELS probes and later reading comprehension scores. Riedel (2007) tested 1, 518 first grade, predominantly African American students in the Memphis City School district. Participants were first-grade students in the Memphis City Schools district during the 2003–2004 school year. As part of a Reading Excellence Grant, students were assessed with both DIBELS and a grant specific assessment, GRA+DE. Riedel (2007) found that the DIBELS ORF subtest was able to predict proficiency on the GRA+DE reading comprehension subtest with 71-80% accuracy. A study of 34,855 students in the first, second, and third grades in Florida showed that DIBELS ORF scores had predictive validity with students' scores on the Stanford Achievement Test-10th Edition; students who had higher DIBELS ORF scores had greater predictive validity (Petscher & Kim, 2011).

Nonsense Word Fluency (NWF)

DIBELS' Nonsense Word Fluency is a measure of decoding where a student's ability to use letter-sound correspondence and decode novel, pseudo-words. Unlike reading real words, nonsense word fluency controls for exposure effects through having students read made-up words such as 'hap'. In the DIBELS assessment system, nonsense word fluency is measured from Fall of Kindergarten to Spring of third grade. DIBELS-8th edition's nonsense word lists only include phonetic consonant and vowel combinations present in English, with more frequently used spelling patterns appearing more often in the testing materials. (DIBELS 8-Administration Guidelines, 2018-2019). Example words include, "bace, zart, melb, scap, brold, geap, foddy, cotalm, and fudlerk" (DIBELS 8-Administration Guidelines, 2018-2019, pg. 19).

When testing Nonsense Word Fluency, the probes contain two practice items. An assessor presents the student with the word 'hap' explaining, "look at this word, it's a make-believe word. Watch me read the word: /h/ /a/ /p/, 'hap', I can say the sounds of the letters, /h/ /a/ /p/ or I can read the whole word 'hap'" DIBELS 8-Administration Guidelines, 2018-2019, pg. 64). The student then has the opportunity to read a new pseudo-word, 'lum' with the accessor providing correction for errors. Following the teaching items, the student is presented with a list of nonsense words to read aloud. All students receive one minute to read as many as they are able; second graders' lists contain 100 nonsense words. The assessor follows along in the list as the student reads aloud, with the assessor marking every incorrectly read word with a slash (/). If the student hesitates for more than 3 seconds on a word, the assessor provides the word

and marks it as incorrect. If a student does not read any words correctly within the first five words, the test is discontinued. The nonsense word probe results in two scores, the sum of the correct letter sounds and the sum of words read or recoded correctly (DIBELS 8-Administration Guidelines, 2018-2019).

DIBELS-8 NWF has strong concurrent alternate form reliability for second grade, with the median reliability being .89–.94 for word recoded correctly. Median DIBELS delayed alternate form reliability, correlating benchmark forms, was .74–.87 for NWF, words recoded correctly, in the second grade. NWF test-retest reliability between Fall and Spring was .62–.80 in the second grade. Mean second grade intercept reliability for NWF is .41–.99 for words recoded correctly while the slope reliability is .41–.99. (DIBELS-8 Technical Manuel, 2020).

Word Reading Fluency (WRF)

DIBELS Word Reading Fluency is a measure of alphabetic principal and fluency in text. Probes include a mixture of decodable words and non-decodable sight words such as 'the', 'we' and 'of' DIBELS 8-Administration Guidelines, 2018-2019). Word Reading Fluency word lists included items from Kucera and Francis (1967), Dale and O'Rourke (1981), Lund and Burgess (1996), as well as Brysbaert and Biemiller (2017), excluding words that were unlikely to be known by the average student.   At each grade level, Word Reading Fluency probes only include words that children typically acquired orally to reduce the likelihood of students being tested on novel words. Forms begin with more common words, words becoming more difficult as the student progresses through

the form. Grade 2 probes includes up to three syllables in tested words (DIBELS 8-Administration Guidelines, 2018-2019).

When testing Word Reading Fluency, the assessor presents the student with the probe and then gives the standardized instruction, "please read from this list of words. Start here and go across the page. When I say 'begin', point to each word and read it the best you can. If you get stuck, I will tell you the word, so you can keep reading. Put your finger on the first word. Read? Begin." (DIBELS 8-Administration Guidelines, 2018-2019, pg. 72). Students receive one minute to read as many words as they are able with second graders receiving a list of 130 words. Like oral reading fluency and nonsense word fluency, the assessor follows along, marking incorrectly read words with a slash (/). If the student hesitates for more than 3 seconds on a word, the assessor provides the word and marks it as incorrect. If a student does not read any words correctly within the first line, the probe is discontinued. Word reading fluency probes provide only one score, the total number of words the student read correctly in one minute. Unlike in nonsense word fluency or oral reading fluency, correct sounds without recoding or blending are not scored and are marked incorrect (DIBELS 8-Administration Guidelines, 2018-2019).

DIBELS-8 has strong concurrent alternate form reliability for second grade word reading fluency (WRF), with the median reliability being .93–.97. Median DIBELS delayed alternate form reliability, correlating benchmark forms, was .88–.94 for WRF in the second grade. WRF test-retest reliability between Winter and Spring was .93-97 in the second grade with other test-retest reliabilities not being available. Mean second grade

intercept reliability for WRF is .96–.98 while the slope reliability is .37–.99 (DIBELS-8 Technical Manuel, 2020).

Research Questions

1. Do schools differ in second-grade reading CBM scores in the fall on the basis of proportion of students living in poverty and proportion of historically underserved students?

2. Do schools differ in second-grade reading CBM fall to spring change on the basis of proportion of students living in poverty and proportion of historically underserved students?

Hypotheses

I hypothesize that we will see gaps at each time point and that those gaps will widen over time.

1. In accordance with National Center for Educational Statistics (2019)'s data on achievement gaps in literacy, schools with higher proportions of students living in poverty or higher proportions of historically underserved groups will have statistically significant lower reading scores during fall.

2. In accordance with research on the Matthew Effect including Francis et al. (1996) and Stanovich (2000), schools with higher proportions of students living in poverty or higher proportions of historically underserved groups will have statistically significant lower CBM fall to spring change in scores.

CHAPTER III

METHODS

Data Set and Participants

This study utilized school reading benchmark data from the DIBELS Data System with an observational design with extant data. The DIBELS Data System is a not-for-profit repository of school based CBM data that began in 2001 and is managed by the Center on Teaching and Learning (CTL) at the University of Oregon. Schools and districts across the United States and internationally, utilize DIBELS curriculum-based measures then upload benchmark and progress monitoring scores for students from Kindergarten through the 8th grade (CTL, n.d.a). Serving thousands of schools, the DIBELS Data System now includes information from 20% of elementary schools nationally (CTL, n.d.b). The Data System allows participating schools to generate progress monitoring reports, analyze response to intervention data, and set literacy goals for students (CTL, n.d.c).

This study included benchmark data from DIBELS Data System users that completed assessments for second grade students during the 2018-2019 school year. The DIBELS Data System includes schools' National Center for Education Statistics (NCES) IDs which allow for linking to school-level demographic information. DIBELS-8 Data System uses uploaded a total of 15,398 applicable unique assessment results into the system. As the DIBELS Data System allows for private schools, tutoring centers, or other similar organizations to upload data, not all applicable users with relevant data were public schools. A total of 73 DIBELS Data System users completed both Fall and Spring benchmark data for second grade students during the 2018-2019 school year. NCES IDs

were verified or identified by user's addresses and organization name with the publicly available NCES Common Core of Data (CCD) School Search. Five users appeared to be tutoring centers or non-profit organizations and were excluded for not being schools.

As the primary conceptualization of this study is to understand how the public-school education system is currently ameliorating societal inequalities, it centers on the analysis of public-school reading data. Of the 68 remaining schools, 21 schools were private schools and were therefore excluded from the sample. As these schools do not fit within the study's core analysis of public education data, private schools were excluded from final analyses.

Three DDS users in the study had classified themselves in the database as being separate schools within the same charter school district but were considered by NCES to be the same charter school, sharing one NCES ID number. These three users were combined to be one school within the sample, bringing the total number of public schools to 44. NCES considered charter schools to be public schools and they are included in the NCES Common Core of Data (CCD) as such. NCES defines "a public charter school is a publicly funded school that is typically governed by a group or organization under a legislative contract—a charter—with the state, the district, or another entity" (NCES, 2020b). Of the schools in the final analytic sample, four were identified as being charter schools.

According to NCES statistics published in 2019, by the 2017-2018 school year, charter schools comprised 7% of all public schools (NCES, 2020b). Approximately 56% of all charter schools in the United States were elementary schools, with only 23% being

secondary schools, the remaining enrolled both elementary and secondary students.

Most charter schools enrolled less than 300 students (45% compared to only 28% of

non-charter public schools). Charter schools in the United States are also more likely to

serve predominantly students of color, with almost half enrolling at least 50% Hispanic

or Black students compared to only a quarter of non-charter public schools doing so

(NCES, 2019a).

Universal Screening

Universal screening is an educational practice in literacy considered a "critical

first step in identifying students who are at risk for experiencing reading difficulties"

(Gersten et al., 2008, pg. 17). The What Works Clearinghouse considers universal

screening to have a moderate level of research evidence and recommends that all

students be screened at the beginning of the school year and more frequently for

students considered at-risk (Gersten et al., 2008). For this study, the analytic sample

includes schools that used DIBELS as a universal screening tool, meaning they assessed

at least 85% of their enrolled second grade students at both the fall and spring time

points. Schools' total number of second grade students administered DIBELS probes in

both Fall and Spring during the 2018-2019 school year were compared to their overall

second grade population according to NCES. Schools were excluded from analysis if they

used DIBELS to assess less than 85% of students to prevent the introduction of

confounding, unmeasurable variables as school's selection processes for testing some

students and not testing others are unknown.

Additionally, not each school tested each student on every applicable measure (NWF, ORF, and WRF) during both time points. Schools were included in the analytic sample for specific measures if they assessed enough students on that measure. Of the 44 schools, 23 schools did not use any of the three DIBELS probes as universal screening measures and were dropped from the sample. One school only enrolled one second-grade student; this school was dropped from the sample due to its incredibly small size. Only students who had both fall and spring benchmark data were included. Of the original schools, 22 completed universal screening measures. See Table 2 for information on the number of included schools. As not all schools used each measure as universal screening, the number of schools included in the analyses changes based on the measure.

Dependent Variables

This study includes each school's average (mean and median) second grade DIBELS score for the Fall and Spring time points. Fall scores and gains from Fall to Spring serve as the dependent variables of interest for both research questions one and two. Scores from three different DIBELS subtests were analyzed as continuous variables: Nonsense Word Fluency- Correct Letter Sounds, Nonsense Word Fluency-Words Read Correctly, Oral Reading Fluency, and Word Reading Fluency. Only students with both fall and spring scores in DDS contributed to the mean and median school scores.

Independent Variables

This study includes analyses of two independent variables: concentration of historically underserved students and concentration of poverty. Schools' National

Center for Education Statistics school identification numbers were cross-referenced with each school's listing in the Common Core of Data (CCD) online database, which gives publicly accessible information on enrollment characteristics. Demographic variables for each school were compiled from their listings in the CCD database.

Concentration of Students from Historically Underserved Groups

The variable 'Concentration of Historically Underserved Students' was calculated using the Common Core of Data (CCD) demographic information on each school's second-grade population during the 2018-2019 school year. The CCD is a publicly available database that includes data on all public elementary schools, secondary schools, and school districts. The CCD publishes data annually and is searchable by school year (CCD, 2019).

Concentration of students from historically underserved groups is a continuous variable calculated at the grade level based on the proportion of 2nd-grade students enrolled during the 2018-2019 school year who are from racial or ethnic groups identified by the U.S. Department of Education as having statistically significant lower reading scores than their more racially privileged, white peers. Based on 2019 4th grade NAEP literacy scores, Black, Hispanic, and American Indian / Alaska Native students have statistically significantly lower reading scores than their white peers (National Center for Educational Statistics, 2019). For each school, Concentration of Historically Underserved Students was first calculated using the second-grade population of historically underserved students (Black, Hispanic, or American Indian / Alaska Native) divided by

the total second-grade student population. The concentration was grand mean-centered for the sample before data analysis.

School Concentration of Poverty

Title I Status. Concentration of students living in poverty was first examined as a categorical variable representing a school's Title I status with three categories: none, Title I School, and Title I Schoolwide. Title I is a federal U.S. government program aimed at promoting academic achievement for economically deprived students under the Federal Title I of the Elementary and Secondary Education Act, as amended by the Every Student Succeeds Act (ESEA). Title I provides "schools with high numbers or high percentages of children from low-income families to help ensure that all children meet challenging state academic standards (U.S. Department of Education, 2016). The Common Core of Data (CCD) online database provides school-level information based on three categories: none, Title I School and Title I Schoolwide. To be eligible for Title I Schoolwide programs, at least 40% of the enrolled student population is living in poverty (Whalen, 2015). Federal Title I funding at these schools can be used to serve all enrolled students. Schools that are labeled as Title I Schools are those that have targeted assistance programs. Targeted assistance programs are aimed at schools whose student population meets the following criteria. First, that they enroll at least 10 children living in poverty. That the proportion of children living in poverty enrolled at their school is at least 5% of the student population. These schools provide only targeted academic programs for students who are eligible for free and reduced lunch or those not academically performing at grade level (Illinois State Board of Education, n.d.;

Department of Education, 2018). This information was transformed into a categorical

variable in SPSS with 1- Ineligible, 2- Title I, and 3- Title I Schoolwide. However, of the 22

schools in the analytic sample, none were ineligible for Title I services, and only three

schools fell in the Title I School category, with the 19 remaining schools eligible for Title I

Schoolwide programs. Due to the representation of Title I categories be severely

imbalanced in the sample, analysis focused on proportion of students receiving free or

reduced-price lunch (FARMS) as the indicator of school poverty.

Table 2

Universal Screening and Analytic Sample

Measure	n schools using each measure at both time points	n schools completing universal screening	n schools in final sample	n students in final sample
Nonsense Word Fluency (NWF)	35	21	19	837
Oral Reading Fluency (ORF)	37	22	20	863
Word Reading Fluency (WRF)	31	21	19	797
Total	44	22	20	971

Proportion Free and Reduced Lunch. Free and Reduced-Price Lunches are

funded through the National School Lunch Program (NSLP). Through the program, "Free

lunches are available to children in households with incomes at or below 130 percent of

poverty. Reduced-price lunches are available to children in households with incomes

between 130 and 185 percent of poverty" (U.S. Department of Agriculture, 2020). The school proportion of students receiving free or reduced-price lunch was included for all schools by pairing the NCES School ID information from the CCD online database listing. Table 2 shows the number of schools and students in the final analytic sample. Two schools with universal screening data had not reported FARMS data to the CCD and were excluded from the analyses leaving 19 schools who completed Nonsense Word Fluency, 20 who completed Oral Reading Fluency, and 19 who completed Word Reading Fluency. The proportion of students eligible for free or reduced-price lunch was then calculated by dividing by the total number of students enrolled in the school and coded in SPSS as a continuous variable. As this data is reported on the school level rather than by grade, all proportions reflect the total number of students eligible by school. The concentration was grand mean-centered for the final analytic sample before data analysis.

Data Analysis

A descriptive analysis was first completed regarding the analytic sample. Following the descriptive analysis of the data, the primary data analysis for this study was a series of hierarchical multiple regressions, regressing DIBELS mean and median BOY and Change scores for four subtest scores (i.e., NWF-CLS, NWF-WRC, ORF, and WRF) on school concentration of historically underserved students and school concentration of poverty. Following the first set of regressions, a second model was run incorporating an interaction term between the two independent variables. These multiple regressions use the 'enter' method and include regression coefficients and a

95% confidence interval. Multiple regression analyses were run with independent

variables as both proportions and percentages of historically underserved students and

students living in poverty (FARMS). Results were identical for proportions and

percentage values; therefore, only the proportion results are presented. The

assumptions of a multiple regression were checked through inspection of the residuals;

analyses include the Durbin-Watson test and casewise diagnostics. Analyses were

conducted using SPSS version 27. Model fit, descriptive statistics, part and partial

correlations, collinearity diagnostics will also be calculated by SPSS. The study drew on

two primary sources of data, reading data from the DIBELS Data System as well as

demographic and other data from the National Center for Education Statistics (NCES) for

each school. Data was matched across databases using the NCES School ID number.

CHAPTER IV

RESULTS

Descriptive Results

The full analytic sample included 22 schools which used at least one of the applicable DIBELS measures (NWF, ORF, or WRF) as a universal screening measure. The restricted analytic sample included 20 of those schools which reported FARMS data to NCES. Descriptive analyses include information on both the full and restricted samples.

Table 3

Student Enrollment & Students Tested

	Min	Max	Mean	Median	Standard Deviation
Full Sample (n=22)					
Total School Population	107	795	354.27	384.50	200.60
Second Grade Student Population	12	136	50.14	38.50	37.04
NWF N Students	12	132	46.45	34.00	34.78
ORF N Students	12	132	46.64	36.50	33.92
WRF N Students	12	132	45.76	36.00	34.65
Restricted Sample (n=20)					
Total School Population	107	787	336.40	284.50	183.17
Second Grade Student Population	12	132	46.55	38.50	33.23
NWF N Students	12	116	44.05	37.00	30.86
ORF N Students	12	116	43.15	36.50	29.55
WRF N Students	12	116	41.95	36.00	29.91

In the full and restricted sample, the following student enrollment demographics were observed during the 2018-2019 school year. Table 3 shows second grade and full school student enrollment for the analytic sample. Second-grade students assessed by each school at both the Fall and Spring time points contributed to calculating the school's mean and median BOY and Change scores. Only schools which assessed at least 85% of their second-grade population were included in the analytic sample. Table 2 also shows descriptive statistics regarding the number of students contributing to each school's mean and median scores for the schools in the full and restricted analytic samples.

Geographic locations according to U.S. census regions as well as type of district (rural, town, suburban, city) are described for included districts according to NCES locale rules. The full sample of 22 public schools included in the analytic sample were from a total of 14 different states across census regions. The restricted sample of 20 public schools had 12 different states represented. For both samples, a little under half of schools were from the Northeast with New York State had the largest number of schools represented with 5 in both analytic samples. In comparison to national school data, rural schools are somewhat overrepresented in the analytic sample; they comprise over 60% of schools in each analytic sample whereas they only comprised 28% of schools nationwide in 2017 (NCES, n.d.). Table 4 shows school breakdown by region locale for all schools in the full and restricted analytic samples.

Table 4

School Location

	Full Sample (n = 22)	Restricted Sample (n = 20)
U.S. Census Region		
Northeast	9	8
South	2	2
Midwest	4	3
West	7	7
Locale		
City	1	1
Suburb	2	1
Town	4	4
Rural	15	14

Table 5 shows the prevalence of different categories of Title I program eligibility in the full and restricted analytic sample by DIBELS outcome measure. Variation in counts between the outcome measures is due to schools selectively administering subtests. In other words, more schools used ORF universally than WRF and NWF. Given the very low counts of Title I eligible and the lack of ineligible schools compared to schoolwide eligible schools, it was determined at this point to drop Title I eligibility as an independent variable and use the proportion of students eligible for FARMS as a continuous indicator of school concentration of poverty instead. Given the more severe limitations of the Title I categorical variable, an indicator of school concentration of poverty was nonetheless deemed suitable for further analysis. Note again that two schools in the original analytic sample did not report FARMS data to the National Center for Educational Statistics and were dropped from these analyses. Table 5 also shows the

number of schools available for analysis in this restricted sample. All further results are

reported for only this restricted sample of 20 schools.

Table 5

Title I Eligibility

	NWF	ORF	WRF
Title I Analytic Sample	n = 21	n= 22	n=21
Title I Schoolwide Eligible	18	19	18
Title I Eligible	2	3	3
Title I Ineligible	0	0	0
Total Schools = 22			
FARMS Analytic Sample	n = 19	n = 20	n = 19
Title I Schoolwide Eligible	17	18	17
Title I Eligible	2	2	2
Title I Ineligible	0	0	0
Total Schools = 20			

Table 6 reports the descriptive statistics for school concentration of historically

underserved students and poverty. In regard to the proportion of historically

underserved students, the data was positively skewed. Although the mean was .25

when individual school values were examined, 11 schools enrolled less than 10%, and of

that four enrolled no second-grade students from historically underserved backgrounds.

Six schools enrolled half or more of their students from historically underrepresented

groups. In regard to the proportion of students eligible for FARMS, the data was more

normally distributed. The mean was .54 and the median .57. Forty percent of schools

had under 50% of students qualify for Free or Reduced-Price Lunch.

Table 6

School Proportion of Poverty & Historically Underrepresented Students, N = 20

	FARMS	Historically Underrepresented Students
Mean	.54	.25
Median	.57	.08
Standard Deviation	.14	.28
Skewness	-.62	.78
Kurtosis	.30	-1.18
Minimum	.19	.00
Maximum	.76	.75

DIBELS Benchmark Scores

This study uses DIBELS Benchmark scores for Nonsense Word Fluency-Correct Letter Sounds (NWF-CLS), Nonsense Word Fluency-Words Read Correctly (NWF-WRC), Oral Reading Fluency (ORF), and Word Reading Fluency (WRF). The study uses both mean and median fall scores as well as fall to spring mean and median gains. The mean and median scores reflect second-grade students assessed during the appropriate time points in the 2018-2019 school year. The fall to spring mean and median change was calculated through a simple subtraction method in SPSS, by fall DIBELS assessment scores being subtracted from each appropriate spring DIBELS score. Table 7 shows descriptives regarding mean and median NWF-CLS, NWF-WRC, ORF, and WRF scores for the analytic sample. Descriptives regarding the number of students whose scores contributed to schools' mean and median scores at each time point can be seen in Table

8. The smallest school in the analytic sample assessed 12 students while the largest assessed 116, which was consistent across measures.

Nonsense Word Fluency scores are reported as both Correct Letter Sounds (CLS) and Word Read Correctly (NWF-WRC) for all schools that used NWF as a universal screening measure for second-grade students in the analytic sample. As NWF-CLS and NWF-WRC scores are calculated from the same probes, these numbers are consistent for both types of scores. Of the 19 schools which used NWF as a universal screener, a mean of 44.05 students had both fall and spring benchmark scores. Within-school score variation can be seen in Figures 3-6 for NWF. ORF scores reflect the number of correct words read in a minute by students from a standardized story passage. Of the 20 schools which used ORF as a universal screener, a mean of 43.15 students had both fall and spring benchmark scores. Within-school score variation can be seen in Figures 7 and 8 for ORF. WRF reflects the number of correct words read in a minute by students from a list of words. Of the 19 schools which used WRF as a universal screener, a mean of 41.95 students had both fall and spring benchmark scores. Within-school score variation can be seen in Figures 9 and 10 for WRF.

Table 7

Second Grade DIBELS Scores: NWF

	NWF-CLS				NWF-WRC			
N Schools	19				19			
	Mean		Median		Mean		Median	
	BOY	Change	BOY	Change	BOY	Change	BOY	Change
Mean	63.85	38.27	53.47	34.63	19.01	12.03	15.55	11.53
Median	66.77	37.16	54.50	31.00	18.58	12.17	15.00	11.50
Standard Deviation	13.23	15.12	12.48	14.06	2.71	2.86	2.91	2.70
Skewness	-1.83	1.39	-1.41	1.62	.54	-.16	-.001	-.44
Kurtosis	3.47	2.90	2.87	4.79	-.15	-.64	-.47	-.52
Minimum	26.12	19.25	18.00	15.50	14.33	6.57	10.00	6.00
Maximum	78.62	82.41	69.50	79.00	24.76	17.19	21.00	16.00

Note. BOY indicates beginning of year (fall) scores. Change indicates Fall to Spring Gains.

Table 8

Second Grade DIBELS Scores: ORF and WRF

	ORF				WRF			
N Schools	20				19			
	Mean		Median		Mean		Median	
	BOY	Change	BOY	Change	BOY	Change	BOY	Change
Mean	57.36	47.71	55.45	49.12	33.54	18.11	32.18	16.92
Median	56.78	47.10	55.00	47.25	33.67	17.89	33.00	17.00
Standard Deviation	6.00	7.34	9.20	8.45	2.56	3.84	4.79	4.25
Skewness	.47	.45	.85	.16	.56	.34	-.07	.66
Kurtosis	.19	-1.04	.51	-.80	.05	-.30	-.58	.38
Minimum	47.28	38.11	43.00	34.00	29.61	12.06	23.00	11.00
Maximum	71.66	60.96	78.00	64.00	39.55	26.42	41.00	27.00

Note. BOY indicates beginning of year (fall) scores. Change indicates Fall to Spring Gains.

Table 9

Number of Students Contributing to Schools' Mean and Median DIBELS Scores

	NWF	ORF	WRF
Mean	44.05	43.15	41.95
Median	37.00	36.50	36.00
Standard Deviation	30.86	29.55	29.92
Skewness	.97	1.06	1.22
Kurtosis	.25	.66	1.05
Minimum	12	12	12
Maximum	116	116	116

Figure 3

NWF-CLS Beginning of Year Within-School Variability
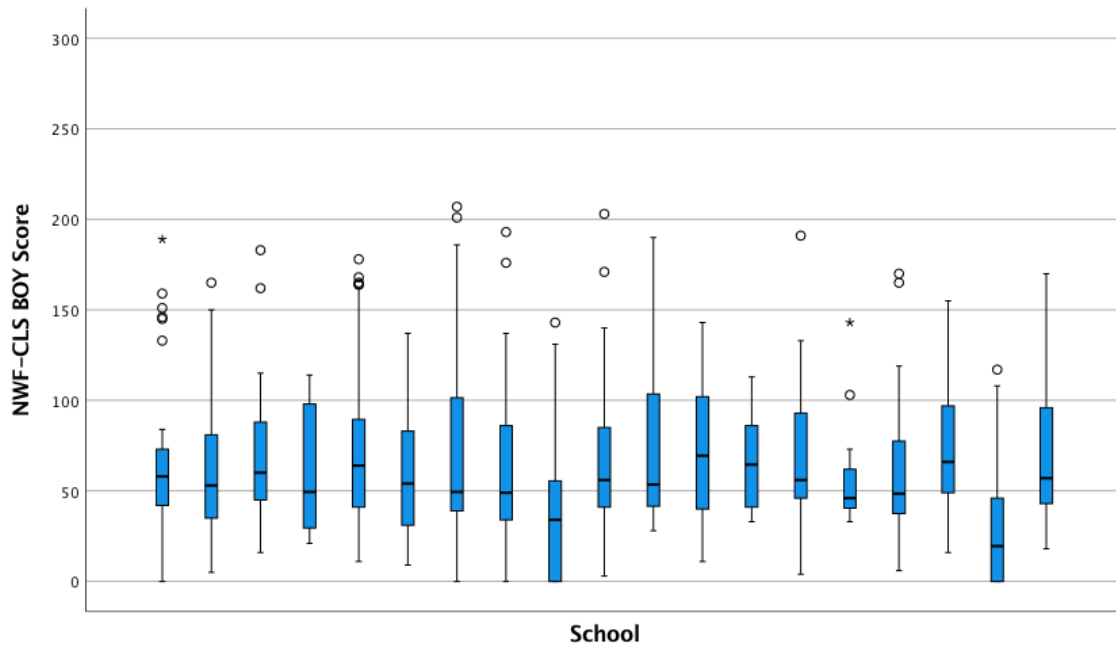
Figure 4

NWF-CLS End of Year Within-School Variability
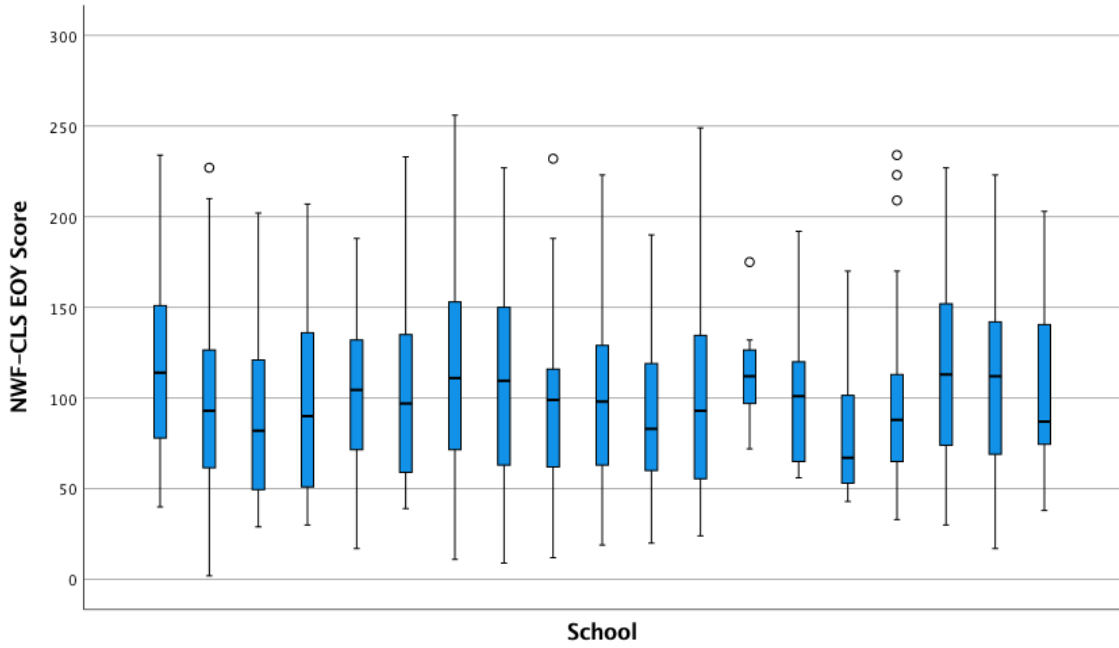
Figure 5

NWF-WRC Beginning of Year Within-School Variability
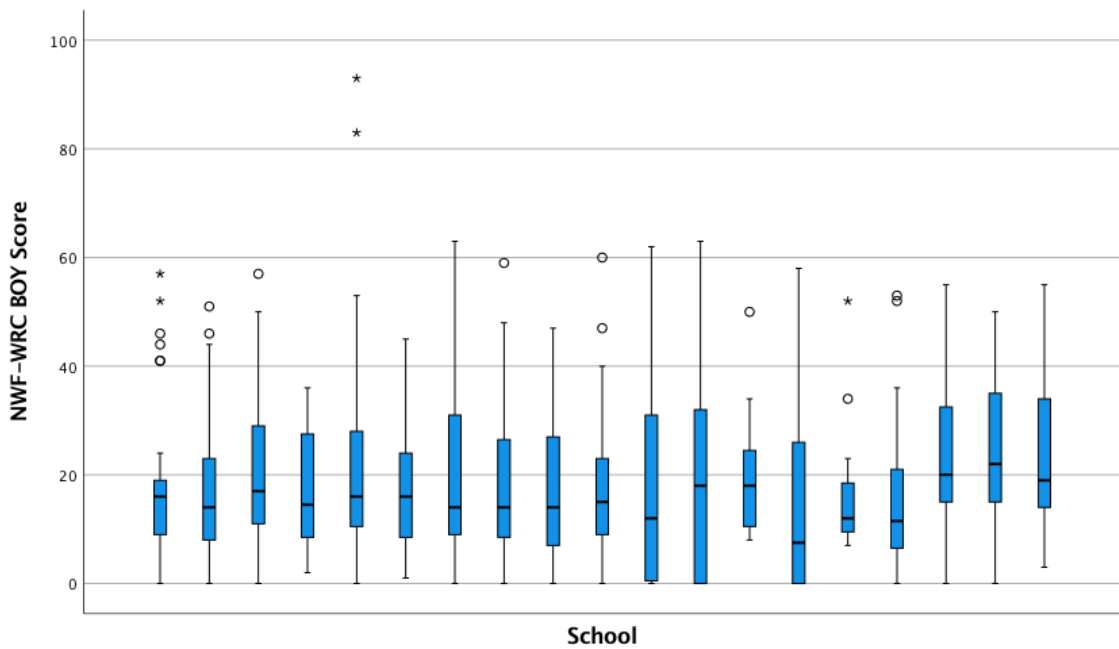
Figure 6

NWF-WRC End of Year Within-School Variability
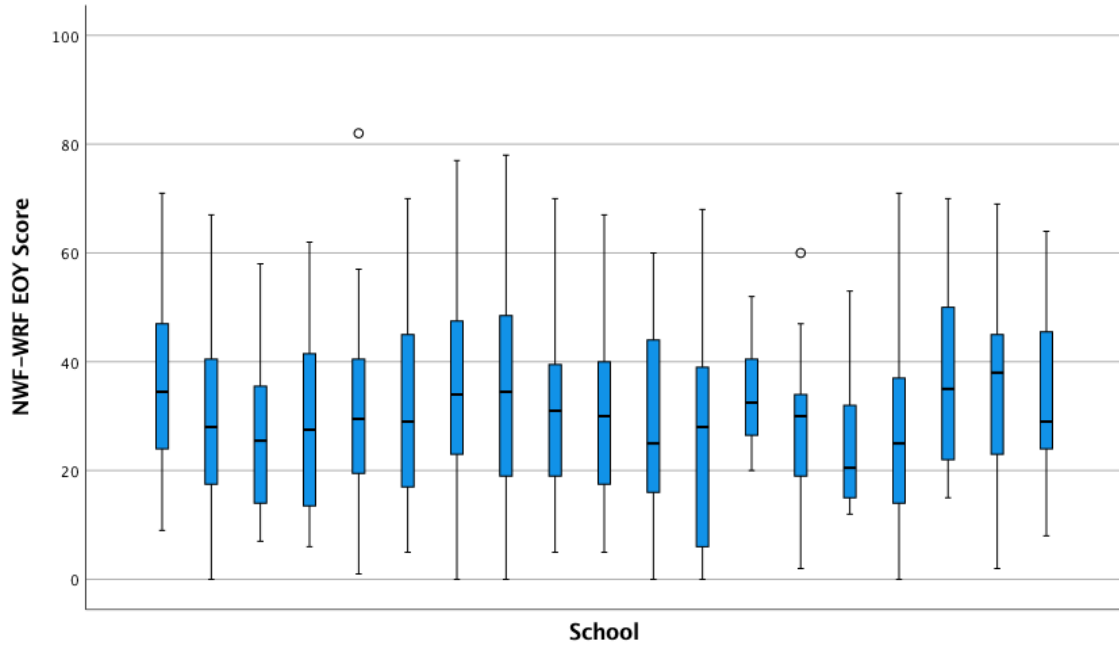
Figure 7

ORF Beginning of Year Within-School Variability
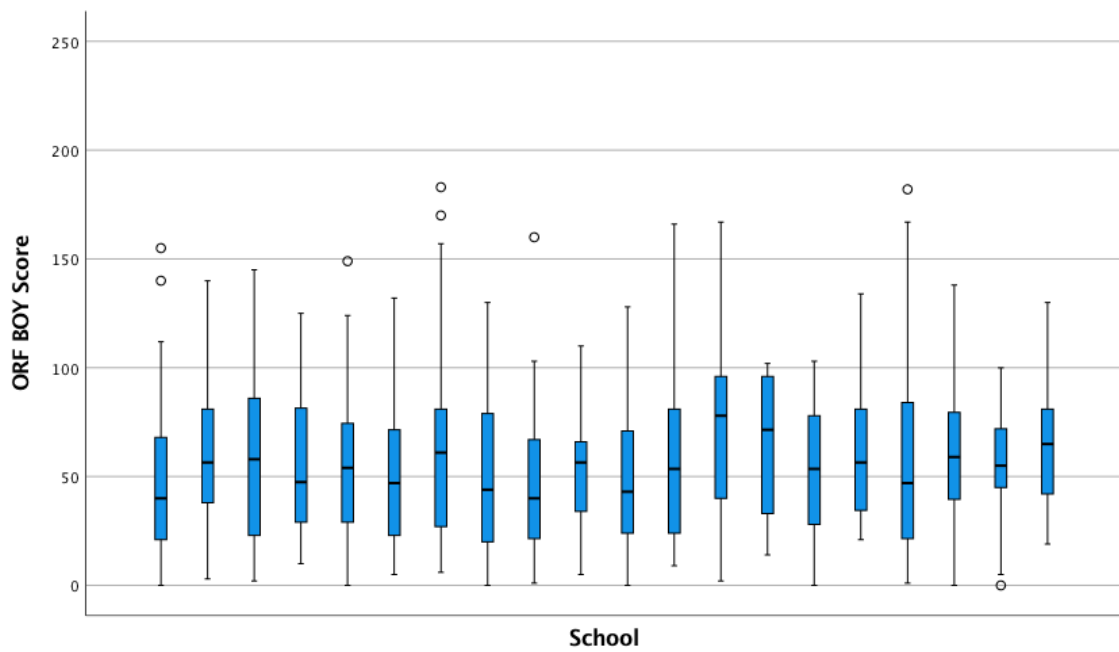
Figure 8

ORF End of Year Within-School Variability
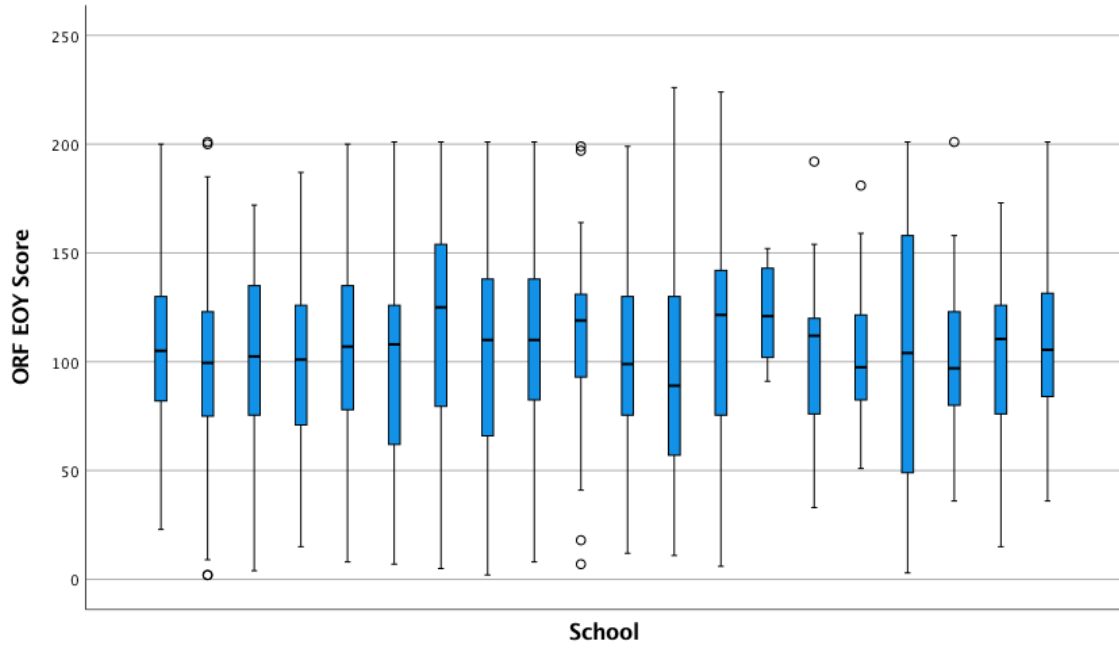
Figure 9

WRF Beginning of Year Within-School Variability
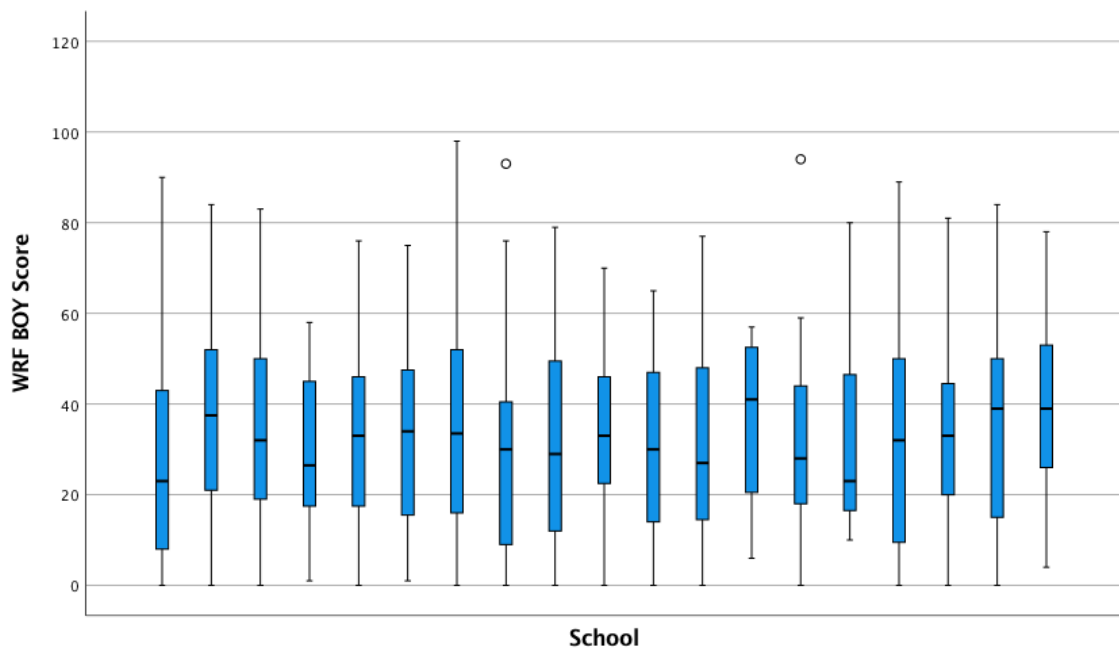
Figure 10

WRF End of Year Within-School Variability
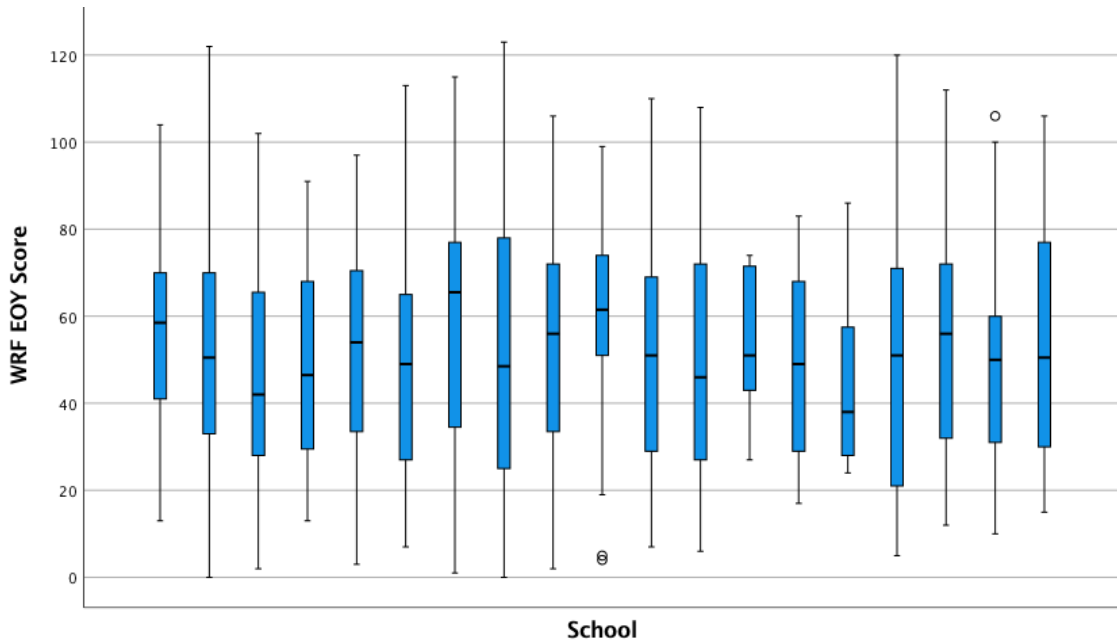
Missing Data

Two-way scatterplots showing schools' percentage of missing data and each of the two independent variables were analyzed to identify potential differential patterns in missing data. School percentage of students tested at both Fall and Spring time periods for each applicable DIBELS-8 measure was plotted against their concentration of historically underserved students or concentration of poverty. A linear regression line with r^2 values were added for each scatterplot (see Appendix D). Scatterplots indicated slightly higher levels of missing data in higher-than-average proportion of FARMS schools, but these relations were quite weak (all $r^2 \leq .025$). No such pattern for schools serving higher proportions of underserved students (all $r^2 < .001$).

Hierarchical Multiple Regression Results

The purpose of this study is to analyze DIBELS Benchmark data for evidence of
the Matthew Effect for schools with above average proportions of racially and
economically historically disadvantaged students. Several hierarchical multiple
regression models were run to investigate whether the concentration of school poverty
(FARMS) and concentration of historically underrepresented students predict beginning
of year scores and beginning to end of year change in scores for schools in the sample.
Multiple regression allows us to determine the proportion of the variation in our
dependent variables (DIBELS Benchmark data) that can be explained by the independent
variables (concentration of school poverty and concentration of students from
historically underserved groups).

The initial sample included all DIBELS Data System users who completed
benchmarking assessments for second-grade students during the 2018-2019 school
year. Sample schools included those who utilized DIBELS 8 as a benchmarking tool
during the 2018-2019 school year for Nonsense Word Fluency (NWF-CLS and NWF-
WRC), Oral Reading Fluency (ORF), and Word Reading Fluency (WRF). A total of 73
DIBELS Data System users completed both Fall and Spring benchmarking data for
second-grade students during the 2018-2019 school year. Twenty-six users were
excluded for being private schools, non-profit organizations, or community centers.
Three users were on different campuses of the same school, and their scores were
merged together. Only public schools that used universal screeners, meaning they
assessed 85% or more of their second-grade students, were included in the analytic

sample for the analyses. Of these schools, none were ineligible for Title I programs, and

only three were eligible for Title I School programs, while the rest of the schools in the

sample were eligible for Title I Schoolwide programs. Due to the limitations in the

variability of Title I School Status as a distinct variable for analysis, further analyses

centered on school proportion of Free and Reduced-Price Lunch (FARMS) as an indicator

of School Concentration of Poverty. Two schools did not report FARMS data to NCES and

were therefore excluded from further analysis.

The final analytical sample included 20 total schools; 19 of those schools

completed NWF benchmarking, 20 completed ORF benchmarking, and 19 completed

WRF benchmarking. The proportion of historically underrepresented students

specifically reflects enrolled second-grade students during the 2018-2019 school year, as

reported by the Elementary and Secondary Information System (ELSI) through the

National Center for Education Statistics (NCES) and was mean-centered for the sample.

The regression procedure for each multiple regression will result in a coefficient

of determination, R^2 is a measure of the variance in the dependent variable explained by

the independent variable, over and above the mean model. Statistically significance of

the model will also be reported, with $p < 0.5$ considered a statistically significant effect

size. Effect sizes of $p < .01$, $p < .05$, and $p < .10$ are reported. Coefficients are interpreted

through the regression equation including the intercept and slope coefficients. Schools'

proportions of historically underserved students and proportion of poverty are

continuous independent variables; the b value of the unstandardized coefficient in the

model represents the change in the dependent variable (DIBELS Score) for one-unit change in the independent variable.

Due to the small sample size and resulting core statistical power, the null hypothesis rejection rule was relaxed and results significant at the p <.10 are reported. Out of the 64 models run, four mean models were statistically significant using a relaxed null hypothesis rejection rule of $p < .10$. The relevant DVs were mean NWF-WRC and WRF at the beginning of the year and median NWF-WRC change.

RQ1: Do schools differ in second-grade reading CBM scores in the fall on the basis of proportion of students living in poverty and proportion of historically underserved students?

Two hierarchical multiple regressions investigated whether proportion of students living in poverty and proportion of historically underrepresented students predict beginning of year median and mean scores for NWF-CLS, NWF-WRC, ORF, and WRF. Results indicated that the proportion of students living in poverty and proportion of historically underrepresented students were overall not significant predictors of beginning of year median and mean scores across applicable DIBELS 8 measures at the $p < .05$ level. Therefore, the null hypothesis could not be rejected. Adding interaction terms did not significantly increase the predictive value of the model at the $p < .05$ level. The results of these analyses for school median and mean scores can be found in Table 9 and 10.

Mean beginning of year DIBELS Scores showed three statistically significant results at the $p < .10$ level. None occurred for median beginning of year scores. For NWF-WRC, Model A was significant, $F(2, 16) = 3.096$, $p = .073$, $R^2 = .279$., and the slope was also significant for FARMS ($B = -8.631$, $t = -2.155$, $p = .407$). Model B was also significant for NWF-WRC, $F(3, 15) = 2.992$, p = .064, $R^2 = .374$. Two slopes were significant proportion of historically underrepresented students ($B = 5.540$, $t = 2.254$, $p = .040$) and FARMS ($B = -7.592$, $t = -1.940$, $p = .071$). For WRF, Model B overall was significant with $F(3, 15) = 2.587$, p = .092, $R^2 = .341$ but without significant predictors. No other models were significant at the $p < .10$ level.

RQ2: Do schools differ in second-grade reading CBM fall to spring change on the basis of proportion of students living in poverty and proportion of historically underserved students?

The second set of hierarchical multiple regression models investigated whether proportion of students living in poverty and proportion of historically underrepresented students predict beginning of year to end of year change in median and mean scores for NWF-CLS, NWF-WRC, ORF, and WRF. Results indicated that the proportion of students living in poverty and proportion of historically underrepresented students were overall not significant predictors of median and mean beginning to end of year change for NWF-CLS, NWF-WRC, ORF, and WRF at the $p < .05$ level. Therefore, the null hypothesis could not be rejected. Adding interaction terms did not significantly increase the predictive

value of the model at the $p < .05$ level. The results of these analyses for school median and mean scores can be found in Tables 9 and 10.

For median fall to spring change in DIBELS scores, NWF-WRC showed statistical significance at the $p < .10$ level. Model A showed statistical significance, $F(2, 16) = 3.012$, $p = .078$, $R^2 = .319$. One slope was statistically significant, FARMS ($B = 7.128$, $t = 1.780$, $p = .094$). No other models were significant at the $p < .10$ level.

Table 10

Median NWF-CLS and NWF-WRC Scores Predicted by Hierarchical Multiple Regression Models using Free and Reduced-Price Lunch and Proportion Historically Underrepresented Students

	NWF-CLS					NWF-WRC				
	R^2	Adjusted R^2	Δ R^2	b	*SE* b	R^2	Adjusted R^2	Δ R^2	b	*SE* b
Restricted Sample			$n = 19$					$n = 19$		
BOY Model A	.045	-.074	.045			.192	.091	.192		
Intercept				54.218**	3.093				15.833**	.664
Underrepresented				4.216	10.854				3.650	2.328
FARMS				-17.820	21.221				-6.515	4.552
BOY Model B	.178	.014	.133			.233	.079	.040		
Intercept				53.440**	3.006				15.933**	.677
Underrepresented				-7.856	12.971				5.201[†]	2.923
FARMS				-23.467	20.655				-5.790	4.654
FARMS * Underrepresented				134.829	86.564				-17.326	19.506
Change Model A	.103	-.009	.103			.274[†]	.183[†]	.274		
Intercept				34.302**	3.378				11.246**	.584
Underrepresented				14.163[†]	11.851				2.732	2.049
FARMS				9.461	23.171				7.128[†]	4.005
Change Model B	.269	.123	.166			.319	.183	.046		
Intercept				35.281**	3.193				11.345**	.592
Underrepresented				29.361	13.780				4.261	2.555
FARMS				16.569	21.943				7.843[†]	4.068
FARMS * Underrepresented				-169.745[†]	91.964				-17.076	17.049

Note. FARMS indicates mean-centered proportion of students eligible for free and-reduced price lunch. Underrepresented indicates mean-centered proportion of students from historically underrepresented student populations. BOY indicates beginning of year DIBELS scores. Change indicates beginning to end change in DIBELS scores. No R^2 or Δ R^2 were significant at the $p < .05$ level. ** $p < .01$. * $p < .05$. [†] $p < .10$.

Table 11

Median ORF and WRF Scores Predicted by Hierarchical Multiple Regression Models using Free and Reduced-Price Lunch and Proportion Historically Underrepresented Students

	ORF					WRF				
	R^2	Adjusted R^2	ΔR^2	b	*SE* b	R^2	Adjusted R^2	ΔR^2	b	*SE* b
Restricted Sample		*n* = 20					*n* = 19			
BOY Model A	.080	-.028	.080			.160	.055	.160		
Intercept				55.941**	2.167				32.310**	1.121
Underrepresented				8.802	7.740				6.841	3.918
FARMS				-9.880	15.223				-2.494	7.722
BOY Model B	.223	.078	.143			.229	.075	.069		
Intercept				55.195**	2.098				32.132**	1.120
Underrepresented				-.238	9.024				3.327	4.925
FARMS				-13.834	14.601				-4.375	7.813
FARMS * Underrepresented				104.497	60.829				38.276	33.080
Change Model A	.086	-.022	.086			.012	-.111	.012		
Intercept				48.539**	1.986				16.842**	1.076
Underrepresented				-7.195	7.091				-1.542	3.762
FARMS				13.131	13.947				1.740	7.415
Change Model B	.164	.007	.078			.032	-.162	.020		
Intercept				48.031**	2.000				16.757**	1.111
Underrepresented				-13.342	8.603				-3.215	4.886
FARMS				10.442	13.920				.844	7.750
FARMS * Underrepresented				71.058	57.991				18.224	32.816

Note. FARMS indicates mean-centered proportion of students eligible for free and-reduced price lunch. Underrepresented indicates mean-centered proportion of students from historically underrepresented student populations. BOY indicates beginning of year DIBELS scores. Change indicates beginning to end change in DIBELS scores. No R^2 or ΔR^2 were significant at the $p < .05$ level. ** $p < .01$. * $p < .05$. † $p < .10$.

Table 12

Mean NWF-CLS and NWF-WRC Scores Predicted by Hierarchical Multiple Regression Models using Free and Reduced-Price Lunch and Proportion Historically Underrepresented Students

	NWF-CLS					NWF-WRC				
	R^2	Adjusted R^2	ΔR^2	b	*SE* b	R^2	Adjusted R^2	ΔR^2	b	*SE* b
Restricted Sample		*n* = 19					*n* = 19			
BOY Model A	.032	-.088	.032			.279[†]	.189	.279		
Intercept				64.511**	3.301				19.379**	.584
Underrepresented				-.075	11.582				3.319	2.049
FARMS				-16.289	22.643				-8.631*	4.006
BOY Model B	.118	-.058	.086			.374[†]	.249	.095		
Intercept				63.850**	3.301				19.522**	.570
Underrepresented				-10.341	14.242				5.540*	2.458
FARMS				-21.091	22.680				-7.592[†]	3.914
FARMS * Underrepresented				114.667	95.051				-24.807	16.405
Change Model A	.104	-.088	.104			.235	.139	.235		
Intercept				37.685**	3.630				11.685**	.634
Underrepresented				13.629	12.739				1.399	2.226
FARMS				15.766	24.905				8.592[†]	4.351
Change Model B	.246	.095	.142			.265	.118	.030		
Intercept				38.658**	3.488				11.770**	.651
Underrepresented				28.750[†]	15.050				2.719	2.809
FARMS				22.839	23.966				9.210[†]	4.474
FARMS * Underrepresented				-168.901	100.442				-14.738	18.749

Note. FARMS indicates mean-centered proportion of students eligible for free and-reduced price lunch. Underrepresented indicates mean-centered proportion of students from historically underrepresented student populations. BOY indicates beginning of year DIBELS scores. Change indicates beginning to end change in DIBELS scores. No R^2 or ΔR^2 were significant at the $p < .05$ level. ** $p < .01$. * $p < .05$. [†] $p < .10$.

Table 13

Mean ORF and WRF Scores Predicted by Hierarchical Multiple Regression Models using Free and Reduced-Price Lunch and Proportion Historically Underrepresented Students

	ORF					WRF				
	R^2	Adjusted R^2	ΔR^2	b	*SE* b	R^2	Adjusted R^2	ΔR^2	b	*SE* b
Restricted Sample		*n* = 20					*n* = 19			
BOY Model A	.013	-.104	.013			.209	.110	.209		
Intercept				57.545**	1.464				33.691**	.580
Underrepresented				1.206	5.229				4.050[†]	2.029
FARMS				-4.565	10.285				-3.313	3.999
BOY Model B	.114	-.052	.102			.341[†]	.209	.132		
Intercept				57.135**	1.461				33.559**	.552
Underrepresented				-3.757	6.285				1.454	2.429
FARMS				-6.736	10.169				-4.703	3.854
FARMS * Underrepresented				57.371	42.364				28.274	16.319
Change Model A	.031	-.083	.031			.055	-.063	.055		
Intercept				47.371**	1.775				18.011**	.952
Underrepresented				-2.999	6.340				-3.168	3.329
FARMS				8.060	12.470				2.084	6.562
Change Model B	.057	-.120	.026			.059	-.130	.004		
Intercept				47.117**	1.845				17.978**	.991
Underrepresented				-6.076	7.936				-3.821	4.359
FARMS				6.714	12.840				1.735	6.915
FARMS * Underrepresented				35.572	53.494				7.108	29.280

Note. FARMS indicates mean-centered proportion of students eligible for free and-reduced price lunch. Underrepresented indicates mean-centered proportion of students from historically underrepresented student populations. BOY indicates beginning of year DIBELS scores. Change indicates beginning to end change in DIBELS scores. No R^2 or ΔR^2 were significant at the $p < .05$ level. ** $p < .01$. * $p < .05$. [†] $p < .10$.

Assessing for Assumptions of Multiple Regression

Assumptions of multiple regressions were assessed post analysis. The assumption of independence of residuals, was checked using the Durbin-Watson statistic. The Durbin-Watson statistic can range from 0.0 – 4.0, with values of 2.0 indicating that there is no autocorrelation detected. For the hierarchical multiple regressions, there was independence of residuals, as assessed by the Durbin-Watson statistics which ranged from 0.912 to 2.891.

The assumption of a linear relationship between the dependent variable (DIBELS scores) and each of the independent variables was tested using a scatterplot of the studentized residuals against the unstandardized predicted values to test for a collective linear relationship. See Appendix A for scatterplots for beginning of year and beginning to end of year change DIBELS mean and median NWF-CLS, NWF-WRC, ORF, and WRF scores. Analysis of the scatterplots showed the assumption of linearity was not violated. The assumption of homoscedasticity of residuals states that the residuals of all values of the predicted dependent variable are equal and can also be confirmed through plotting the studentized residuals against the unstandardized predicted values. If data shows homoscedasticity, the spread of the residuals will be constant rather than exhibiting a pattern. Analysis of the scatterplots showed the assumption of homoscedasticity was not violated.

Multicollinearity was assessed through an inspection of correlation coefficients and Tolerance/VIF values. Multicollinearity occurs when a multiple regression's independent variables, this this case concentration of poverty and concentration of

historically underrepresented groups, are correlated with one another. When multi-collinearity occurs, it becomes difficult to statistically understand how each independent variable contributes to understanding the variance explained by the dependent variables, in this case the DIBELS scores. Tolerance values less than 0.1 or correlation coefficients larger than 0.7 indicate potential collinearity issues. An inspection of correlation coefficients showed no correlation coefficients larger than 0.7 between the independent variables. Correlation coefficients ranged from 0.17 to 0.20. Inspection of tolerance values showed no tolerance values less than 0.1. Tolerance values ranged from 0.602 to 0.966. Both inspection of correlation coefficients and tolerance values showed the assumption non-multicollinearity was not violated.

Outliers were assessed using casewise diagnostics and studentized deleted residuals to identify residuals ±3 standard deviations. To determine whether any cases exhibited high leverage, ideally leverage values should be less than 0.2 while leverage values above 0.5 are considered high leverage points. Analysis of the leverage values indicated no values were above 0.5 and that values ranged from 0.02 to 0.36. Cook's Distance was used to check for influential points in the models. Analysis of Cook's Distance values indicate there were no values above 1 and values ranged from 0.000002 to 0.97. Finally, normal distribution of the residuals was checked by using histograms with superimposed normal curves and the Probability-Probability (P-P) Plots. See Appendix B for Histograms and Appendix C for P-P Plots. Visual analysis of the histograms and P-P plots indicated the data was adequately normally distributed.

Casewise diagnostics identified a studentized deleted residual of +3.003 for case 58 (WRF BOY Mean Score). A further examination of the studentized deleted residuals (SDR) indicated three potential outliers. Case 58 (SDR = 5.25 for WRF BOY Mean), Case 52 (SDR = 3.58 for NWF-CLS Change Mean and SDR = 4.14 for NWF-CLS Change Median), and Case 36 (SDR = -3.30 for NWF-WRC Begin Mean and SDR= -3.43 for NWF-WRC Begin Median). Due to the presence of three potential outliers, Cases 36, 52, and 58, the hierarchical multiple regressions were excluded for applicable multiple regressions and the regressions were re-run without the potential outliers. Due to the small sample size, excluding these outliers reduced the sample to 17 and 18 schools respectively, further underpowering the analyses. It should be noted that these schools may not represent true outliers and instead could be part of a larger pattern in school DIBELS scores that may be present in an analysis with a larger sample of schools. The assumptions of multiple regression were not violated for these new hierarchical linear regressions.

For NWF-WRC Begin Mean, excluding Cases 52 and 58 resulting in both BOY Model A and Model B being statistically significant at the p < 0.05 level. In BOY Model A for NWF-WRC Mean Scores $F(2,14)=4.33$, $p = .034$, $R^2 = .382$, both proportion of underrepresented students and proportion of students receiving free and reduced-price lunch were additionally significant predictors. BOY Model B was also significant $F(3,13) = 4.09$, $p = .03$, $R^2 = .486$, however proportion of underrepresented students was a significant predictor at the $p < .05$ level. However, for NWF-WRC Begin Median, neither was significant at the $p < .05$ level. For BOY NWF-WRC Median Scores Model A was significant at the $p < .10$ level with $F(2, 14) = 3.385$, $p = .06$, $R^2 = .326$ with proportion of

63

students receiving free and reduced-price lunch being a significant predictor. Model B

was not significant for median scores. Results of the follow-up regression for NWF-CLS

Beginning of Year Scores excluding cases 52 and 58 can be seen in Tables 11 and 12 .

     After excluding Case 58 from the analysis of WRF BOY Mean, both BOY Model A

and BOY Model B were statistically significant. For BOY Model A, $F(2,15)= 5.034$, $p=.02$,

$R^2 = .402$ with proportion of underrepresented students adding predictive value to the

model. For BOY Model B, $F(3, 14) = 9.573$, $p=.001$, $R^2 = .672$ with proportion of students

receiving free and reduced price lunch and the interaction term adding predictive value

to the model. This interaction term is the only statistically significant interaction term

throughout the study at the $p <.01$ level. Results of the follow-up regression for WRF

Beginning of Year Mean Scores excluding case 58 in comparison to the original multiple

regression results can be seen in Table 13.

Table 14

NWF-WRC Beginning of Year Mean Scores

	Sample Without Outliers n = 17					Analytic Sample n = 19				
	R^2	Adjusted R^2	Δ R^2	b	SE b	R^2	Adjusted R^2	Δ R^2	b	SE b
BOY Model A	.382*	.156	.382*			.279[†]	.189	.279		
Intercept				19.554**	.517				19.379**	.584
Underrepresented				4.537*	1.915				3.319	2.049
FARMS				-7.384*	3.344				-8.631*	4.006
BOY Model B	.486*	.091	.104			.374[†]	.249	.095		
Intercept				19.795**	.512				19.522**	.570
Underrepresented				7.060*	2.392				5.540*	2.458
FARMS				-6.854[†]	3.183				-7.592[†]	3.914
FARMS * Underrepresented				-23.149	14.306				-24.807	16.405

Note. FARMS indicates mean-centered proportion of students eligible for free and reduced price lunch. Underrepresented indicates mean-centered proportion of students from historically underrepresented student populations. All R^2 values indicated adjusted R^2

** p <.01. * p <.05. [†] p <.10.

Table 15

NWF-WRC Beginning of Year Median Scores

	Sample Without Outliers n = 17					Analytic Sample n = 19				
	R^2	Adjusted R^2	Δ R^2	b	SE b	R^2	Adjusted R^2	Δ R^2	b	SE b
BOY Model A	.326[†]	.341	.326[†]			.192	.091	.192		
Intercept				16.005**	.546				15.833**	.664
Underrepresented				4.959*	2.022				3.650	2.328
FARMS				-4.835	3.532				-6.515	4.552
BOY Model B	.351	.298	.025			.233	.079	.040		
Intercept				16.124**	.581				15.933**	.677
Underrepresented				6.208*	2.717				5.201[†]	2.923
FARMS				-4.572	3.616				-5.790	4.654
FARMS * Underrepresented				-11.458	16.252				-17.326	19.506

Note. FARMS indicates mean-centered proportion of students eligible for free and reduced price lunch. Underrepresented indicates mean-centered proportion of students from historically underrepresented student populations. All R^2 values indicated adjusted R^2

** $p <.01$. * $p <.05$. [†] $p <.10$.

Table 16

WRF Beginning of Year Mean Scores

	Sample Without Outliers n = 18					Analytic Sample n = 19				
	R^2	Adjusted R^2	ΔR^2	b	SE b	R^2	Adjusted R^2	ΔR^2	b	SE b
BOY Model A	.402*	.091*	.402*			.209	.110	.209		
Intercept				33.410**	.435				33.691**	.580
Underrepresented				4.341**	1.501				4.050†	2.029
FARMS				-5.396†	3.005				-3.313	3.999
BOY Model B	.672**	.147**	.271**			.341†	.209†	.132		
Intercept				33.234**	.337				33.559**	.552
Underrepresented				1.287	1.459				1.454	2.429
FARMS				-7.187**	2.362				-4.703	3.854
FARMS * Underrepresented				33.482**	9.847				28.274	16.319

Note. FARMS indicates mean-centered proportion of students eligible for free and reduced price lunch. Underrepresented indicates mean-centered proportion of students from historically underrepresented student populations.

** *p* <.01. * *p* <.05. † *p* <.10.

Illustrating the Complex Patterns of Results

Graphs showing estimated values for DIBELS-8 scores as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals were created to illustrate the complex, though largely non-significant associations in the multiple regression models. Graphs for Nonsense Word Fluency-Words Read Correctly Mean Beginning of Year Scores and for Word Reading Fluency Beginning of Year Mean Scores are shown for both the full sample and sample without outliers (see Figures 11-14 respectively) since these were the models that displayed significance after the removal of outliers. Additional graphs for the other models are included in Appendix E.

Each graph shows below average (minus one standard deviation), average, and above average (plus one standard deviation) for FARMS (represented as three points along the x-axis) and for underrepresented students (represented by the three lines).

For mean NWF-CLS scores (Figures 11 and 12) at the beginning of the year, the pattern of differences between the full and limited sample changes. That is, for example, while the full sample estimates no effect of proportion FARMS for schools with above-average proportions of underrepresented students (the dotted line in both figures), the restricted sample demonstrates a negative association for FARMS for these same schools. Similarly, while the full sample results suggests that schools with below-average proportions of underrepresented students (the solid line in both figures), the restricted sample demonstrates no association for FARMS for these same schools. These contrasting results suggest that the NWF-WRC results are too volatile based upon sampling decisions to allow for clear inferences to be drawn.

In contrast, for mean WRF scores at the beginning of the year, the pattern of

results is highly consistent regardless of whether the outlier school is omitted. The

patterns here show that for schools serving below-average proportions of

underrepresented students (the solid line in both graphs), the higher the proportion of

FARMS, the lower the predicted WRF score. The difference between below-average and

above-average FARMS schools serving below-average underrepresented students

amounts to about 4 points in the full sample (see Figure 13) and just over 5 points in the

restricted sample (see Figure 14). This same pattern is similar but less dramatic for

schools serving average proportions of underrepresented students (the dashed line in

both figures). In contrast, for schools serving above-average proportions of

underrepresented students, there is less than 1-point difference between below-

average and above-average FARMS schools, suggesting that concentration of poverty

makes little difference in these schools.

Figure 11.

Estimated mean NWF-WRC at BOY as a function of school proportion of students from

underrepresented groups and receiving free and reduced-price meals (Full Sample)
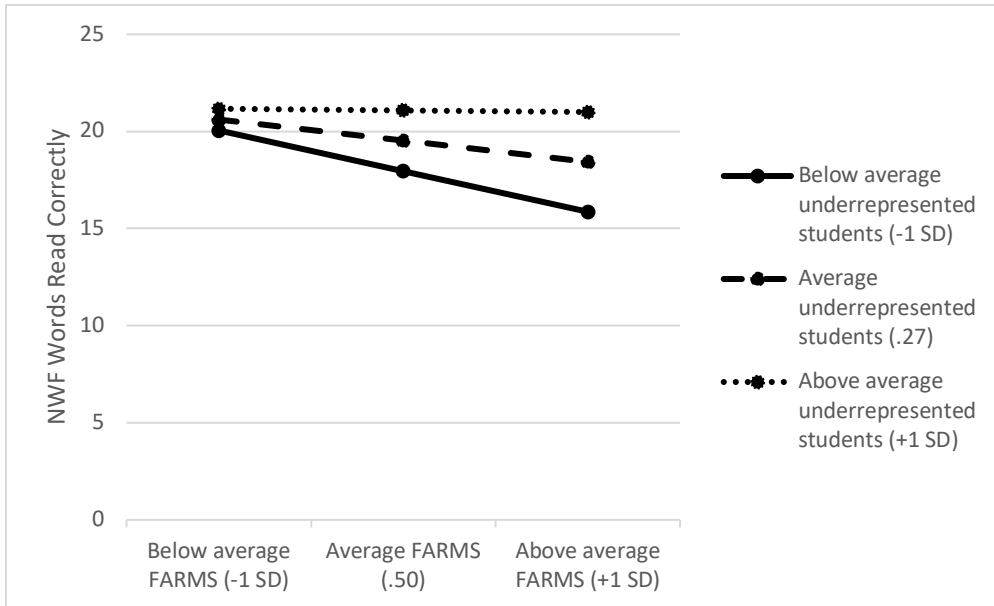
Figure 12.

Estimated mean NWF-WRC at BOY as a function of school proportion of students from

underrepresented groups and receiving free and reduced-price meals (Without Outliers)
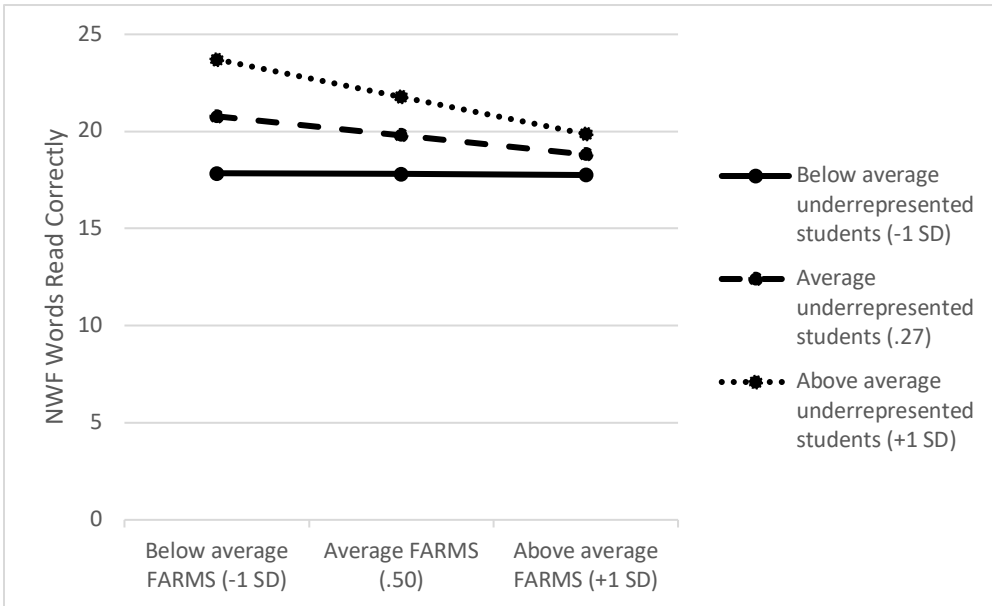
Figure 13.

Estimated mean WRF at BOY as a function of school proportion of students from

underrepresented groups and receiving free and reduced-price meals (Full Sample)
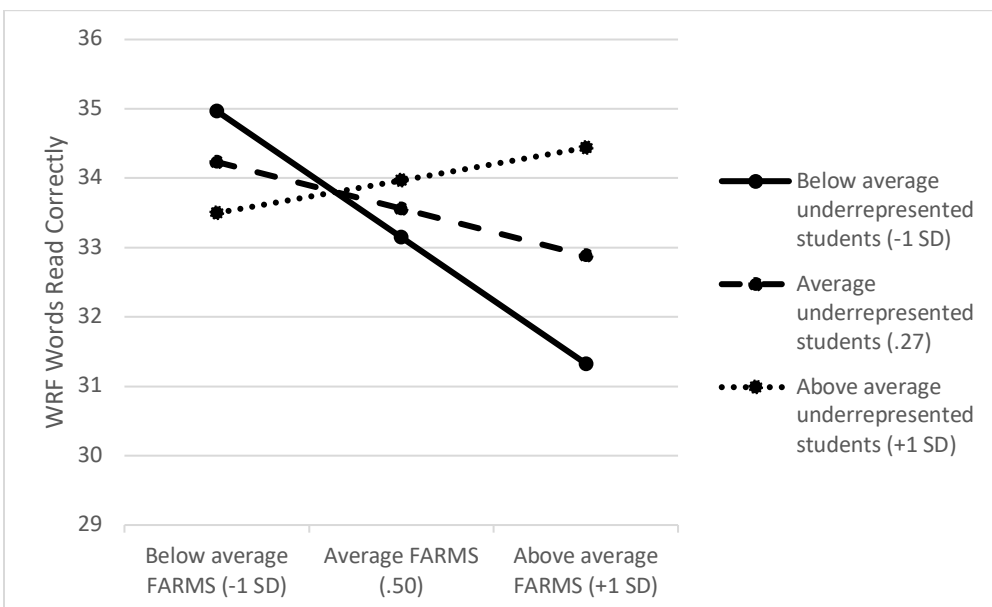
Figure 14.

Estimated mean WRF at BOY as a function of school proportion of students from

underrepresented groups and receiving free and reduced-price meals (Without Outliers)
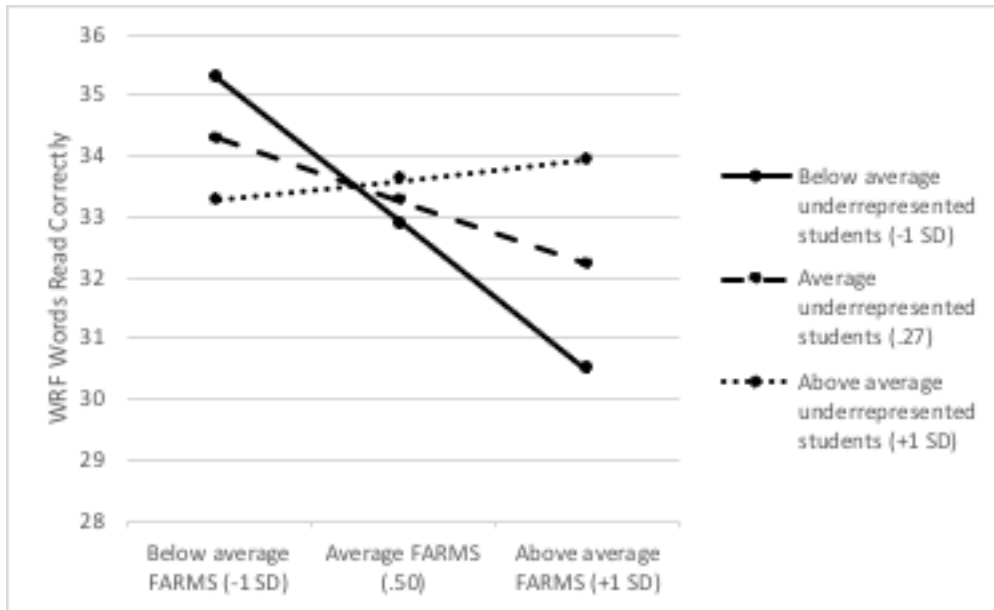
CHAPTER V

DISCUSSION AND CONCLUSION

Summary

Reading difficulties are correlated with school-based academic problems in later grades and general adverse life outcomes such as high school graduation rates and future participation in the economic system (Kame'enui et al., 2000; Fiester, 2013). Furthermore, many students who struggle with reading in the early elementary years do not catch up to their more skilled peers (Butler et al., 1985; Fiester, 2010; Snow et al., 1998). The purpose of this study was to examine DIBELS-8 within-year benchmark data for the presence of Matthew Effects. The Matthew Effect is an educational theory that children who begin their academic careers with lower reading scores are unable to catch up to their peers, instead falling farther and farther behind as their schooling progresses (Francis et al., 1996; Stanovich, 2000; Walberg & Tsai, 1983).

More specifically, this study used a series of hierarchical multiple regressions to analyze mean and median school reading scores for three DIBELS measures: oral reading fluency, word reading fluency, and nonsense word fluency. Although there has been some evidence of Matthew Effects in literacy for students as well as ongoing evidence of achievement / opportunity gaps for students from specific racial and socioeconomic groups, there has been limited research on schools as the unit of analysis (Logan & Petscher, 2010; Morgan et al., 2008). Through looking at the Matthew Effect in the aggregate, on a school comparison level, we can instead analyze the Matthew Effect as a social issue reflecting widespread and historical systemic disadvantages.

The sample included public schools which used the applicable DIBELS-8 measures as universal screening tools for second grade students during the 2018-2019 school year. The analysis of school mean and median scores compared schools on the basis of proportion of historically underrepresented student populations and proportion of students living in poverty in order to systemic inequalities in education. For the conducted analyses, proportion of historically underrepresented student populations and proportion of students living in poverty served as the independent variables. For proportion of historically underrepresented students, the variable represented number of second grade students identified as Black, Hispanic, or Native American / Alaskan Native, meaning those from racial demographic groups identified by NAEP as experiencing achievement / opportunity gaps in literacy. Proportion of students living in poverty represented the schoolwide proportion of students receiving free or reduced-price lunch assistance. School beginning of year mean and median DIBELS-8 scores and beginning of year to end of year change scores served as the dependent variables. Two hierarchical multiple regression models were conducted, with an interaction term in the second model.

Discussion of Key Findings

This study focused on benchmark scores from the 2018-2019 school year, the first year of DIBELS-8 adoption. At the time of this dissertation, schools are currently transitioning from previous version of DIBELS to the 8th edition. Schools represented in this sample were early adopters of DIBELS-8 and may not be representative of the entire population of school engaging in progress monitoring through curriculum-based

measures. Of these early adopters, not all used DIBELS-8 as a universal screening measure; 22 of the 44 public schools assessed fewer than 85% of their students with any DIBELS-8 measures. As schools were in a transition year from older versions of DIBELS, it is not possible to know if schools were using other CBM measures or version as universal screeners, transitioning some classrooms to DIBELS-8 and not others, or if they were using DIBELS-8 with specific populations of students. Interestingly, rural schools were overrepresented as early adopters of DIBELS-8 universal screening with 70% of schools the analytic sample being located in rural areas, not including one geographically remote school which was dropped from the analysis due to only enrolling a single second grade student.

Beginning of Year Scores

In accordance with National Center for Educational Statistics (2019)'s data on achievement gaps in literacy, this study hypothesized schools with above average proportions of students living in poverty or above average proportions of historically underserved groups would have statistically significant lower mean and median reading scores in the fall. Although results are not statistically significant, a pattern emerged across hierarchical multiple regression models for all DIBELS-8 subtests which showed schools serving high proportions of students living in poverty did have lower beginning of year mean and median scores. For example, the first hierarchical multiple regression run showed that schools which served a above average proportion of students who qualified for free and reduced-price lunch had a median score of 17.820 fewer correct letter sounds in NWF benchmarking than schools serving more financially privileged

populations. This pattern was consistent but not statistically significant, likely in part

due to the small sample size. Schools serving above average proportions of students

from historically underrepresented groups had more variable patterns emerge in scores.

Likewise, these differences were not statistically significant but unlike concentration of

poverty, concentration of historically underrepresented students did not indicate clear

patterns in lower reading scores at the beginning of the year. Median and mean

beginning of year scores were generally slightly higher across subtests in the first model

(except for mean NWF-CLS), but this pattern became more variable in the second model

which introduced the interaction term although still not statistically significant.

Beginning of Year to End of Year Change in Scores

This study hypothesized that schools with above average proportions of students

living in poverty or above average proportions of historically underserved groups would

have statistically significant lower CBM fall to spring change in scores in accordance with

prior research conducted by Francis et al. (1996) and Stanovich (2000). Matthew Effects

necessitate statistically significant lower fall scores and lower rates of growth for schools

based on concentration of poverty or proportion of historically underserved students

(Pfost et al., 2014). Although not statistically significant, results were not consistent with

evidence for a Matthew Effect across DIBELS-8 subtests. For schools serving above

average proportions of students living in poverty, results instead indicated potential

evidence for a compensatory model. Schools which had above average proportions of

students eligible for FARMS had positive but not statistically significant effects for Fall to

Winter change in scores across DIBELS-8 subtests. For example, for ORF, schools with

above average rates of FARMS had a mean Fall to Winter change score that was 8.060

words read correctly than schools serving lower than average numbers of FARMS

eligible students. Results for schools serving above average proportions of historically

underserved students were more mixed, with higher Fall to Winter change in NWF-CLS

and NWF-WRC scores but negative outcomes for ORF and WRF.

Matthew Effects versus a Compensatory Model

According to Pfost et al. (2014)'s meta-analysis, a Compensatory Model indicates

originally lower-scoring schools having greater growth over time than their more

privileged comparison schools. This higher rate of growth would allow the closure of the

achievement gap, as the difference in average scores would decrease across the three

time points. Pattern A in Figure 15 shows a Matthew Effect in comparison to Pattern B

showing a Compensatory Model.

Figure 15

Matthew Effects v. Compensatory Model (Pfost, Hattie, Dorfler, and Artelt, 2014, pg.

206).
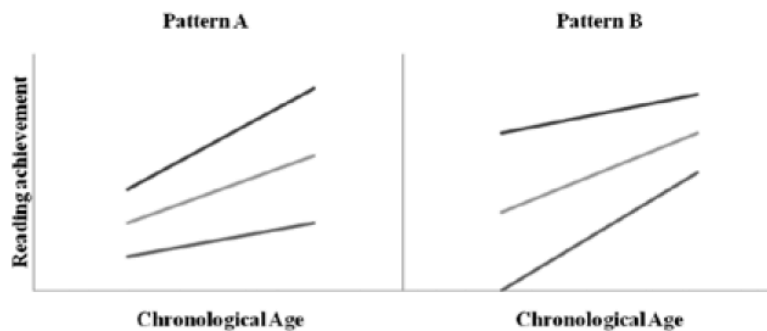
This study hypothesized that Matthew Effects would be found in the analysis of

school mean and median literacy scores which would have necessitated an increasing

gap between schools that serve more and less privileged populations. For Matthew Effects to be present, schools with above average proportions of students from underserved communities or schools with above average proportions of students living in poverty would have had statistically significant lower beginning of year reading scores and statistically significant lower change scores.

Schools which served above average proportions of students receiving free or reduced-price lunch additionally had higher but not statistically significant Fall to Spring change in scores than their lower-FARMS counterparts. Schools in this study which served above average proportions of students living in poverty began the year with lower average reading scores but had higher positive change in scores from Fall to Spring. This finding is in line with Reardon et al. (2019)'s findings that lower prior test scores on the school level did not lead to widening gaps within-district achievement gaps even though they found evidence for persistent poverty related achievement gaps. Additionally, across models, the interaction terms between poverty and underrepresented students were non-significant.

This indicates that on the basis of poverty, scores followed an apparent (although not significant) compensatory pattern rather than representing widening gaps. Most importantly, a compensatory model would indicate gaps in literacy scores are not widening over the course of a single school year, but rather, especially for students living in poverty, the gaps are in the process of closing. However, national reading data from NAEP (2019) show achievement gaps persist for both students living in poverty and students from historically underserved backgrounds.

Previous studies such as Logan & Petscher (2010), Morgan et al. (2008), and Rumberger & Palardy (2005) have given evidence for Matthew Effects related to demographic variables; however, the results of the present study did not replicate these prior finding. Morgan et al. (2008) and Rumberger & Palardy (2005) each found evidence for the Matthew Effect using Hierarchical Linear Modelling. Morgan et al., (2008)'s analysis of demographic risk for 10,587 students who entered kindergarten during the 1998 school year found students discrepancies between more and less privileged students grew between kindergarten and third grade within their sample. Similarly, Rumberger & Palardy (2005) found that both students' own socio-economic status, and the school average socio-economic status contributed to student achievement. This study focused not on student's demographic factors but solely using between-schools focus on school achievement, representing a large methodological shift. Unlike Morgan et al. (2008) and Rumberger & Palardy (2005), this study did not find academic achievement was strongly affected by school-level demographics, at least when analyzing that achievement at the school-level.

The 2019 NAEP 4th grade reading data shows evidence of achievement or opportunity gaps for Black, Hispanic, and American Indian / Alaska Native students as well as those living in poverty. NAEP Scores indicate these students have statistically significantly lower reading scores than their more privileged peers (National Center for Educational Statistics, 2019). One implication for practice of the present study is that although these gaps are present on the individual student level, these gaps become obscured at the very least or are not statistically present when analyzing school scores.

That is, this study shows that schools which serve an above average proportion of these

students do not have statistically significant lower beginning of year or beginning to end

of year change over schools which serve smaller proportions of these students.

However, this indicates schools that serve above average proportions of students living

in poverty or from minority groups are not worse off than other schools by their nature.

However, this does not mean the education system supports disadvantaged students in

closing individual-level achievement and opportunity gaps. Educational questions in

everyday practice regarding judgement of school functioning perhaps need to be

disaggregated based on student risk factors rather than aggregated on the school-level.

Further research into these achievement gaps and how they progress over time should

be instead conducted on the individual level or using an HLM statistical methodology.

The study's small sample size makes it harder to draw conclusions since the

study is under powered, but the hierarchical multiple regressions nevertheless show

interesting patterns. When analyzing the regression results, schools with high

proportions of students receiving free and reduced-price lunch consistently had lower

beginning of year scores across measures. These results were not statistically significant

at the $p < .05$ level, but they do show a non-significant indication that within the small

sample, the schools which served an above average population of students eligible for

FARMS population began the school year with consistently lower reading scores than

those which served a more financially privileged (lower than average proportion eligible

for FARMS) population. This pattern was additionally apparent in further analysis of

NWF-CLS when the regressions were rerun to exclude schools which were potential

outliers. Upon exclusion of outliers, the results were statistically significant at the $p < .05$ level.

If similar non-significant results were shown in a study with an expanded sample with analyses that were not underpowered, it would show evidence that Matthew Effects are not observable at the school level. These findings would potentially be in contrast to Logan and Petscher's 2009 analysis of Florida reading scores. They analyzed CBM reading scores for 175,857 students using a similar school-level of analysis. Their findings indicated CBM growth differentiated on the school-level based on demographic risk, with gaps between schools widening over time. Yet this study found that although non-significant gaps appear at the beginning of the year, these gaps close over the course of the school year.

This study also relied on a method of within-year progress monitoring data while many previous studies were longitudinal in nature. Morgan et al., (2008) and Northrop (2017) found evidence for Matthew Effects using data from the Early Childhood Longitudinal Study (ECLS) from the Institute of Educational Sciences on the basis of risk variables such as demographic risk or prior reading scores respectively. It is possible if this study was expanded over the course of several years, we would begin to see similar effects within the dataset and that Matthew Effects are more cumulative than one year of data can portray.

Limitations

Individual student demographic information was not available for the sample, preventing student centered analyses of race or poverty factors. As student specific

demographic information was not available in the dataset, it is unknown how these factors may have influenced each student's scores over the course of the year. Additionally, the school sample size for this study was heavily limited due to several factors. First, the use of DIBELS-8 scores meant that only early adopters of the new edition could be included in the study. In the end, only 20 schools were included in the sample, which meant the analyses were underpowered. Casewise diagnostics and examination of the studentized residuals identified three schools which were potential outliers in the analyses. With the small sample size, any potential outliers would have an outsized effect on data analysis. However, it is difficult to know if these schools were true outliers or represented larger trends in the data that would have become apparent with a larger sample size. Removal of these outliers and rerunning of the affected multiple regressions led to statistically significant results in most of the models, further underscoring the need for a larger sample. As adoption of DIBELS-8 continues, this limitation will be naturally alleviated for future research with the DDS.

Directions for Future Research

Almost all early adopters of DIBELS-8, who used it as a universal screening tool, were also schools which qualified for Title I Schoolwide program funding from the federal government. In fact, there were no schools in the analytic sample which were ineligible for Title I funding and only two who were eligible for the more limited Title I School funding rather than Title I Schoolwide funding. Title I Schoolwide eligible schools are those whose population includes a minimum of 40% of students living in poverty or

schools which receive a federal waiver (ESSA, 2016). This led to necessitating the use of FARMS as a measure of poverty in the multiple regression analyses.

Title I programing is a potentially interesting variable for future analysis due to its historically close ties to literacy intervention programs. Although Title I funds can be used for a variety of academic programming, they have historically been aligned with math and reading (ESSA, 2016). As schools have large leeway to use their Title I funding, an interesting area of future research could be analysis which includes both Title I and FARMS, especially if such analysis could differentiate between how schools spend their Title I funding dollars. As some schools may spend their funding on reading intervention programs, versus other activities such as summer enrichment or professional development trainings for teachers, it is possible that differences exist in school reading scores depending not only on proportion of students living in poverty but those students' access to specialized federal funding aimed at alleviating some of the negative academic effects of that poverty.

However, further complicating the study of effects of poverty concentration regards very recent changes to federal funding of school lunches. On April 20th, 2021 the U.S. Department of Agriculture announced continued extended funding for school lunches due to the COVID-19 pandemic. According to the U.S. Department of Agriculture's Press Release:

> "Schools nationwide will be allowed to serve meals through USDA's National School Lunch Program Seamless Summer Option (SSO), which is typically only available during the summer months… schools that choose this option will receive higher-than-normal meal reimbursements for every meal they serve" (U.S. Department of Agriculture, 2021)

Due to the Covid-19 pandemic, many more children are currently qualifying and receiving free meals from their education settings and the Federal government is actively promoting "strategies to increase student and family access to meal programs during the school year and over the summer, including specific strategies for underserved students" (Department of Agriculture, 2021). It is currently unknown if changes to the school lunch programs will affect how schools track the number of students living in poverty and therefore the use of FARMS as an indicator of poverty within education research.

As DIBELS-8 has been implemented further, the sample size has grown drastically to 2,550 registered users during the 2020-2021 school year including schools, daycares, and community centers. While we cannot assume all of these DIBELS-8 users completed universal screening or are public schools, it is likely more schools would be available for analysis. As the analyses in this study were unpowered, with a larger sample size, it is possible results would be statistically significant.

An optimal study would look at student benchmark scores during the academic year and the summer period. Due to sampling, this study was limited to benchmark scores within a single school year. The data showed some trends in favor of a compensatory model. The non-significant trends indicated schools with above average proportions of students living in poverty were able to make progress towards catching their students up to more privileged peers within a single school year. However, longitudinal data from prior research shows long-term negative outcomes to low reading scores and that these gaps do not close (Francis et al., 1996; Stanovich, 2000).

Research into the summer slide presents a possible explanation for this phenomenon,

that although students make progress towards closing reading gaps in the schoolyear,

this progress is undone over the summer (Cleano and Neuman, 2009; Alexander et al.,

2007). Alexander et al., (2007) used data from the Baltimore Beginning School Study and

found that students cumulative achievement gains generally reflected within school-

year gains but achievement gaps between higher and lower SES status students

reflected differential summer learning loss. The authors concluded that evidence in

Baltimore of achievement gaps was rather due to lower SES students losing more skills

over the summer than their affluent peers who had access to higher quality summer

programs. Their study centered on students within one large urban area, the DDS has

access to a national sample including students in both urban and rural areas. The

addition of further data from the following Fall would with a larger sample would be

able to show if summer loss in reading scores is contributing to the persistence of

achievement gaps for schools in the DDS.

Potentially the most important avenue for future research into Matthew Effects

in literacy regards the recent COVID-19 public health crisis. In March 2020, most schools

in the United States transitioned to virtual distance learning, with constantly changing

and complex systems of virtual learning, hybrid learning, and traditional but socially

distanced in-person learning. A year into the COVID-19 pandemic, recent research has

focused on questions of learning loss and differing rates of student growth due to both

the educational transition to at-home versus in-person learning but also potentially

lingering effects of pandemic related trauma (Fontenelle-Tereshchuk, 2021; Garcia and

Weiss, 2020). As the COVID-19 pandemic has not yet ended and all students have not transitioned back to traditional in-person learning, full impacts of the pandemic are not yet known. Preliminary research on ORF data has shown Fall 2020 gains in ORF scores but that these gains have not recouped lack of ORF growth in Spring 2020. Furthermore, Domingue et al., (2021)'s preliminary research has shown inequitable impact, as students attending schools with historically lower standardized test scores have been more adversely impacted in comparison to more privileged peers. Although it is likely that further research will show differing effects of the COVID-19 pandemic and potentially exacerbate achievement gaps between students based on demographic risk factors and backgrounds. Thus, with the impact of the COVID-19 pandemic, it is likely we would see increased differences in beginning of year DIBELS-8 scores for schools serving above average concentrations of students living in poverty.

SCATTERPLOT OF REGRESSION PREDICTED VALUES FOR ALL DEPENDENT VARIABLES

Figure 16

Scatterplot of Regression Predicted Value for NWF CLS BOY Mean Score
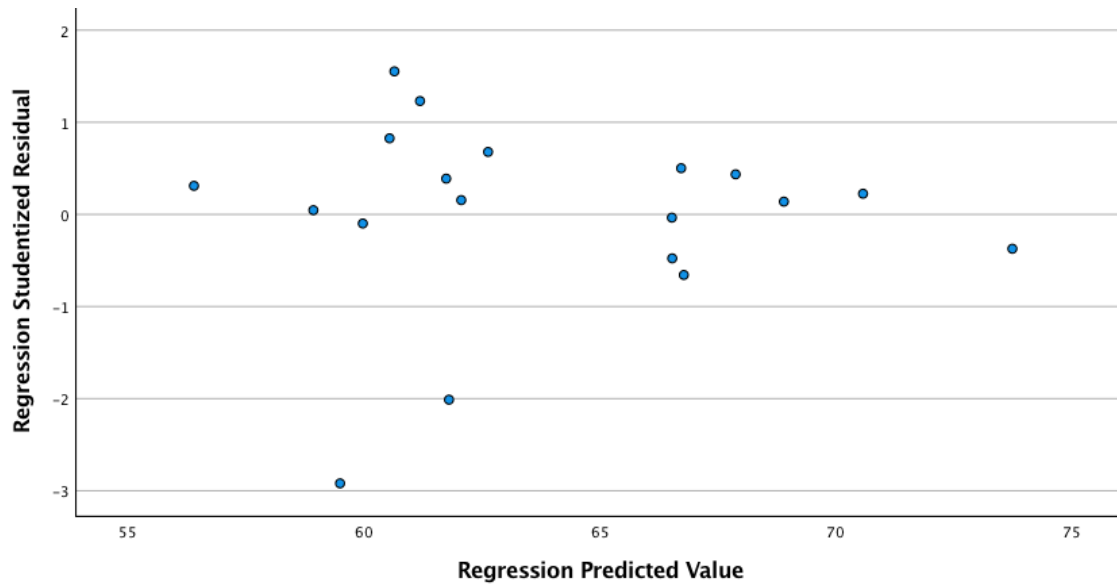
Figure 17

Scatterplot of Regression Predicted Value for NWF-CLS BOY Median Score
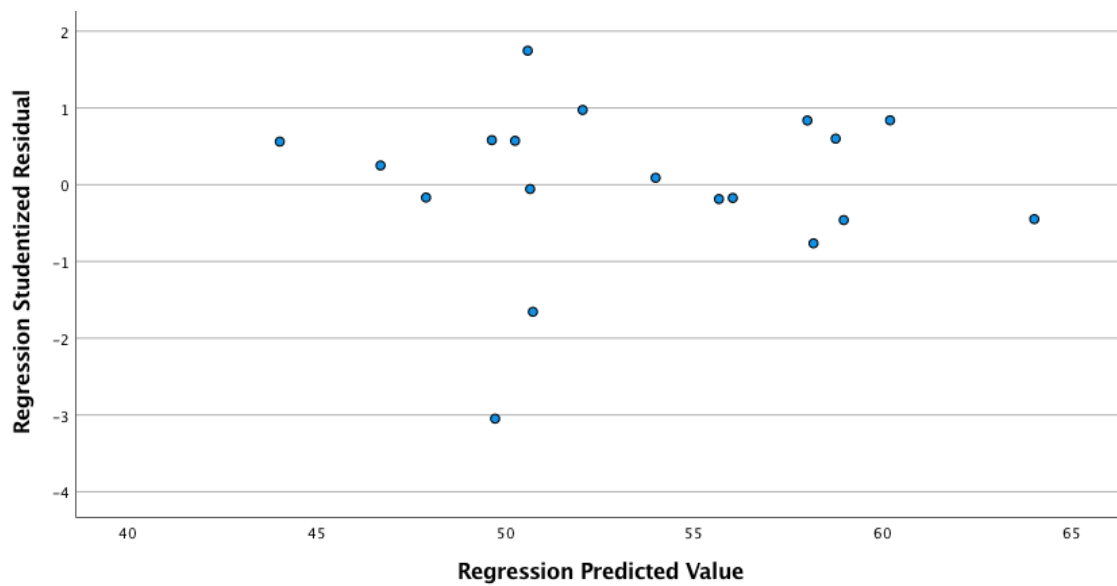
Figure 18

Scatterplot of Regression Predicted Value for NWF CLS Change Mean Score
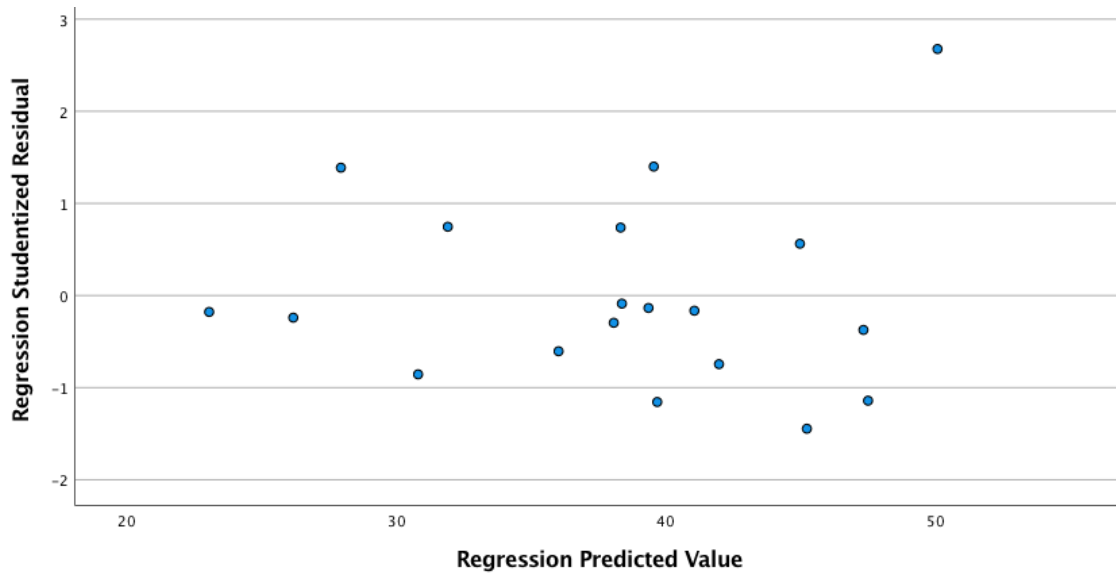
Figure 19

Scatterplot of Regression Predicted Value for NWF CLS Change Median Score
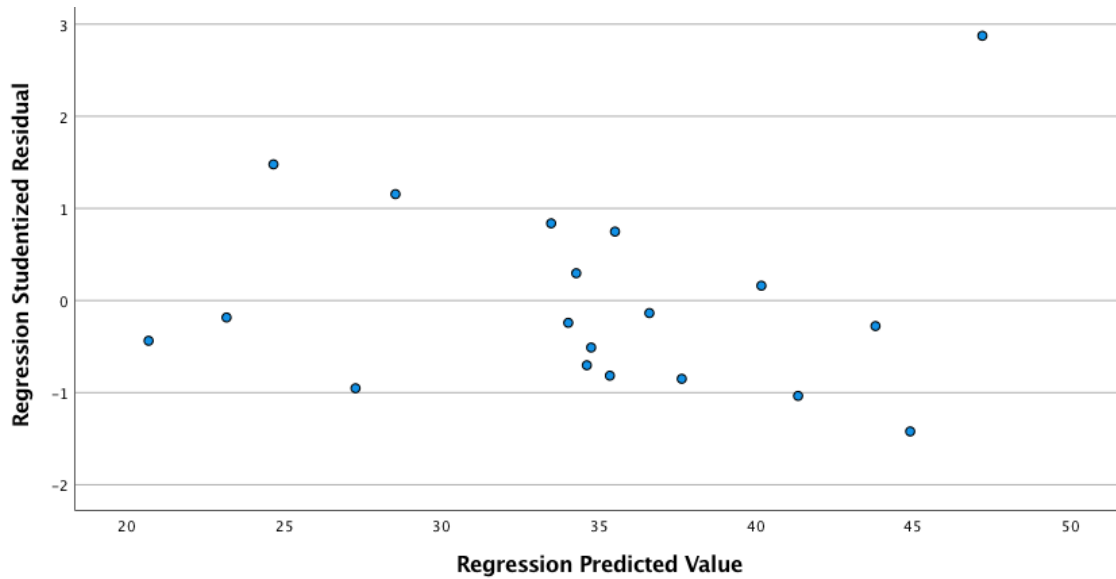
Figure 20

Scatterplot of Regression Predicted Value for NWF-WRC BOY Mean Score
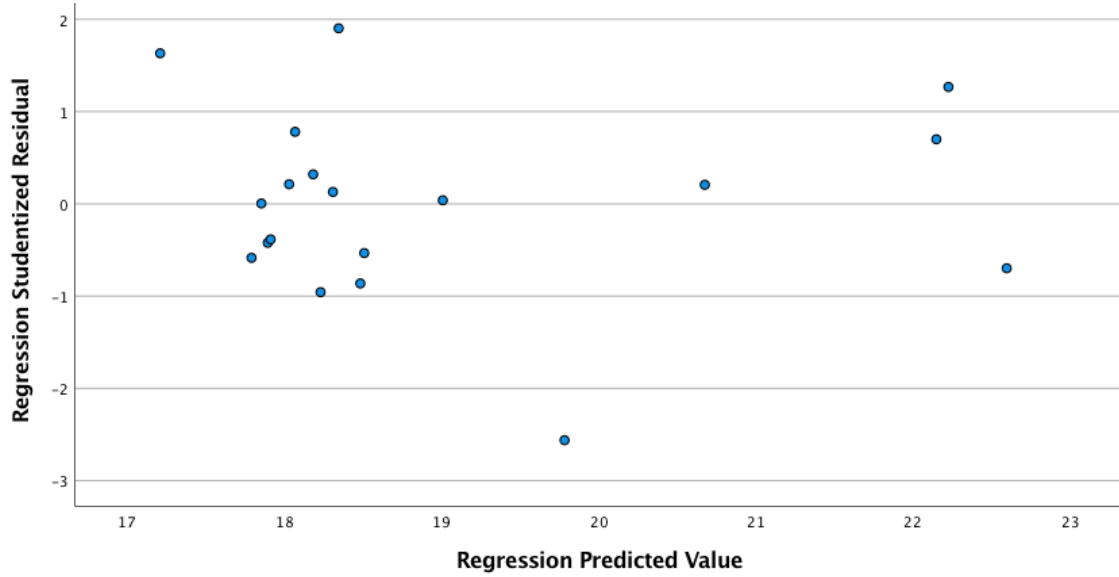
Figure 21

Scatterplot of Regression Predicted Value for NWF-WRC BOY Median Score
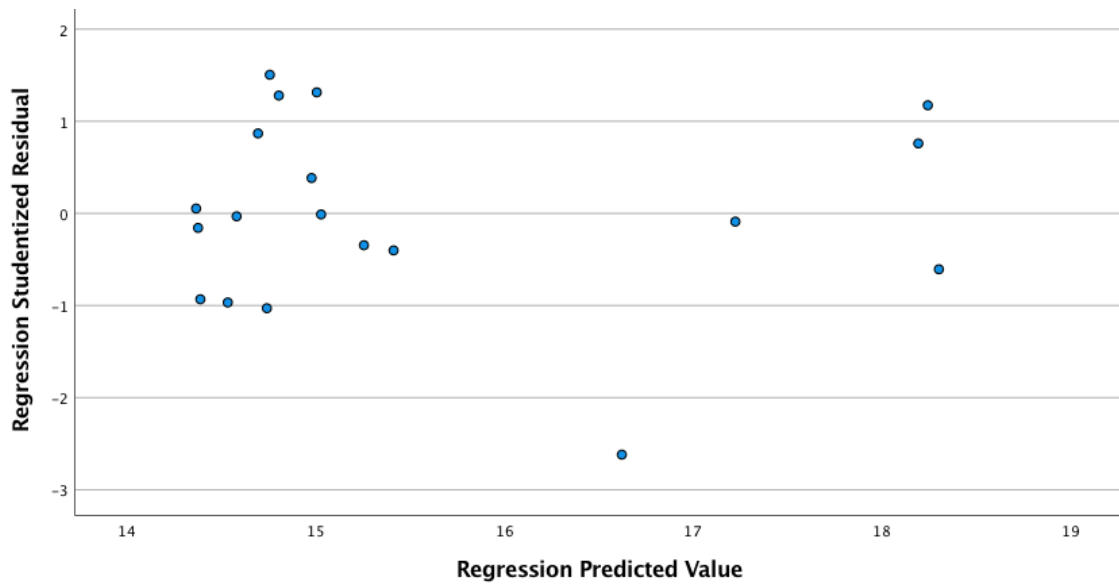
Figure 22

Scatterplot of Regression Predicted Value for NWF-WRC Change Mean Score
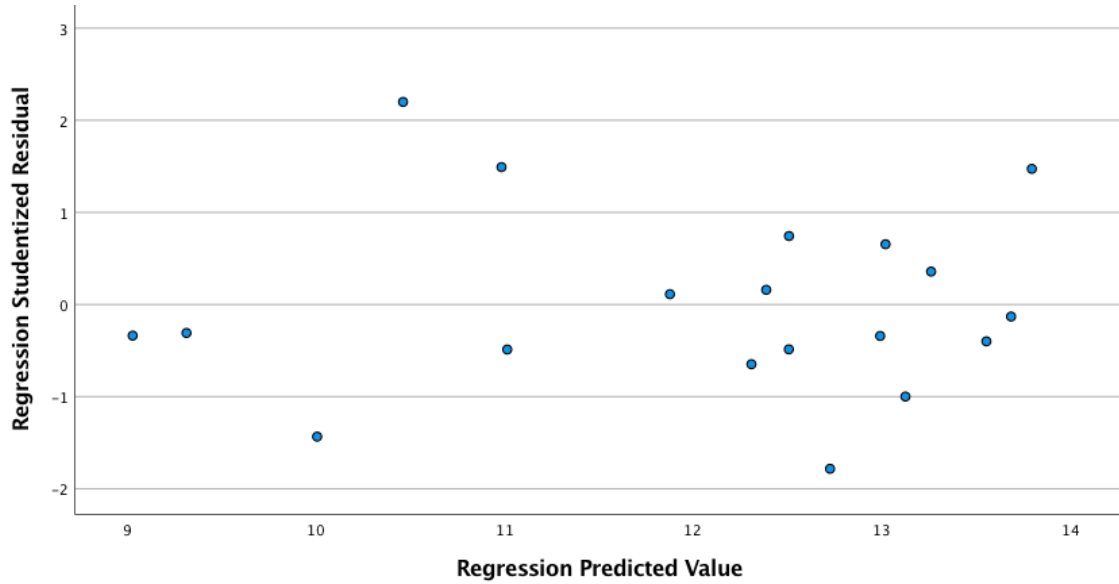
Figure 23

Scatterplot of Regression Predicted Value for NWF-WRC Change Median Score
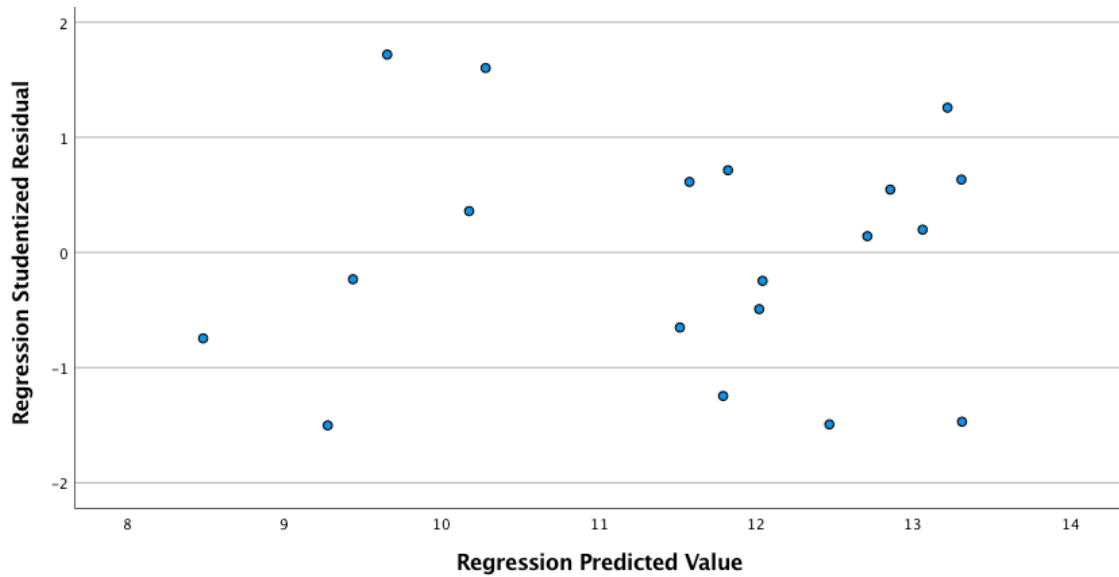
Figure 24

Scatterplot of Regression Predicted Value for ORF BOY Mean Score
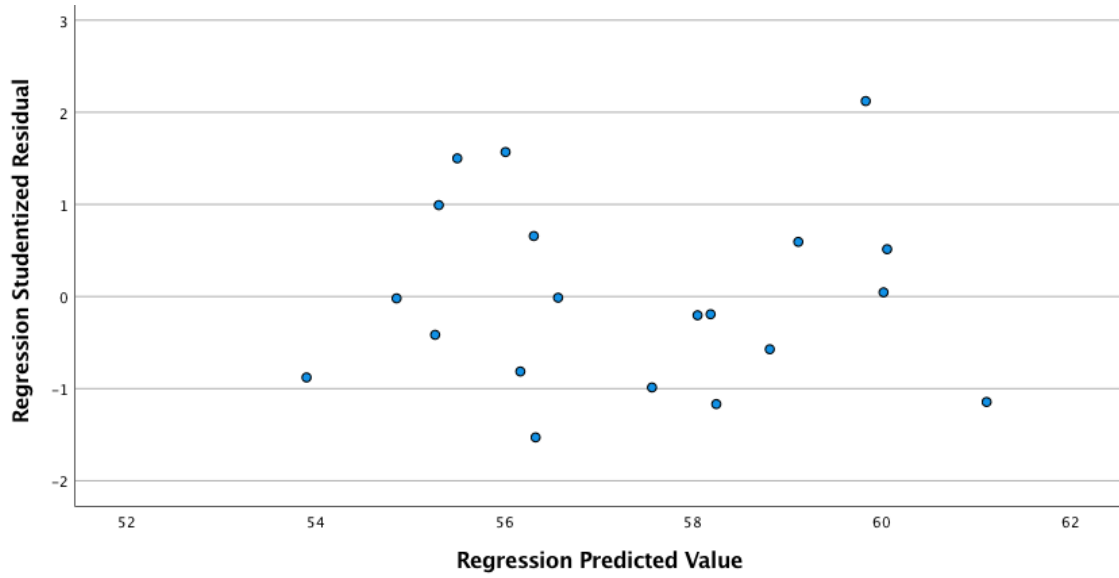
Figure 25

Scatterplot of Regression Predicted Value for ORF BOY Median Score
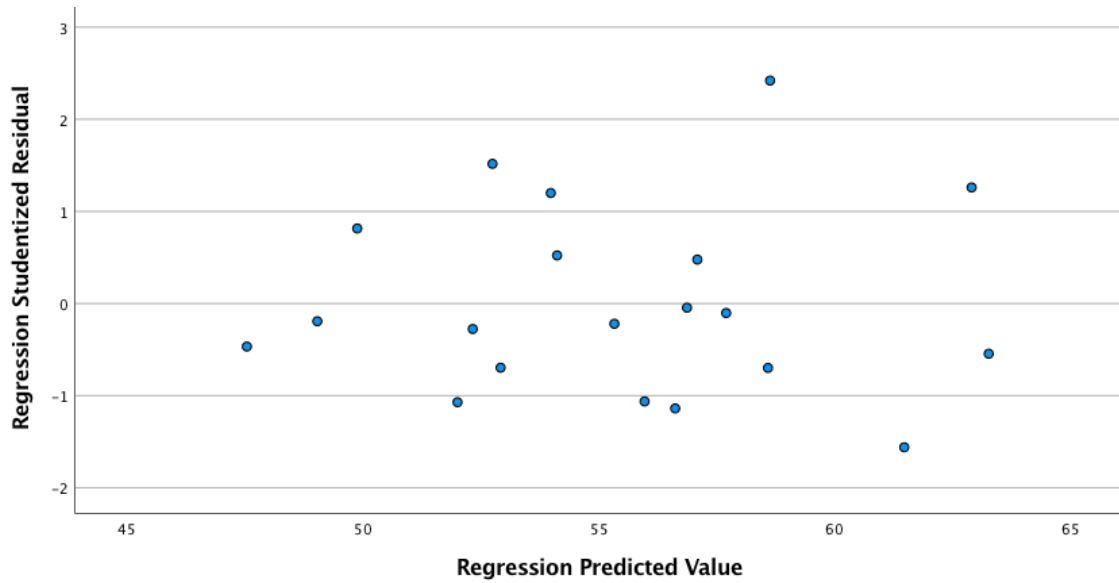
Figure 26

Scatterplot of Regression Predicted Value for ORF Change Mean Score
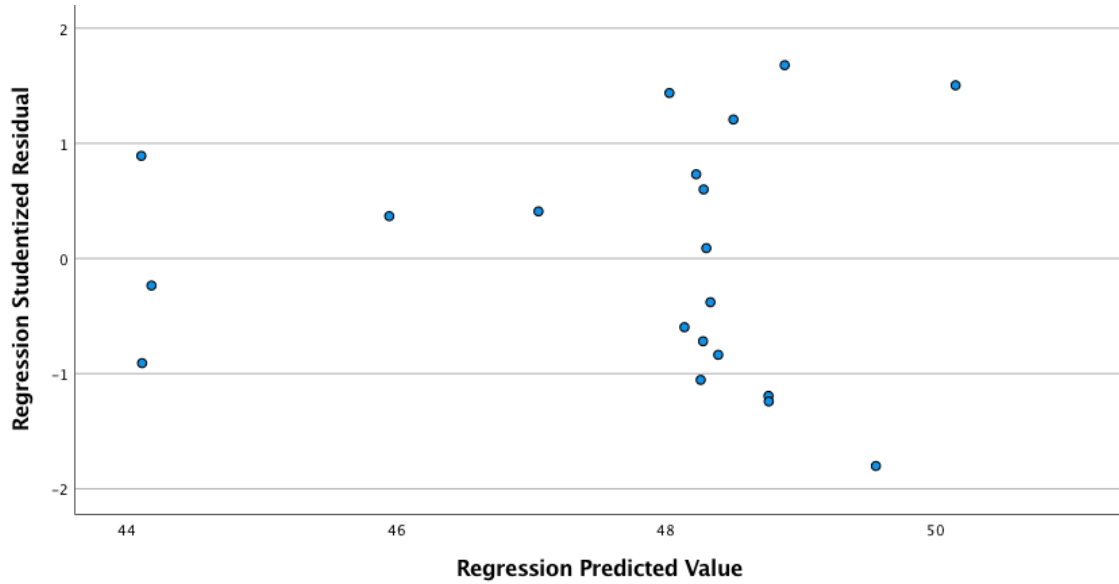
Figure 27

Scatterplot of Regression Predicted Value for ORF Change Median Score
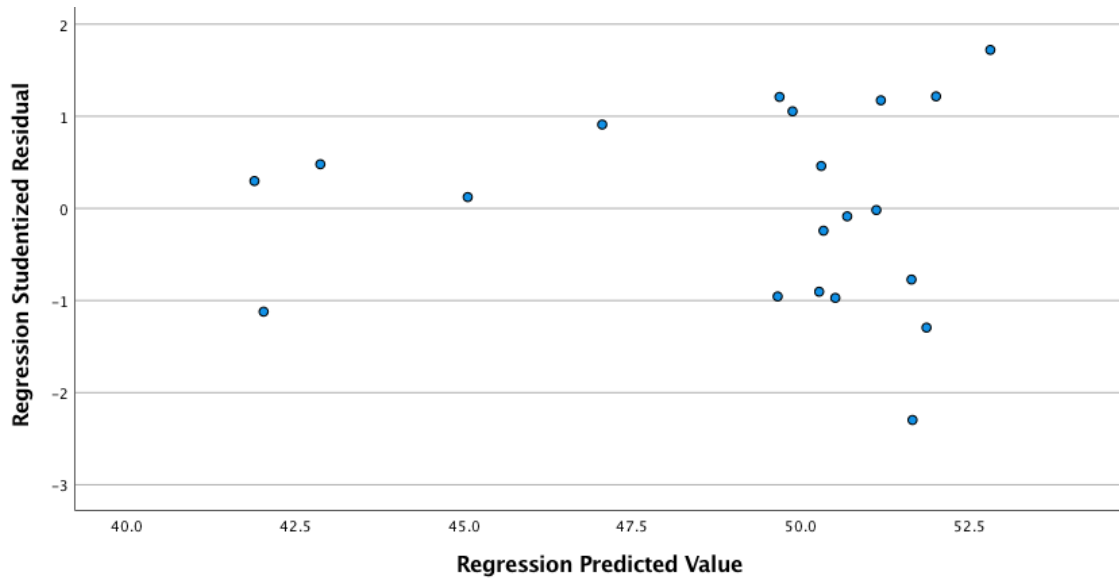
Figure 28

Scatterplot of Regression Predicted Value for WRF BOY Mean Score
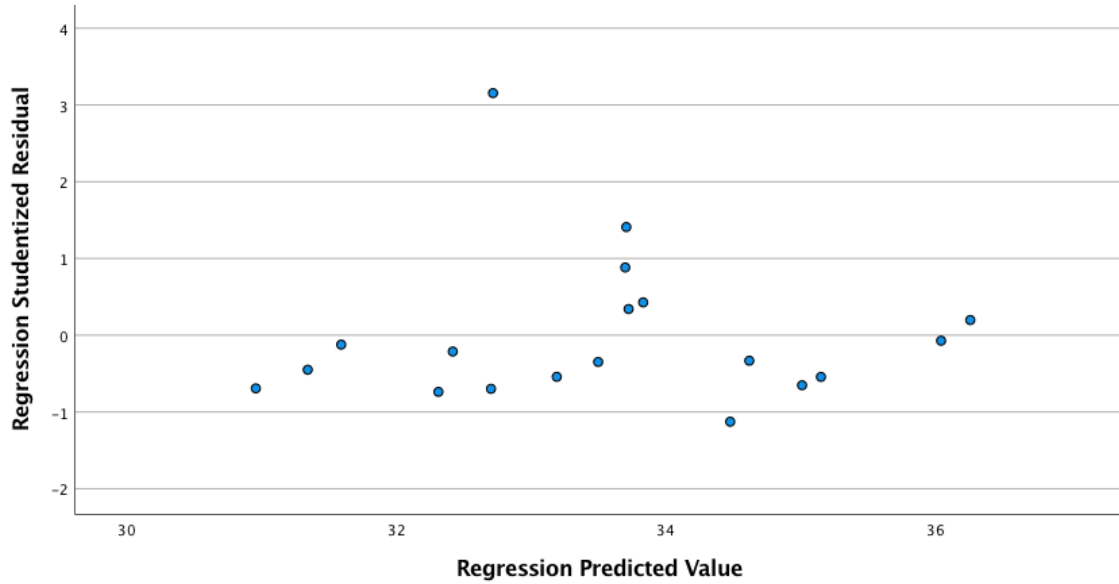
Figure 29

Scatterplot of Regression Predicted Value for WRF BOY Median Score
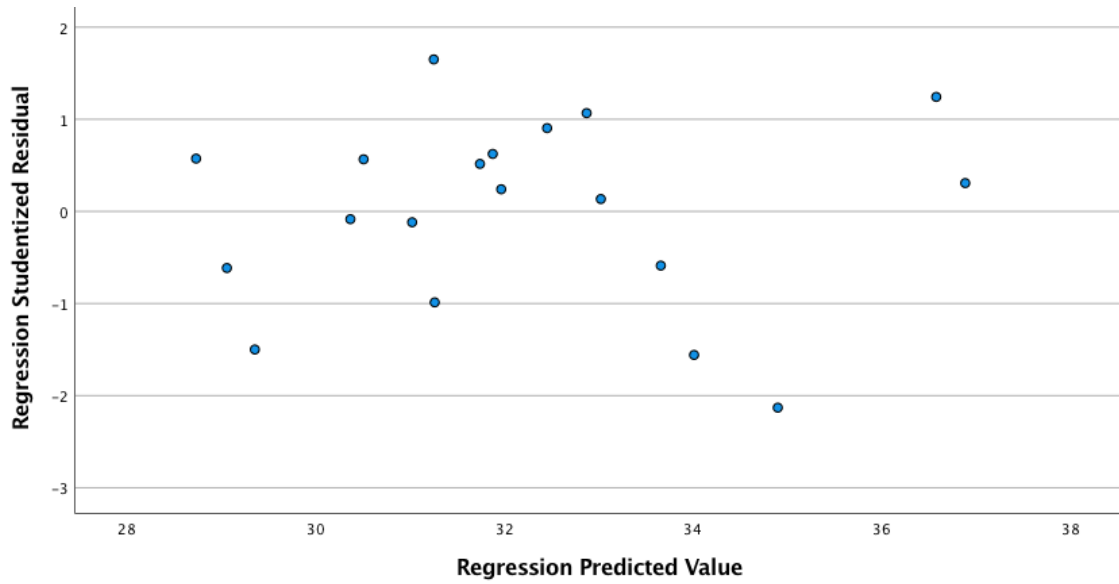
Figure 30

Scatterplot of Regression Predicted Value for WRF Change Mean Score
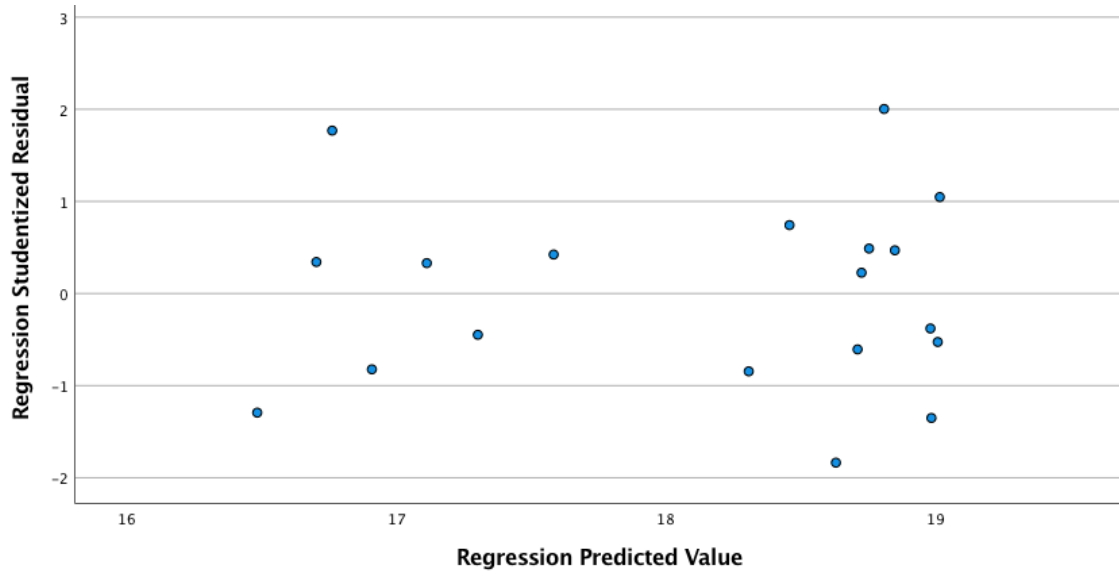
Figure 31

Scatterplot of Regression Predicted Value for WRF Change Median Score
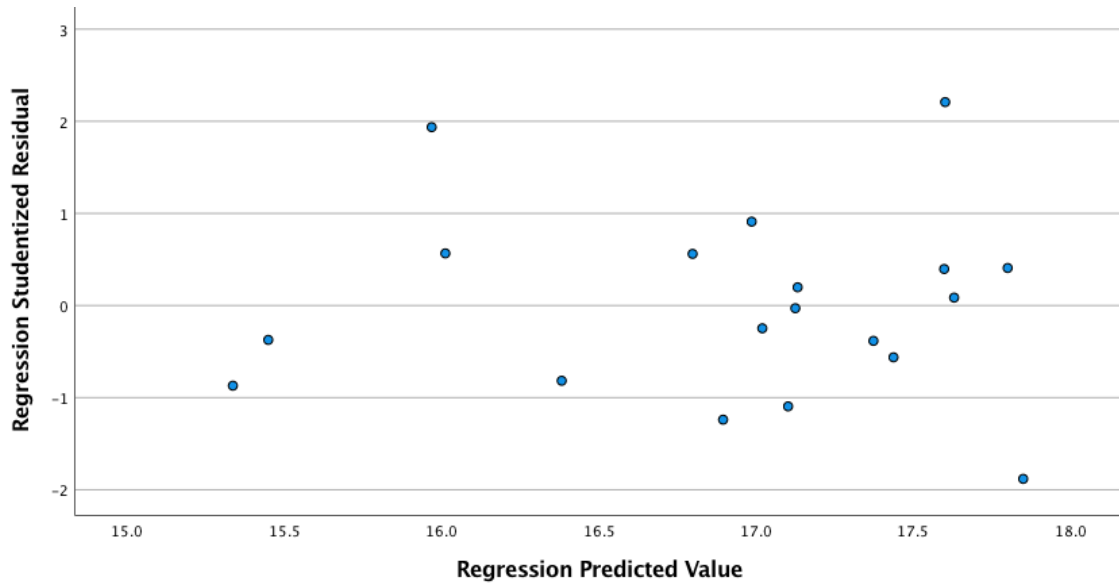
HISTOGRAMS OF REGRESSION STANDARDIZED RESIDUAL FOR ALL DEPENDENT

VARIABLES

Figure 32

Histogram of Regression Standardized Residual for NWF-CLS BOY Mean Score
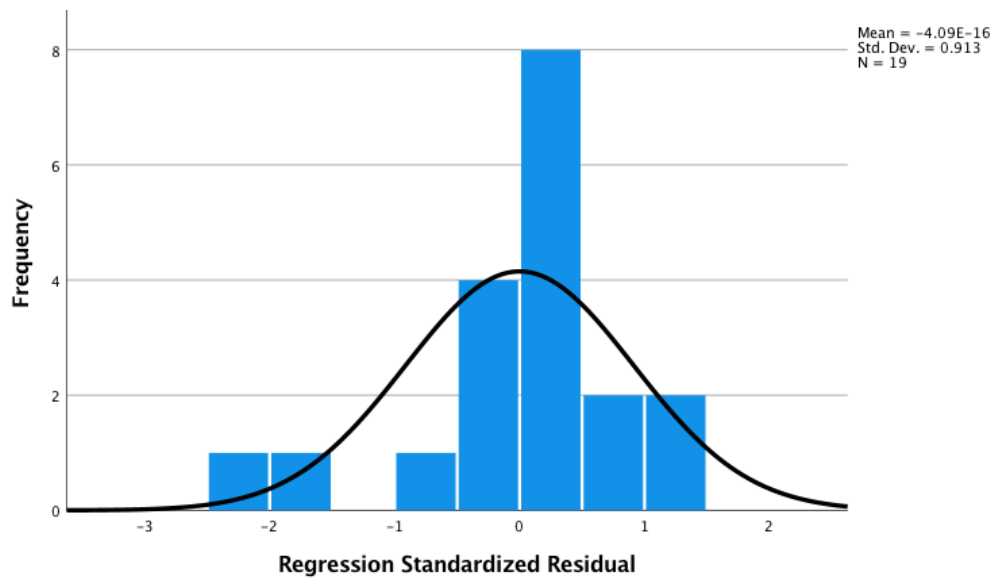
Figure 33

Histogram of Regression Standardized Residual for NWF-CLS BOY Mean Score
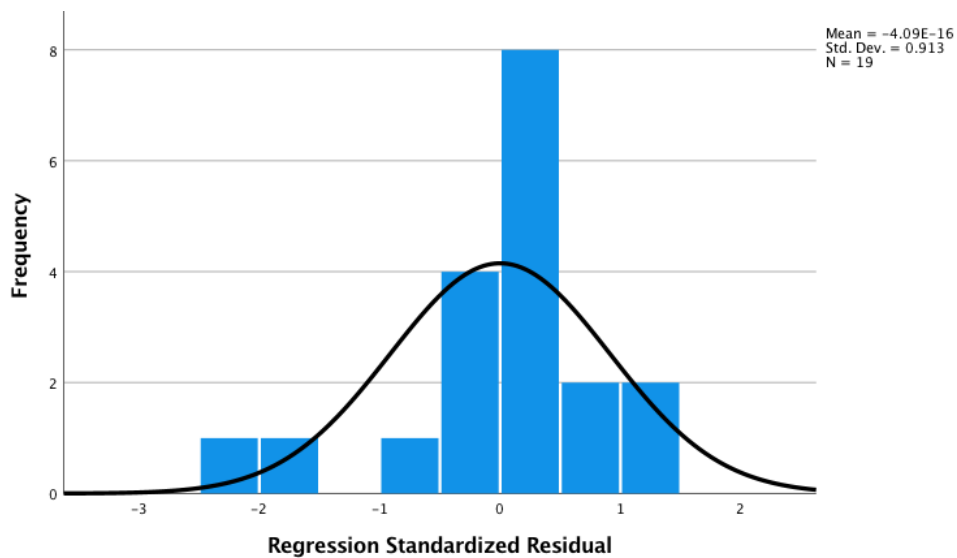
Figure 34

Histogram of Regression Standardized Residual for NWF-CLS Change Mean Score
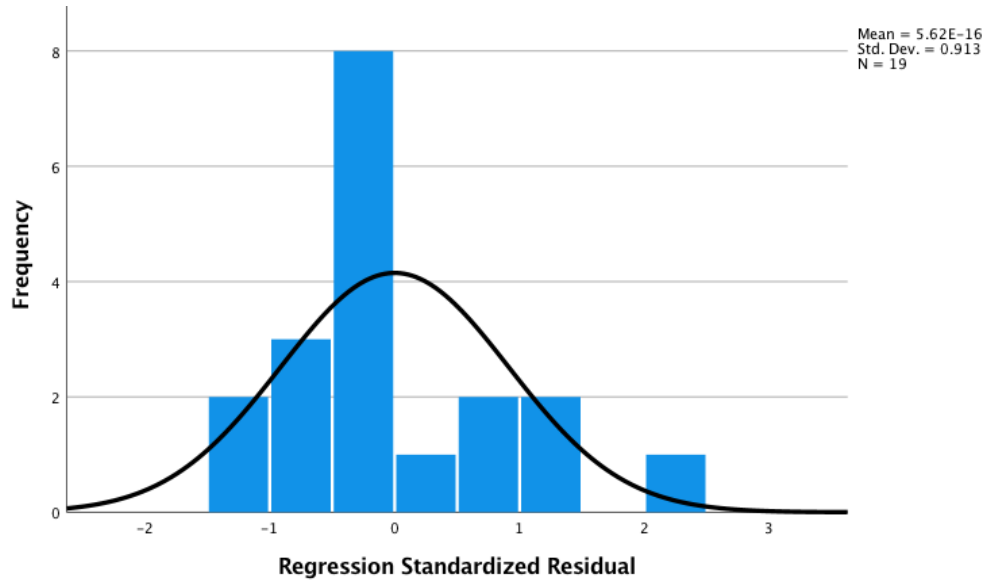
Figure 35

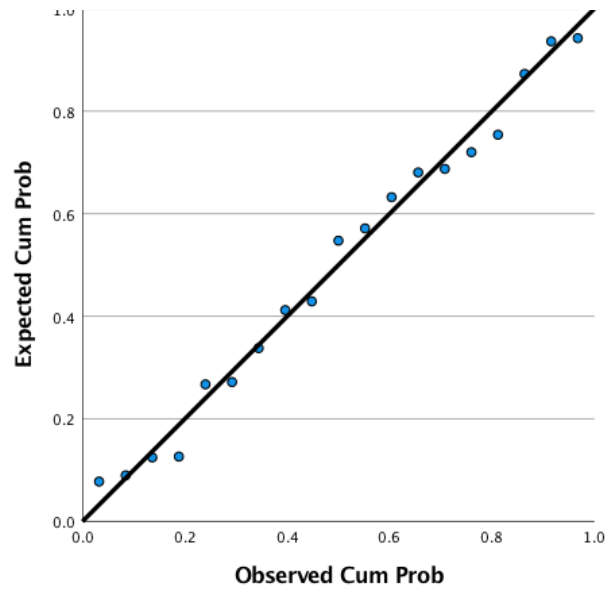*Histogram of Regression Standardized Residual for NWF-CLS Change Median Score*
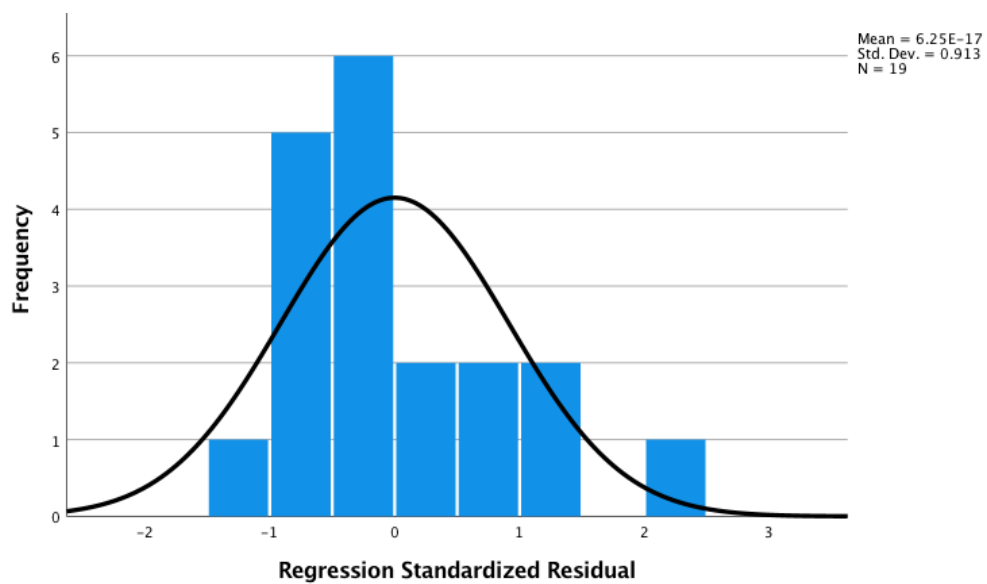
Figure 36

Histogram of Regression Standardized Residual for NWF-WRC BOY Mean Score
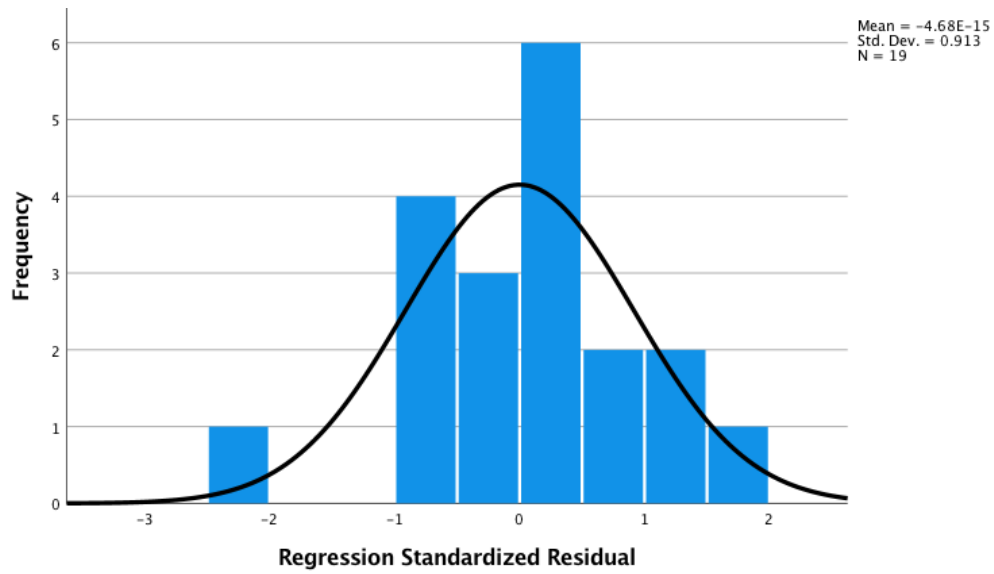
Figure 37

Histogram of Regression Standardized Residual for NWF-WRC BOY Median Score
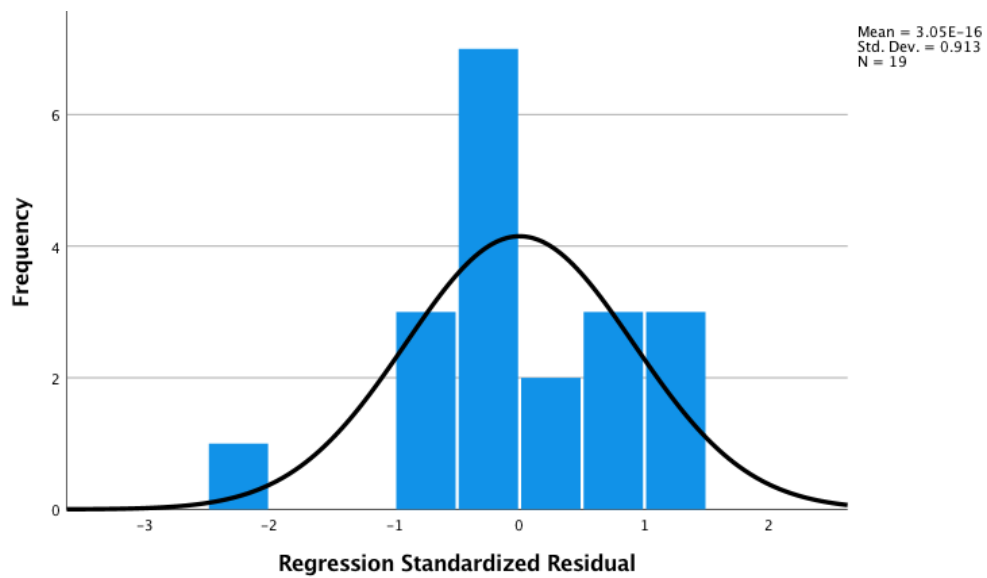
Figure 38

Histogram of Regression Standardized Residual for NWF-WRC Change Mean Score
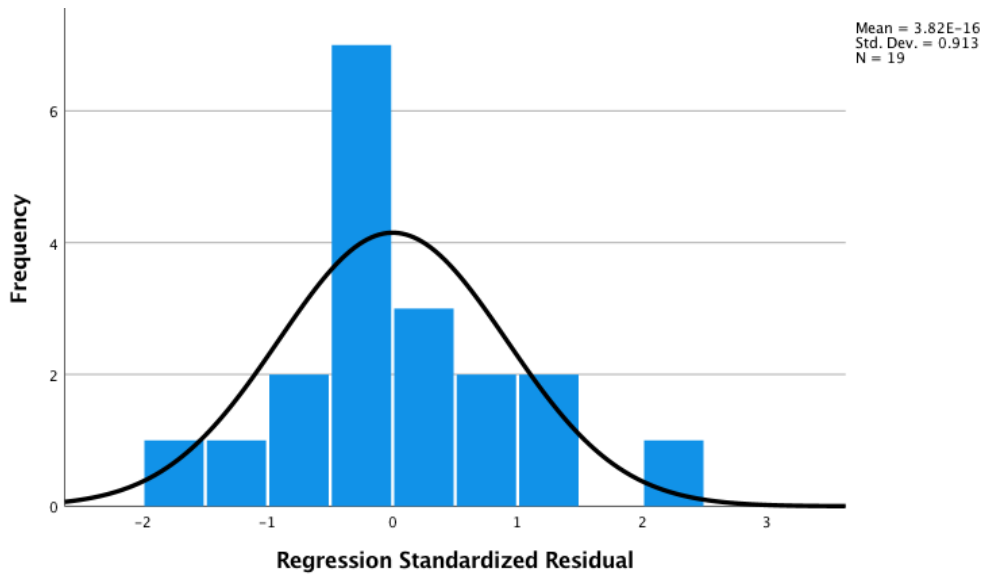
Figure 39

Histogram of Regression Standardized Residual for NWF-WRC Change Median Score
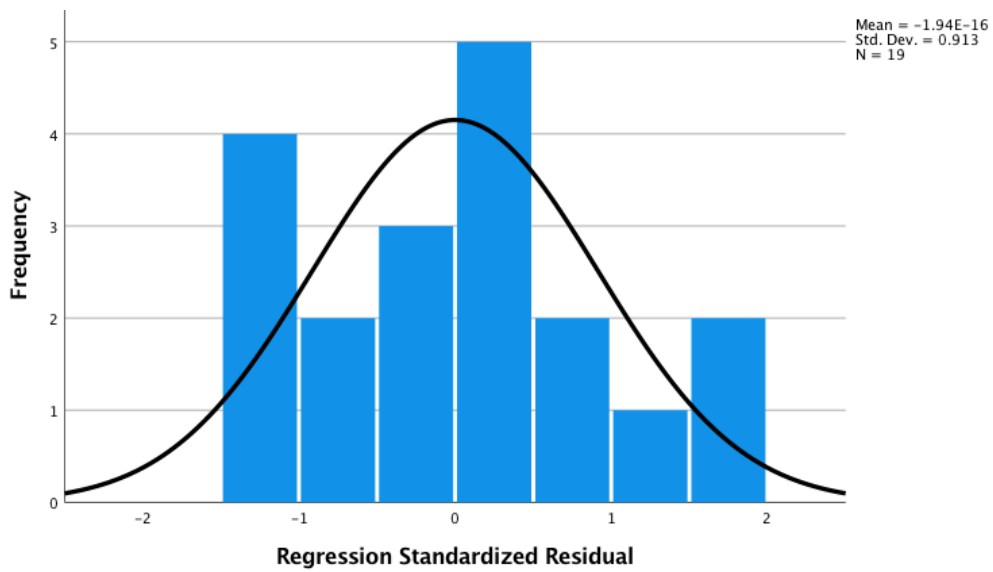
Figure 40

Histogram of Regression Standardized Residual for ORF BOY Mean Score
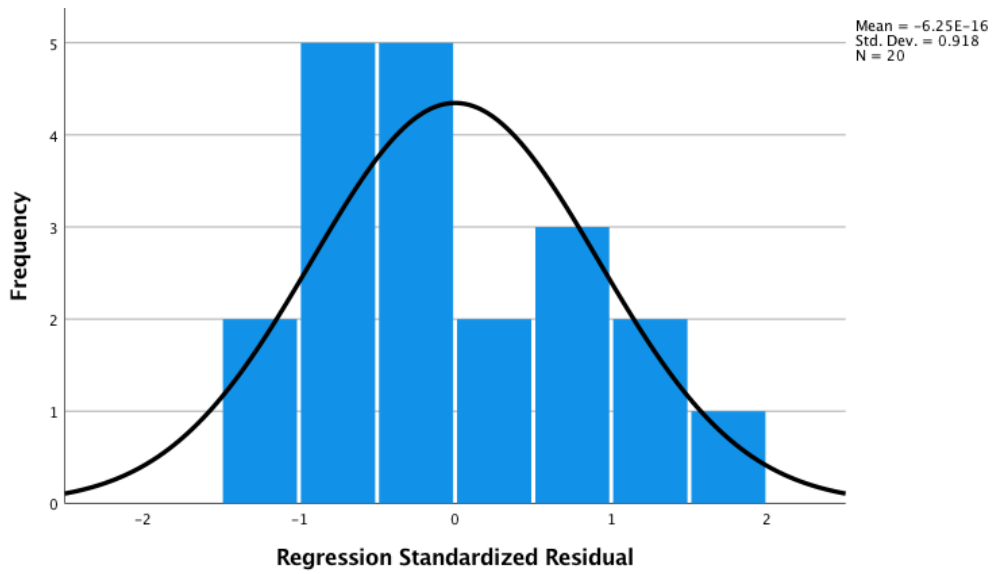
Figure 41

Histogram of Regression Standardized Residual for ORF BOY Median Score
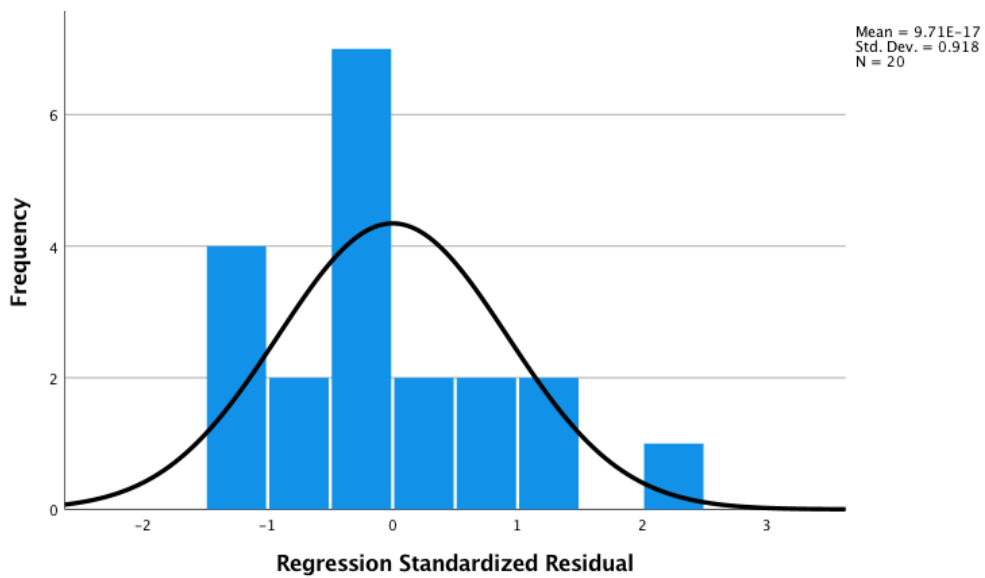
Figure 42

Histogram of Regression Standardized Residual for ORF Change Mean Score
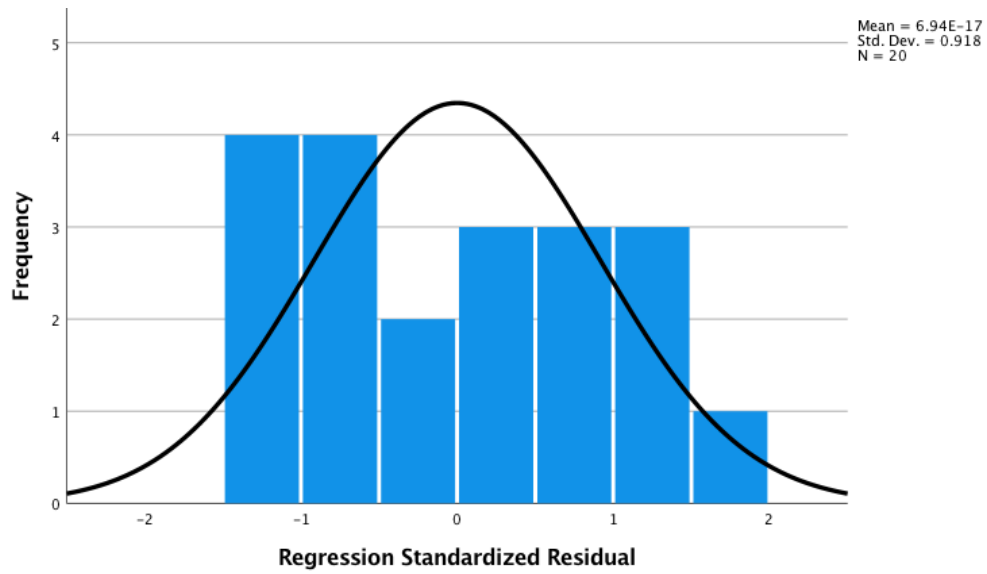
Figure 43

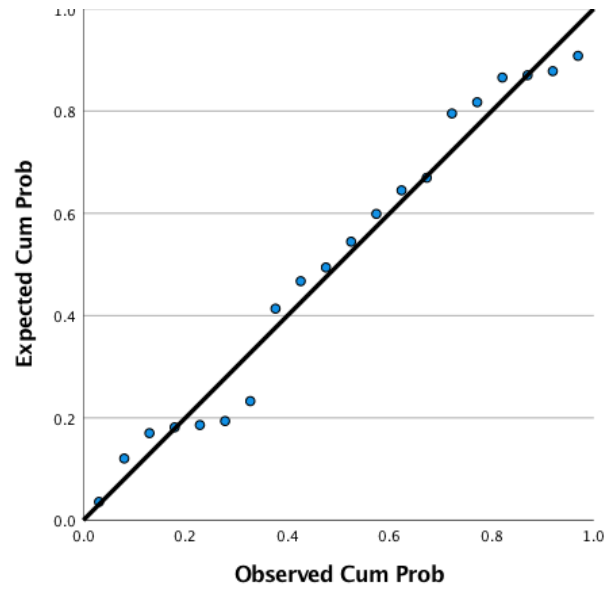*Histogram of Regression Standardized Residual for ORF Change Median Score*
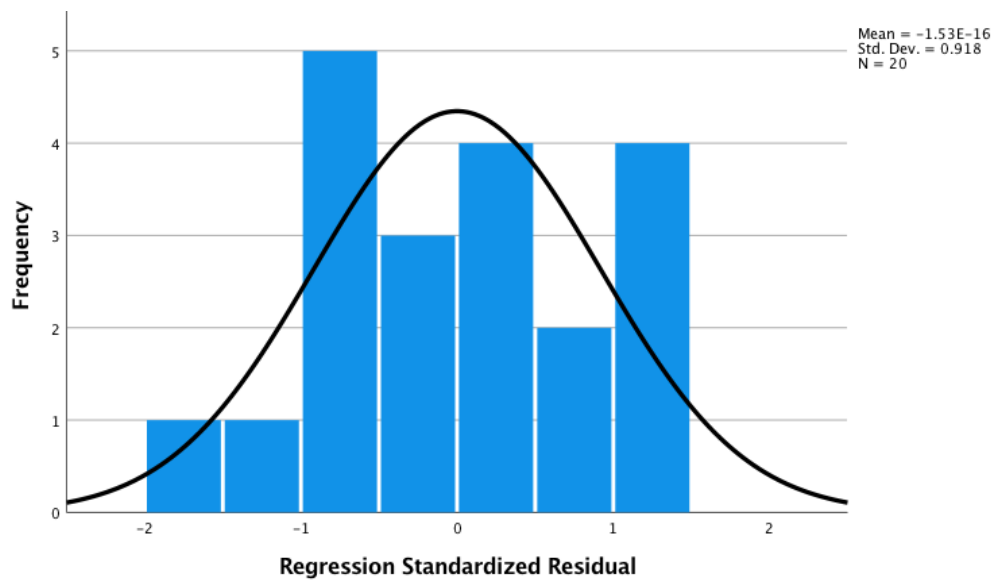
Figure 44

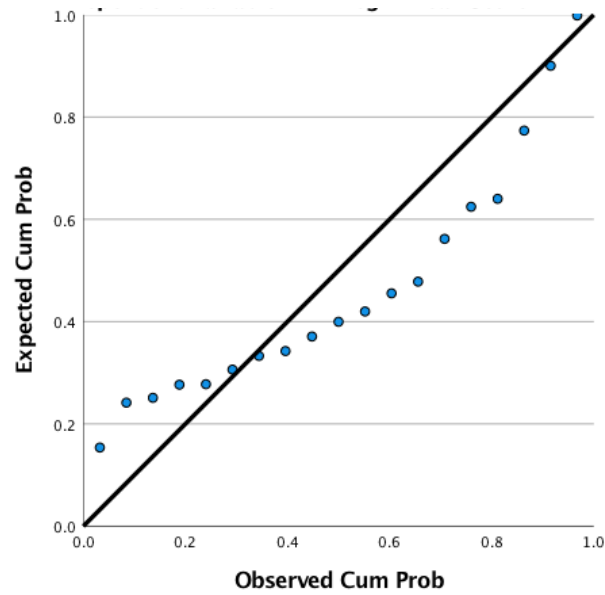*Histogram of Regression Standardized Residual for WRF BOY Mean Score*
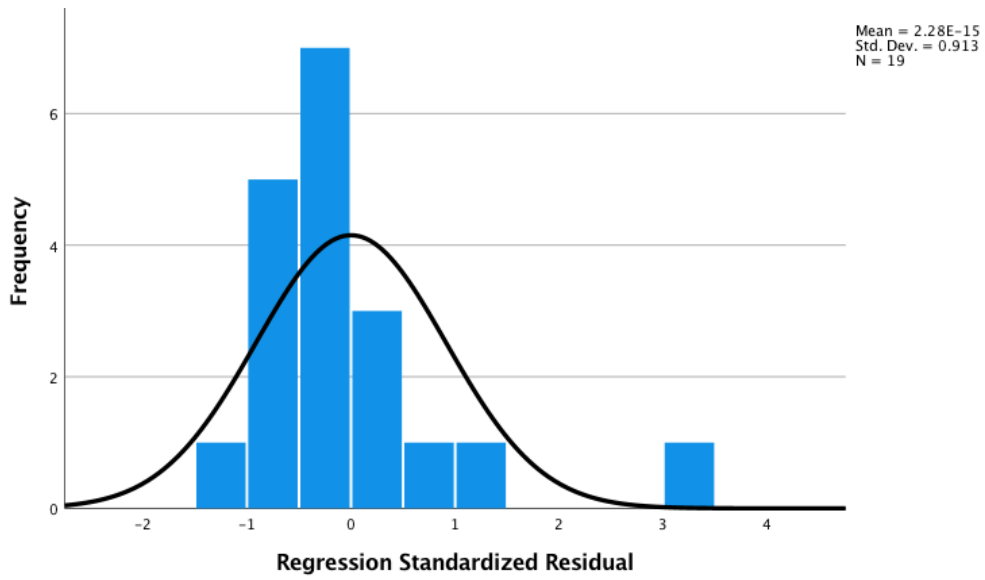
Figure 45

Histogram of Regression Standardized Residual for WRF BOY Median Score
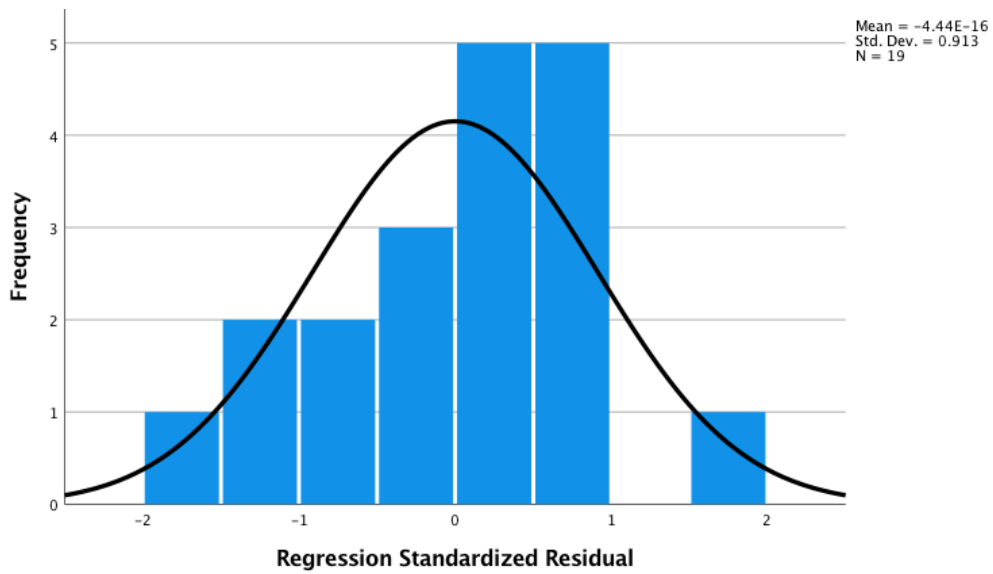
Figure 46

Histogram of Regression Standardized Residual for WRF Change Mean Score
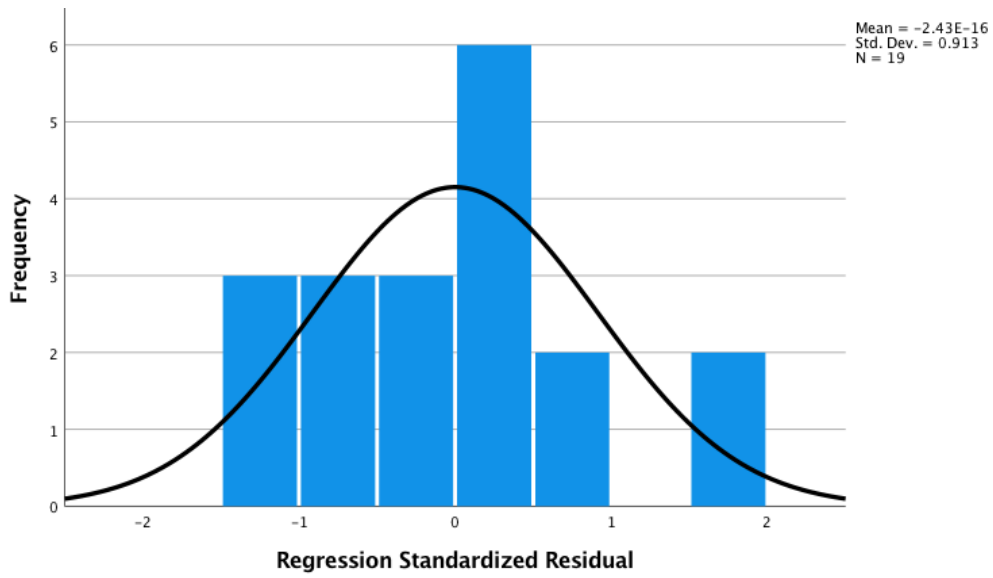
Figure 47

Histogram of Regression Standardized Residual for WRF Change Median Score
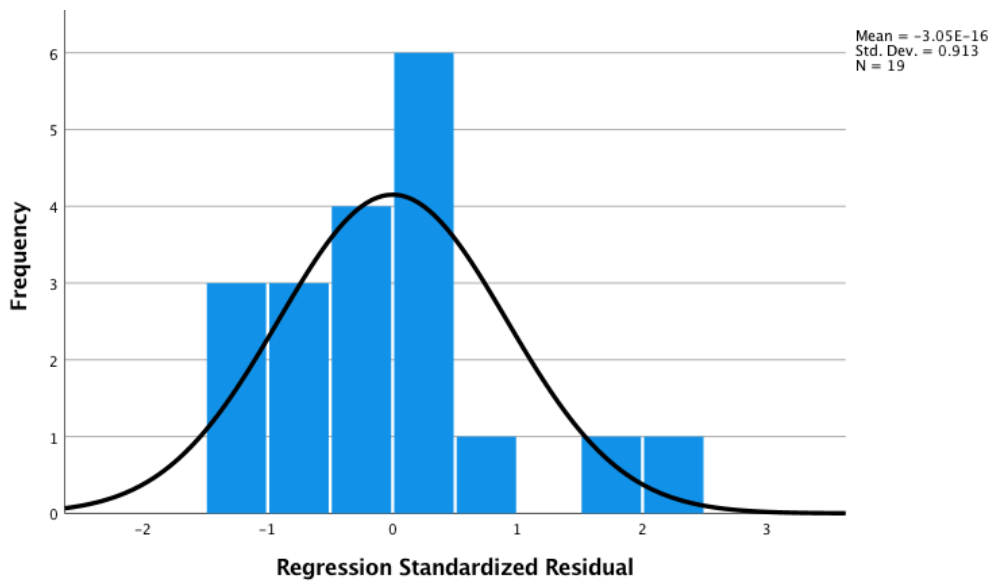
APPENDIX C

NORMAL P-P PLOTS OF REGRESSION STANDARDIZED RESIDUALS FOR ALL DEPENDENT
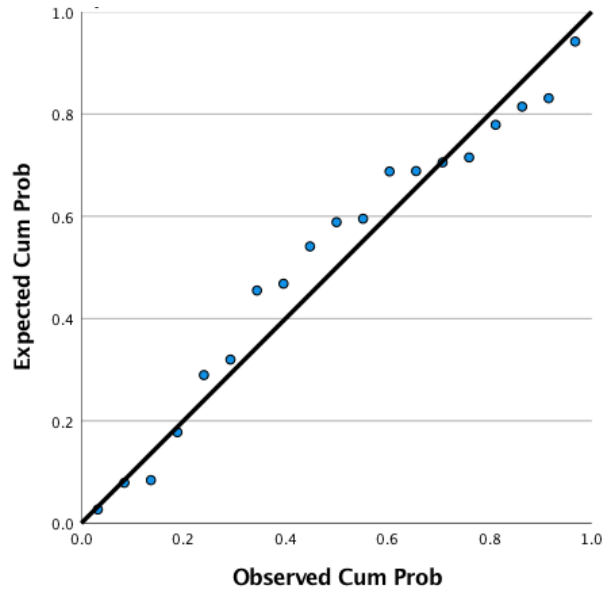
VARIABLES

Figure 48

Normal P-P Plot of Regression Standardized Residual for NWF-CLS BOY Mean Score
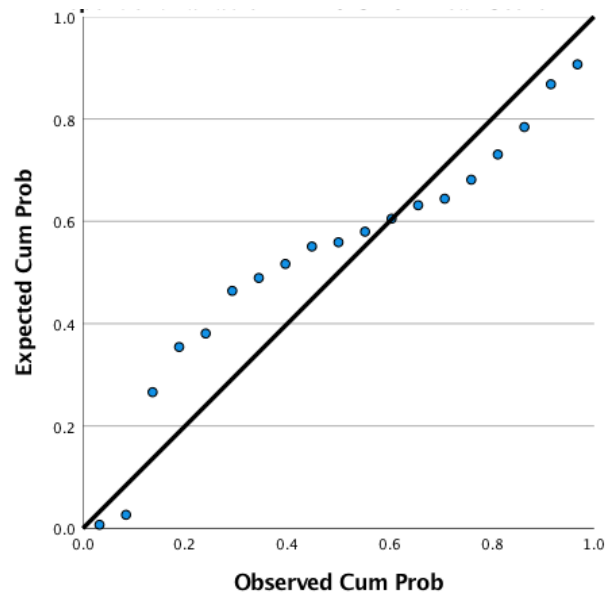
Figure 49

Normal P-P Plot of Regression Standardized Residual for NWF-CLS BOY Median Score
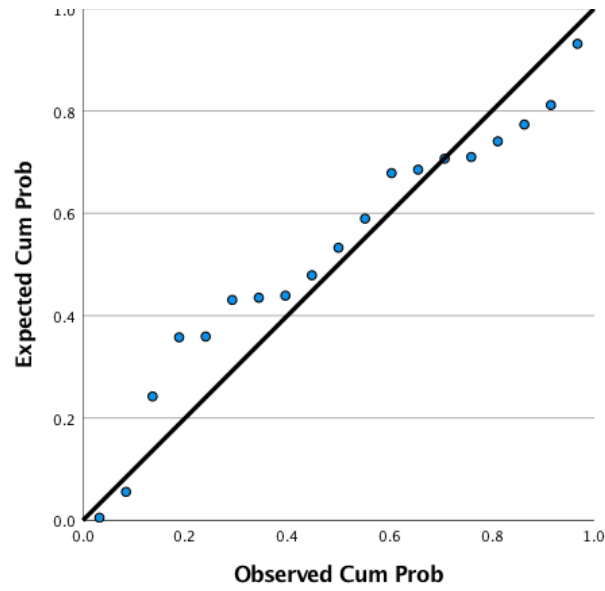
Figure 50

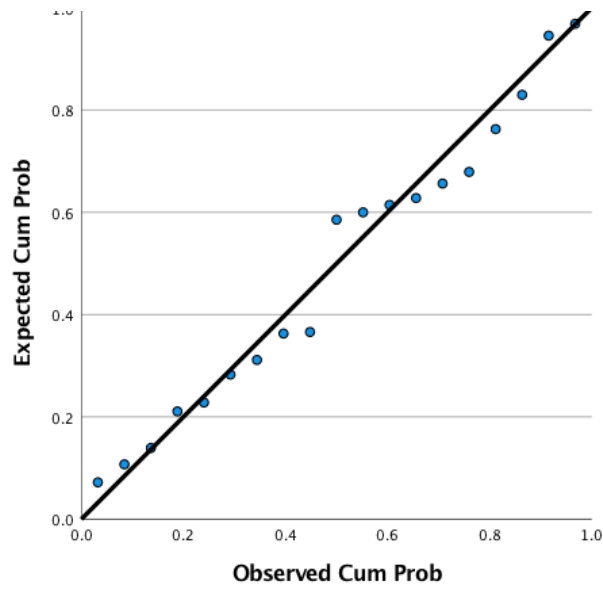*Normal P-P Plot of Regression Standardized Residual for NWF-CLS Change Mean Score*
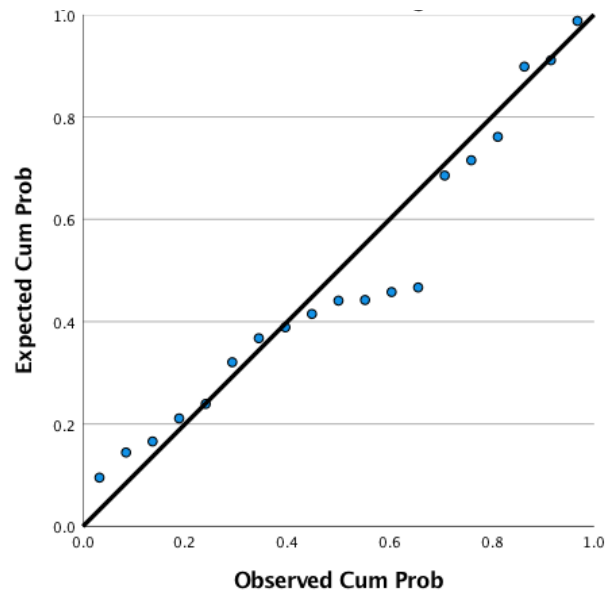
Figure 51

Normal P-P Plot of Regression Standardized Residual for NWF-CLS Change Median Score
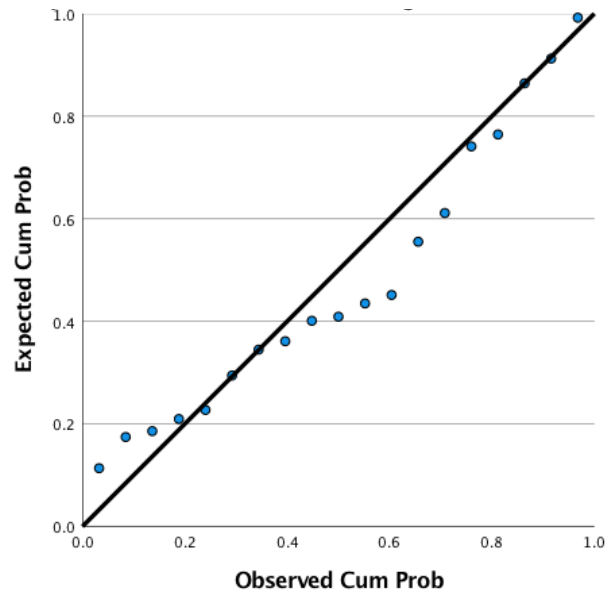
Figure 52

Normal P-P Plot of Regression Standardized Residual for NWF-WRC BOY Mean Score
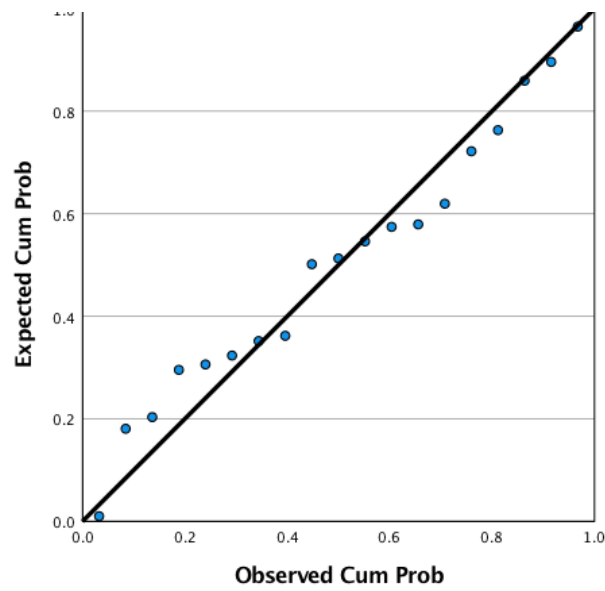
Figure 53

Normal P-P Plot of Regression Standardized Residual for NWF-WRC BOY Median Score
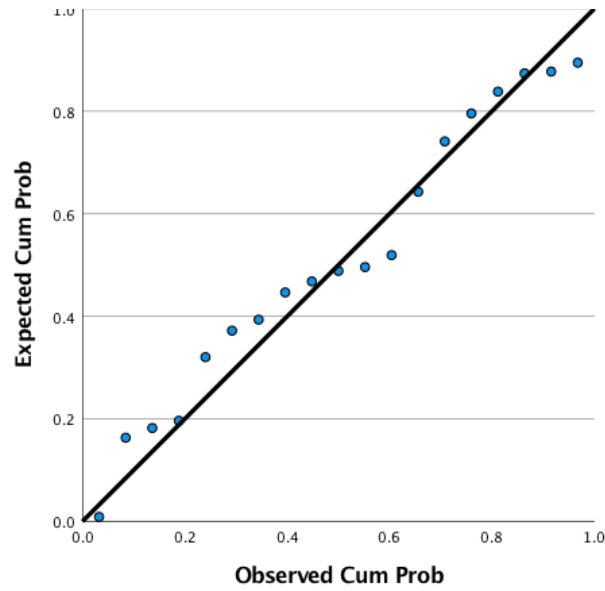
Figure 54

Normal P-P Plot of Regression Standardized Residual for NWF-WRC Change Mean Score
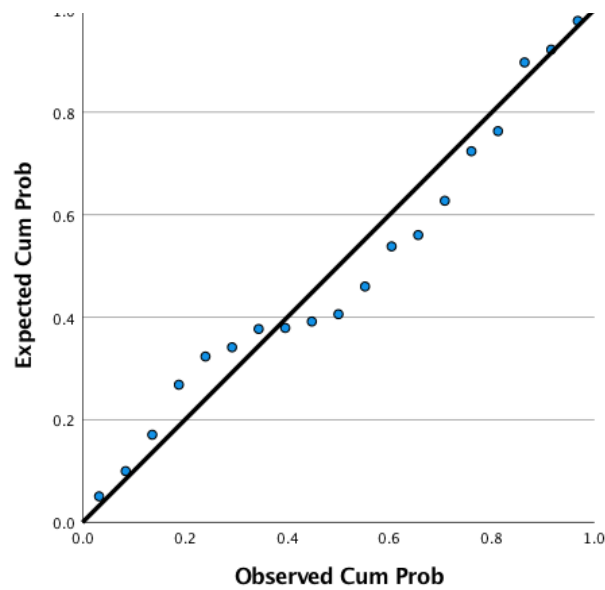
Figure 55

Normal P-P Plot of Regression Standardized Residual for NWF-WRC Change Median

Score

Figure 56

Normal P-P Plot of Regression Standardized Residual for ORF BOY Mean Score
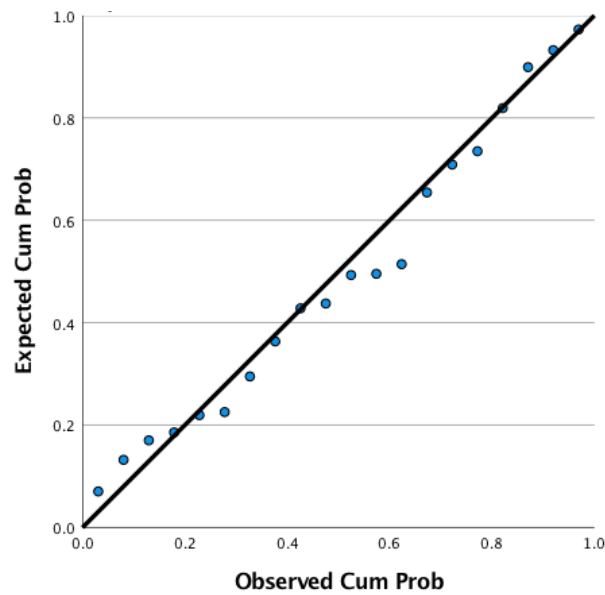
Figure 57

Normal P-P Plot of Regression Standardized Residual for ORF BOY Median Score
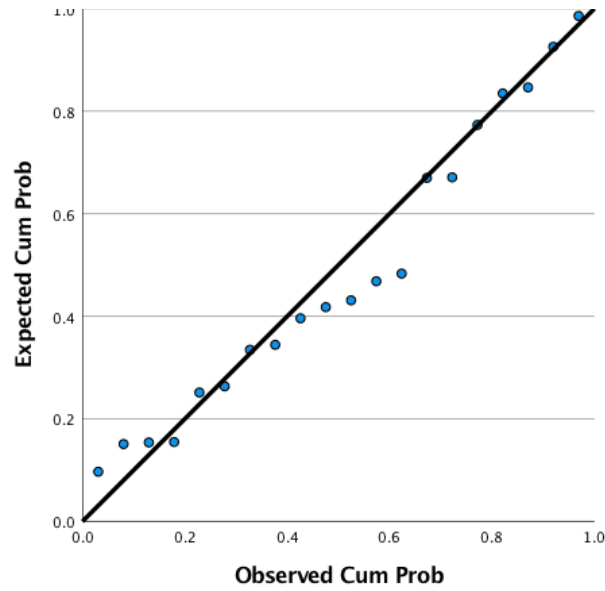
Figure 58

Normal P-P Plot of Regression Standardized Residual for ORF Change Mean Score
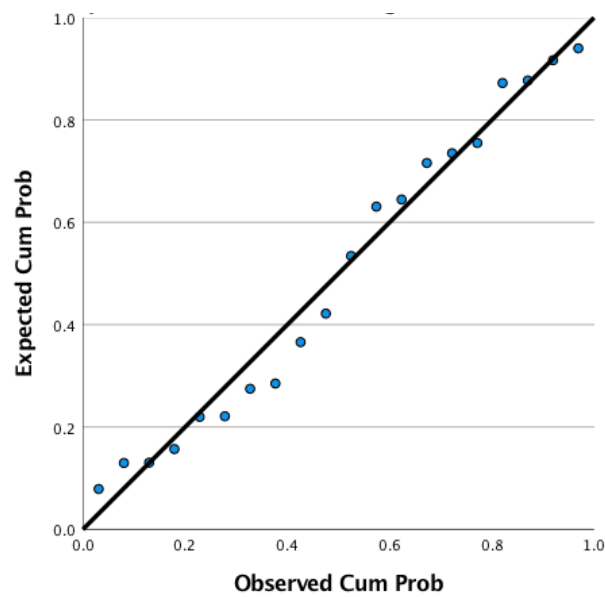
Figure 59

Normal P-P Plot of Regression Standardized Residual for ORF Change Median Score

Figure 60

Normal P-P Plot of Regression Standardized Residual for WRF Begin Mean Score

Figure 61

Normal P-P Plot of Regression Standardized Residual for WRF BOY Median Score

Figure 62

Normal P-P Plot of Regression Standardized Residual for WRF Change Mean Score

Figure 63

Normal P-P Plot of Regression Standardized Residual for WRF Change Median Score
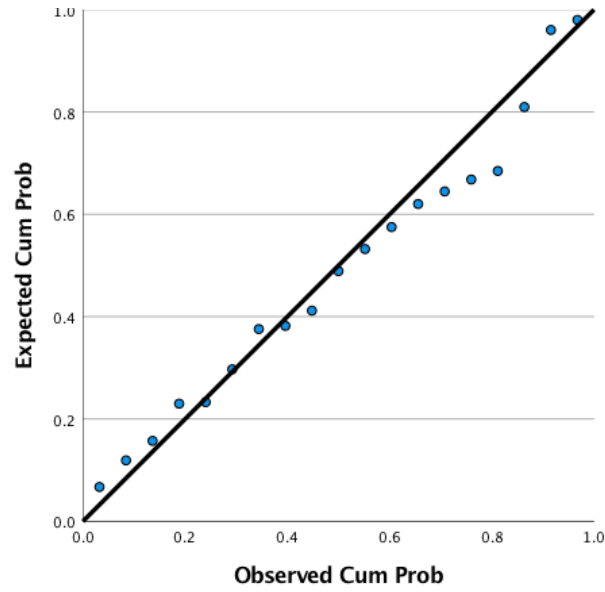
APPENDIX D

SCATTERPLOTS OF SCHOOL PROPORTION OF FARMS BY PROPORTION OF STUDENTS

TESTED FOR EACH DIBELS-8 SUBTEST

Figure 64

NWF: Scatterplot of School Proportion of FARMS by Proportion of Students Tested
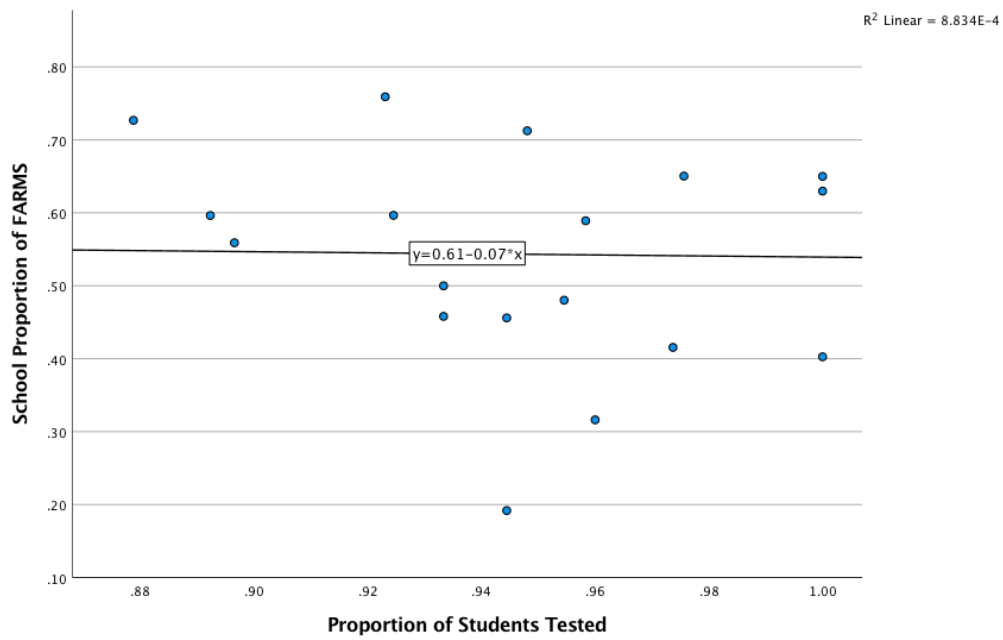
Figure 65

ORF: Scatterplot of School Proportion of FARMS by Proportion of Students Tested
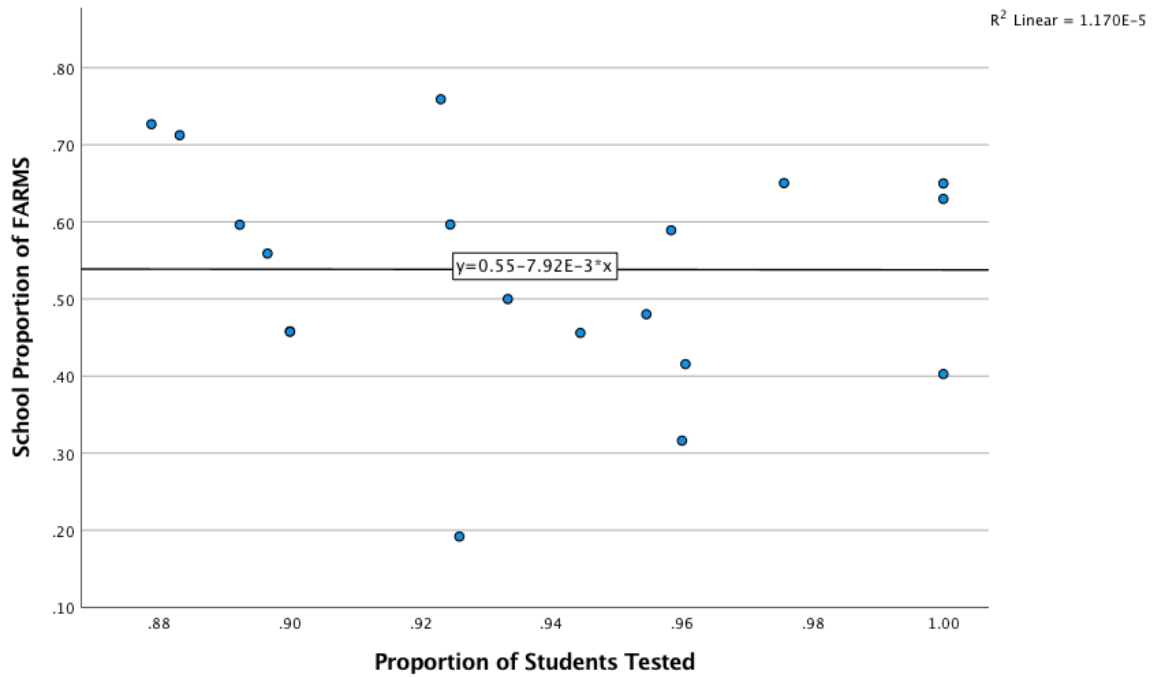
Figure 66

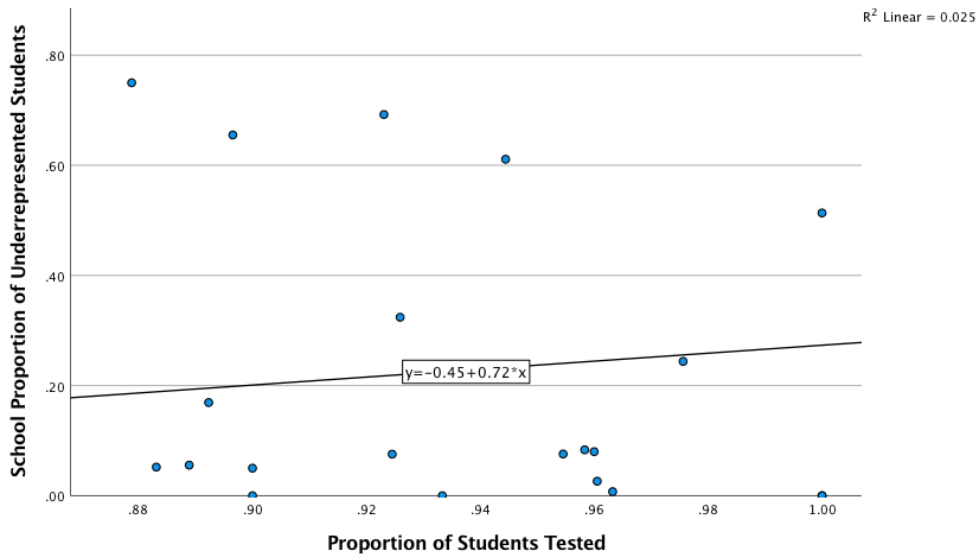*WRF: Scatterplot of School Proportion of FARMS by Proportion of Students Tested*
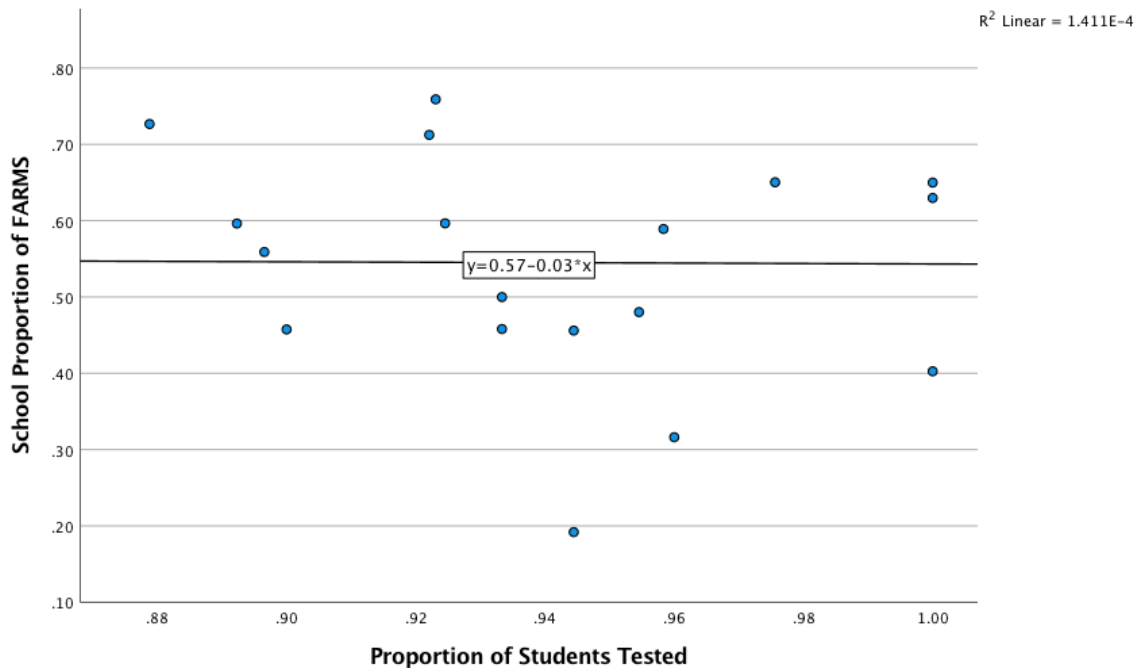
Figure 67

NWF: Scatterplot of School Proportion of Underrepresented Students by Proportion of
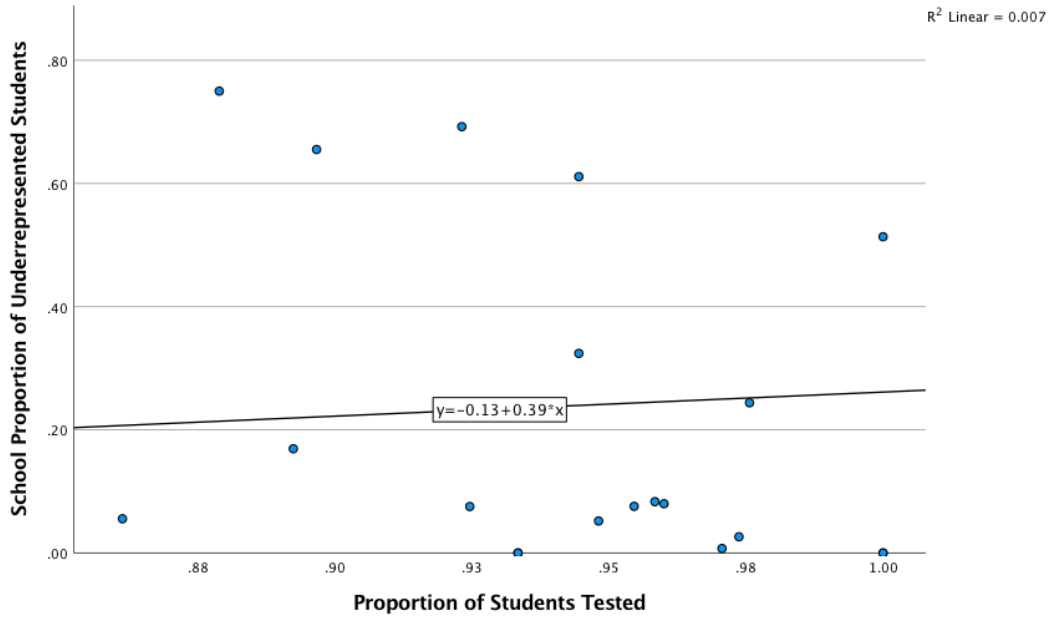
Students Tested

Figure 68

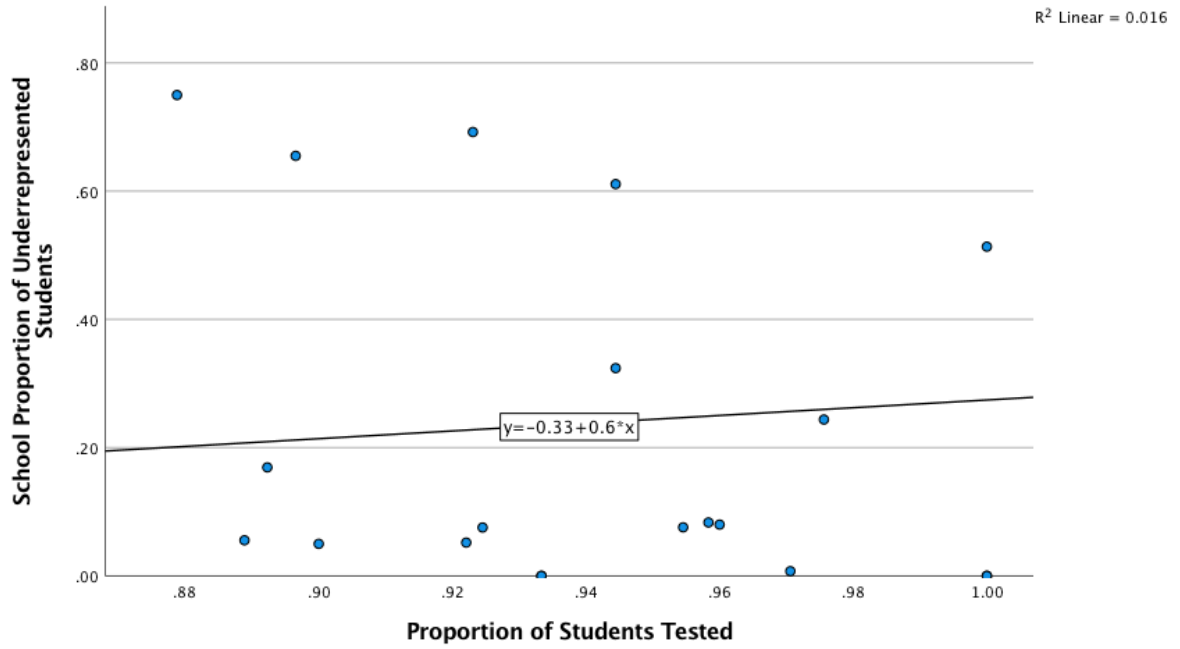*ORF: Scatterplot of School Proportion of Underrepresented Students by Proportion of*

Students Tested

Figure 69

WRF: Scatterplot of School Proportion of Underrepresented Students by Proportion of

Students Tested

SCHOOL PROPORTION OF STUDENTS FROM UNDERREPRESENTED GROUPS AND

RECEIVING FREE AND REDUCED-PRICE MEALS.

Figure 70

Estimated mean NWF-CLS at BOY as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
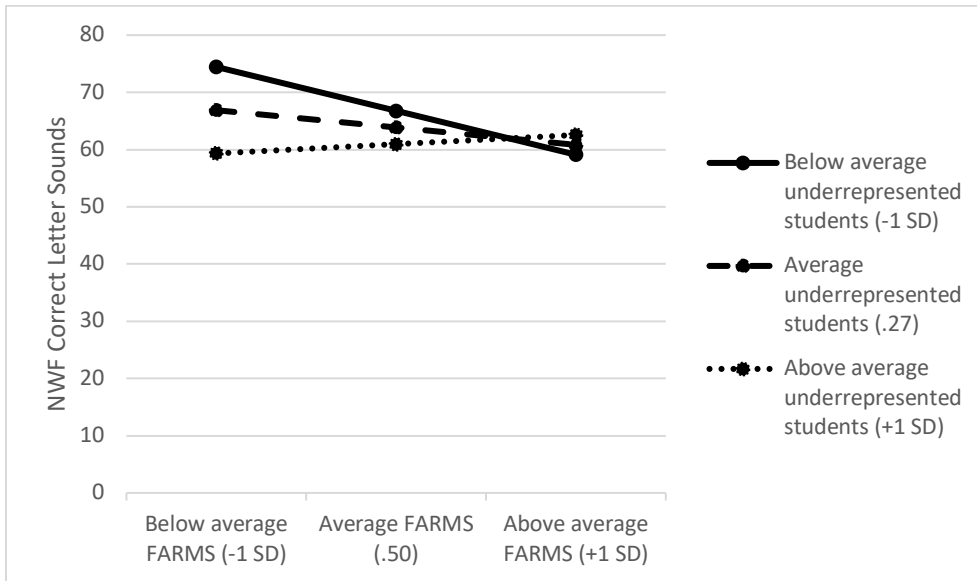
Figure 71

Estimated mean NWF-WRC at BOY as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
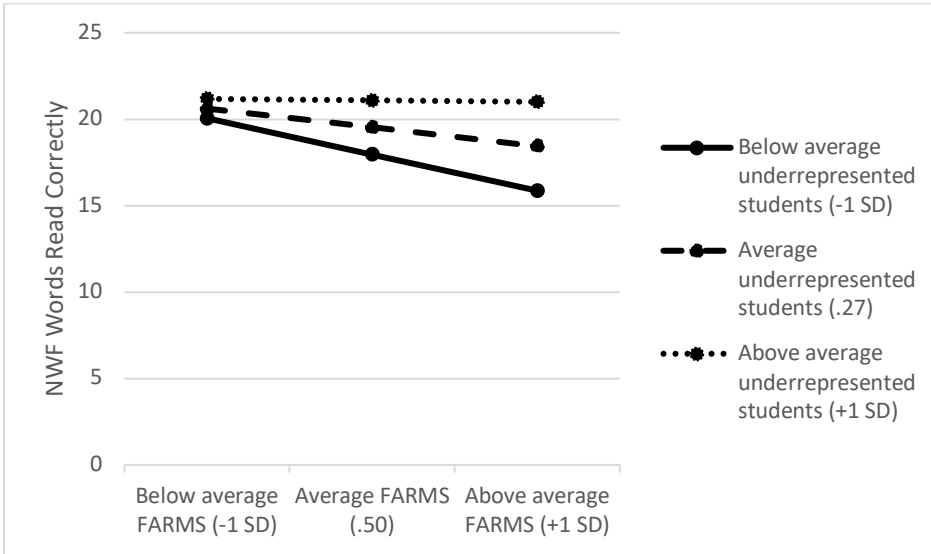

Figure 72

Estimated mean ORF at BOY as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
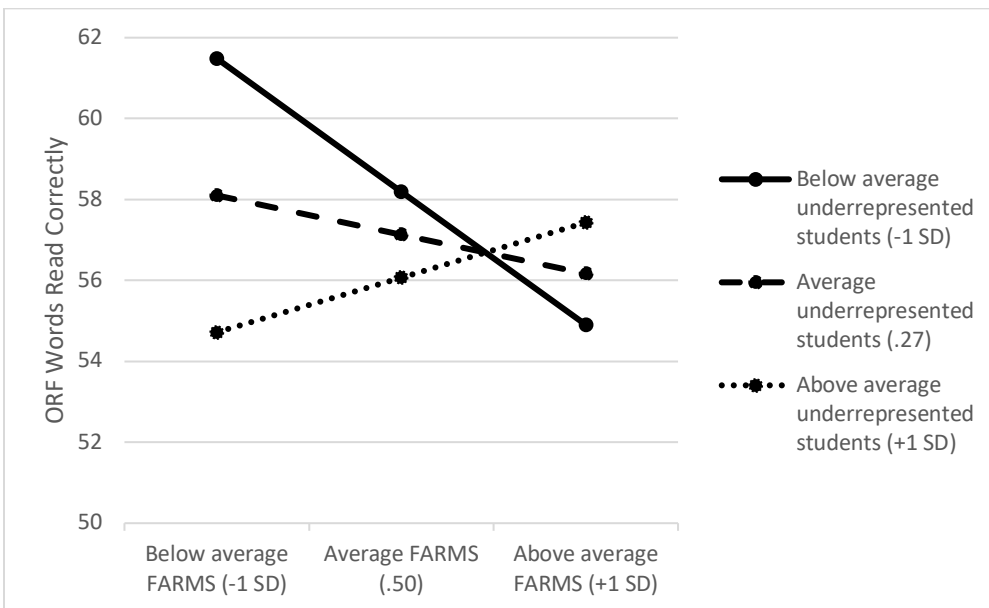
Figure 73

Estimated mean WRF at BOY as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
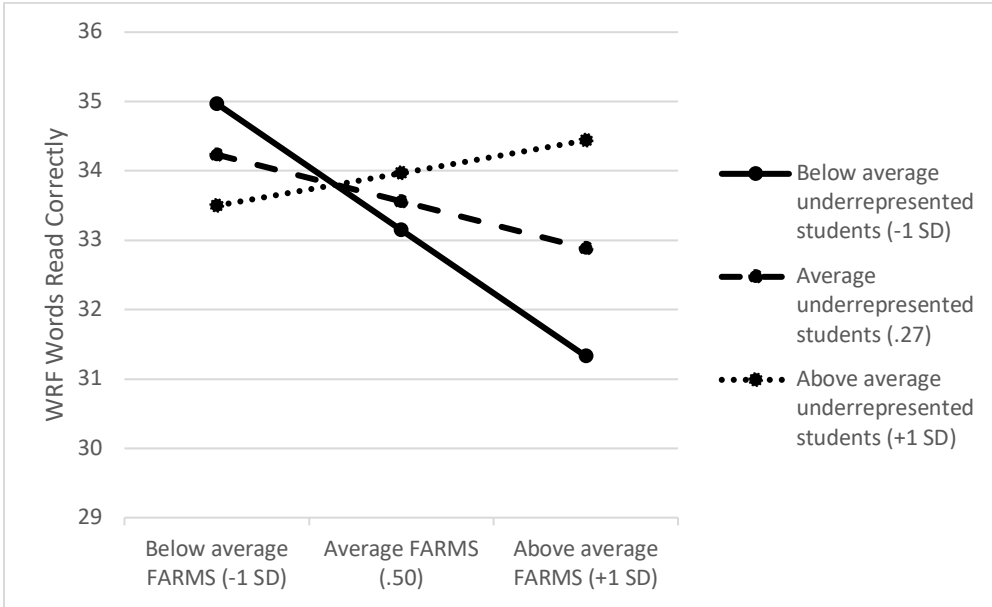

Figure 74

Estimated mean NWF-CLS change as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.

Figure 75

Estimated mean NWF-WRC change as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
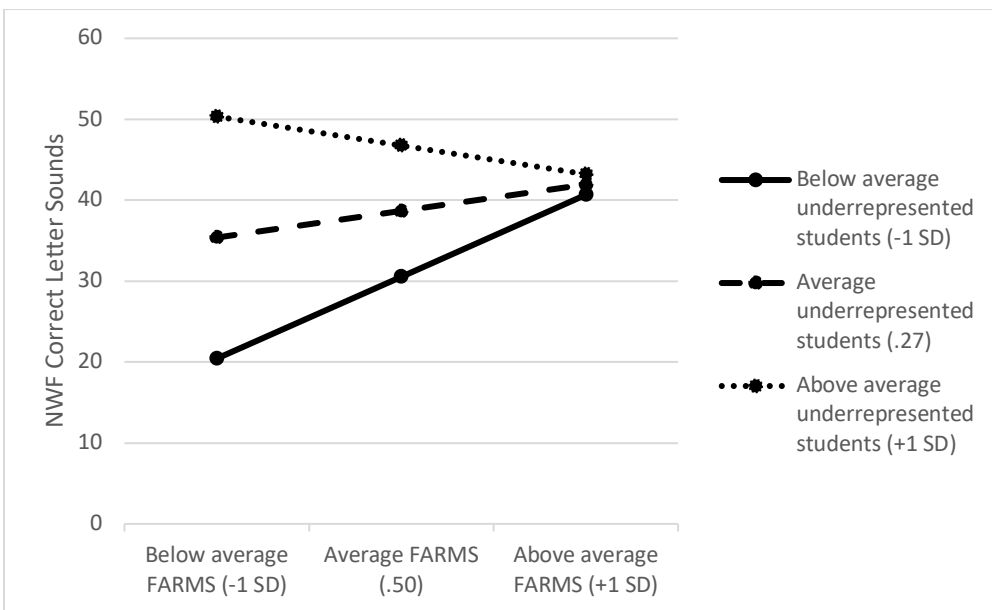

Figure 76

Estimated mean ORF change as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.

Figure 77

Estimated mean WRF change as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
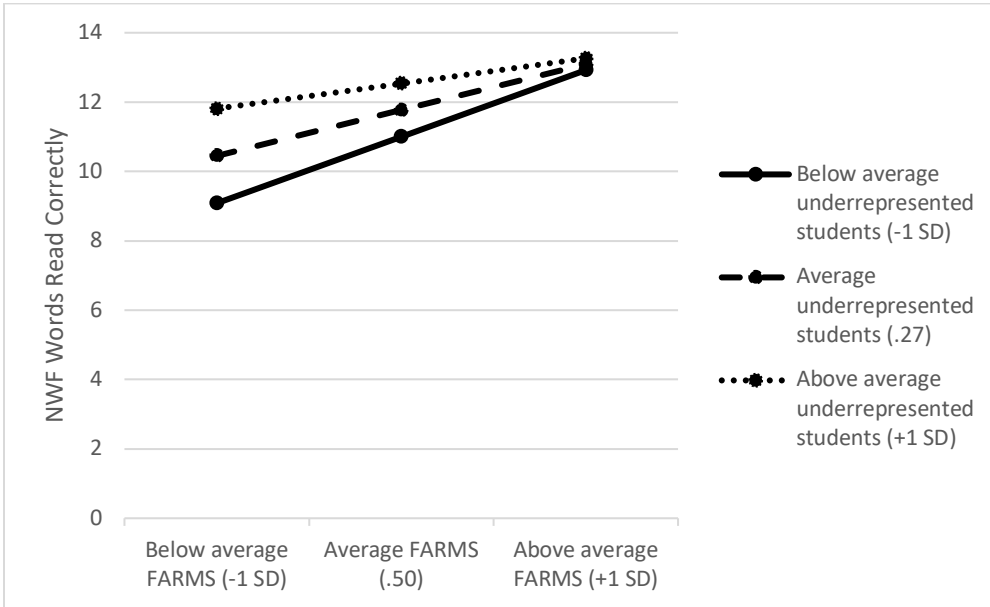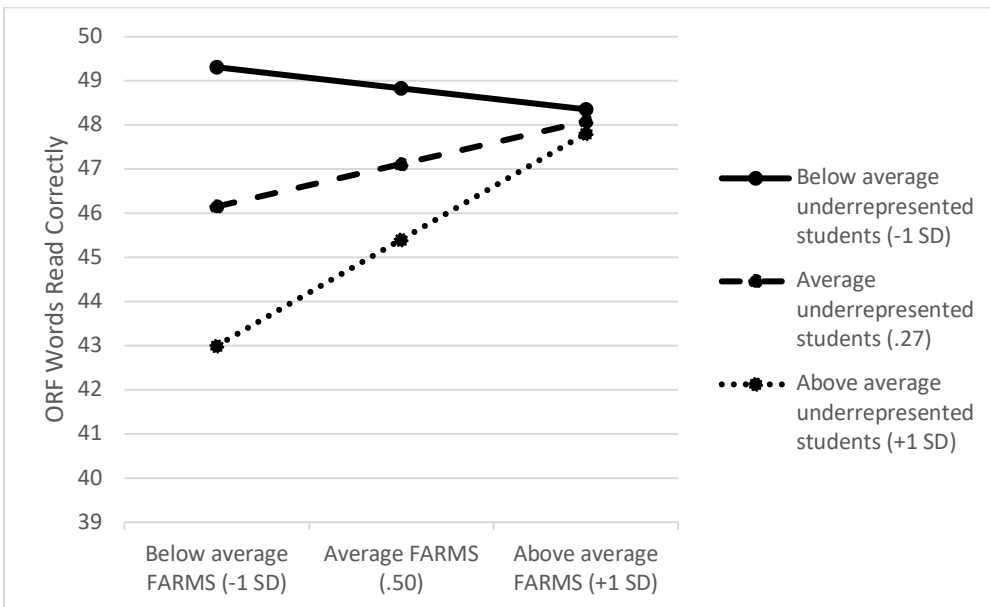

Figure 78

Estimated median NWF-CLS at BOY as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.

Figure 79

Estimated median NWF-WRC at BOY as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
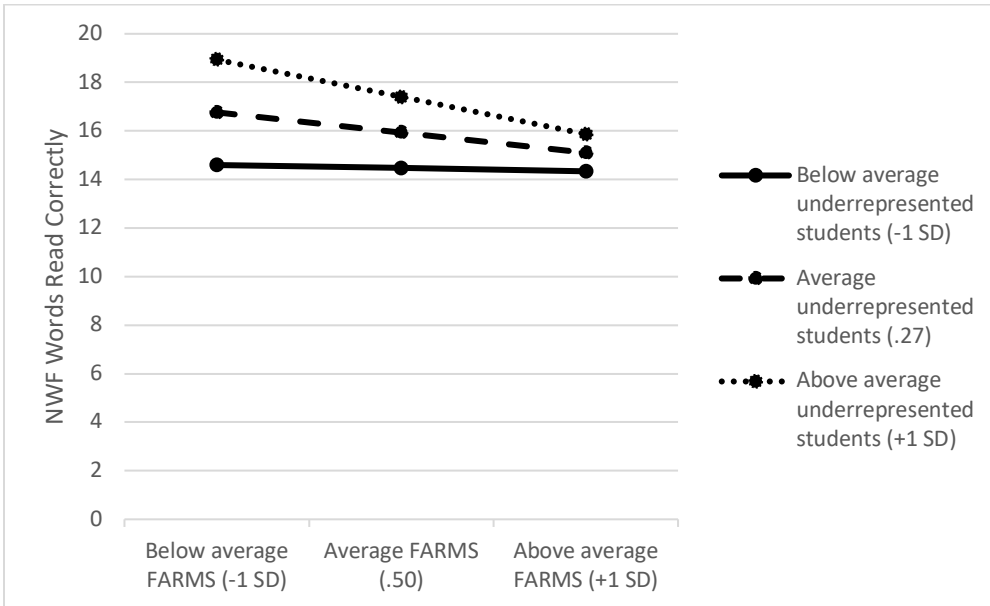
Figure 80

Estimated median ORF at BOY as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
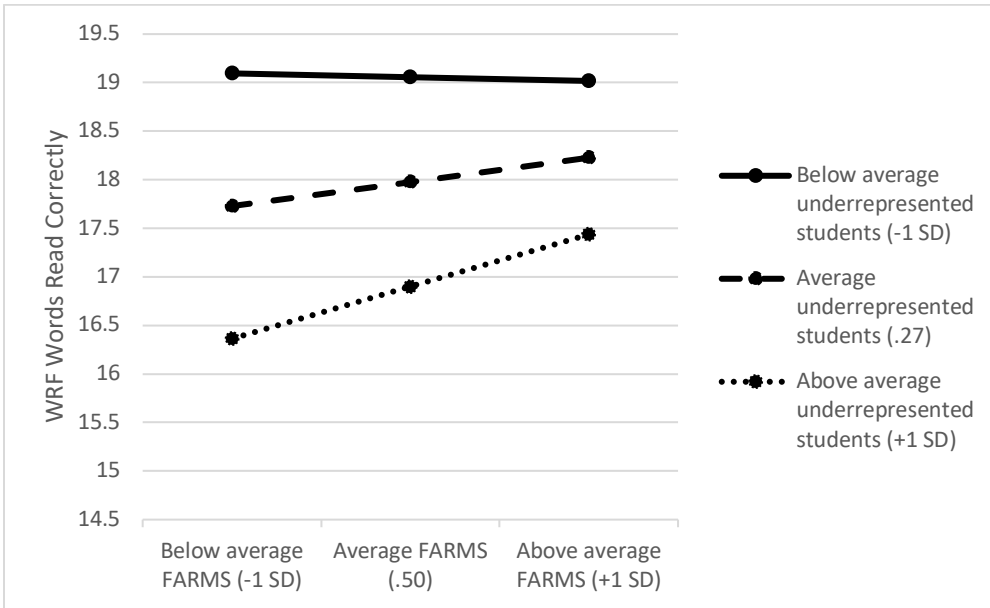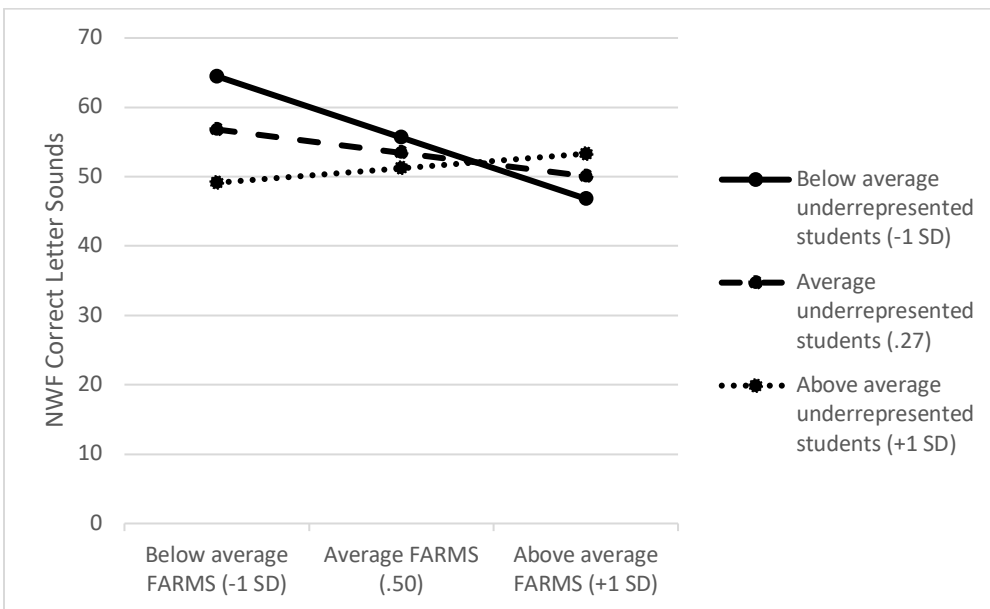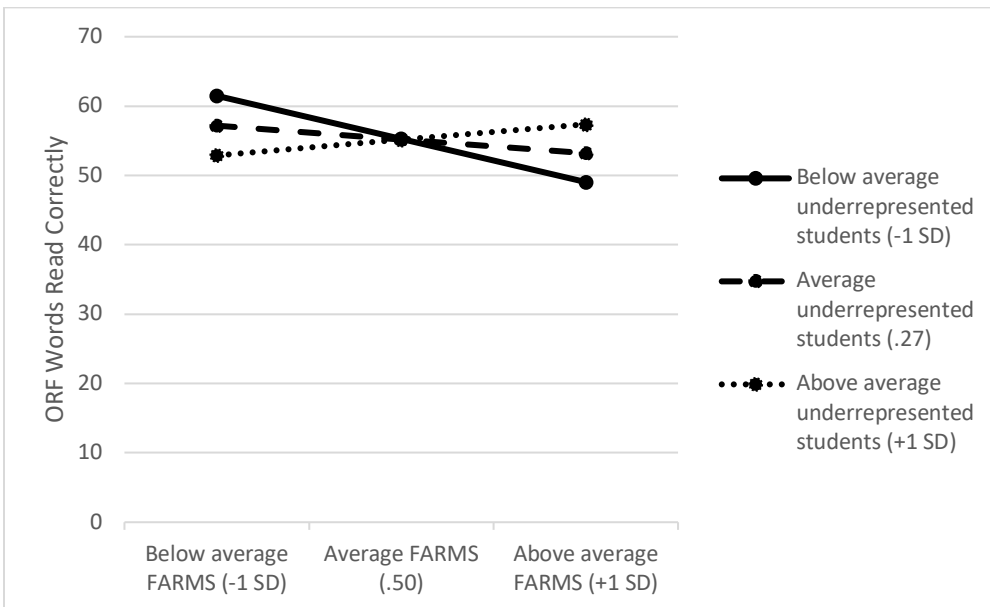
Figure 81

Estimated median WRF at BOY as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.
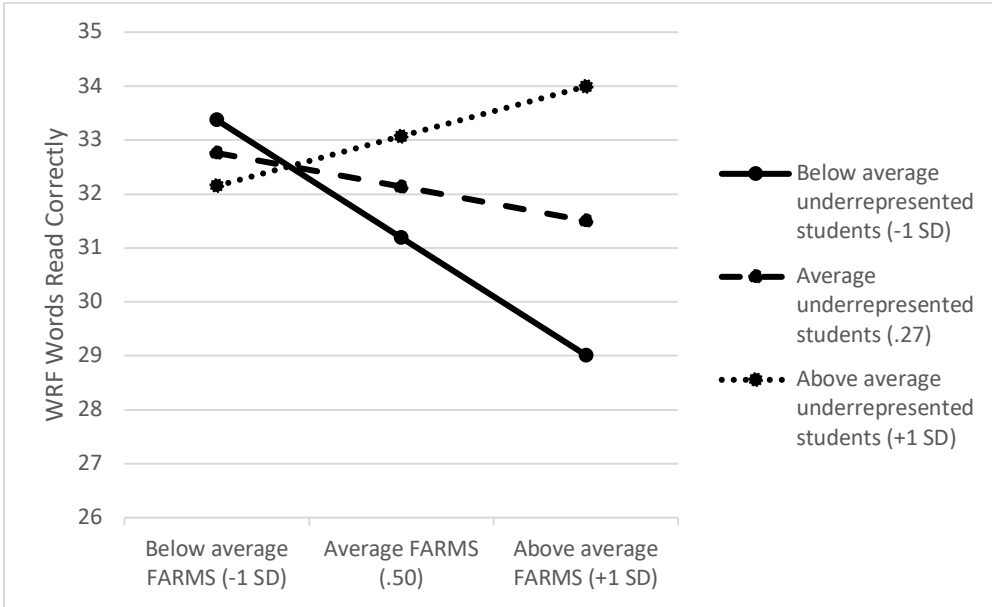
Figure 82

Estimated median NWF-CLS change as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.

Figure 83

Estimated median NWF-WRC change as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.

Figure 84

Estimated median ORF change as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.

Figure 85

Estimated median WRF change as a function of school proportion of students from underrepresented groups and receiving free and reduced-price meals.

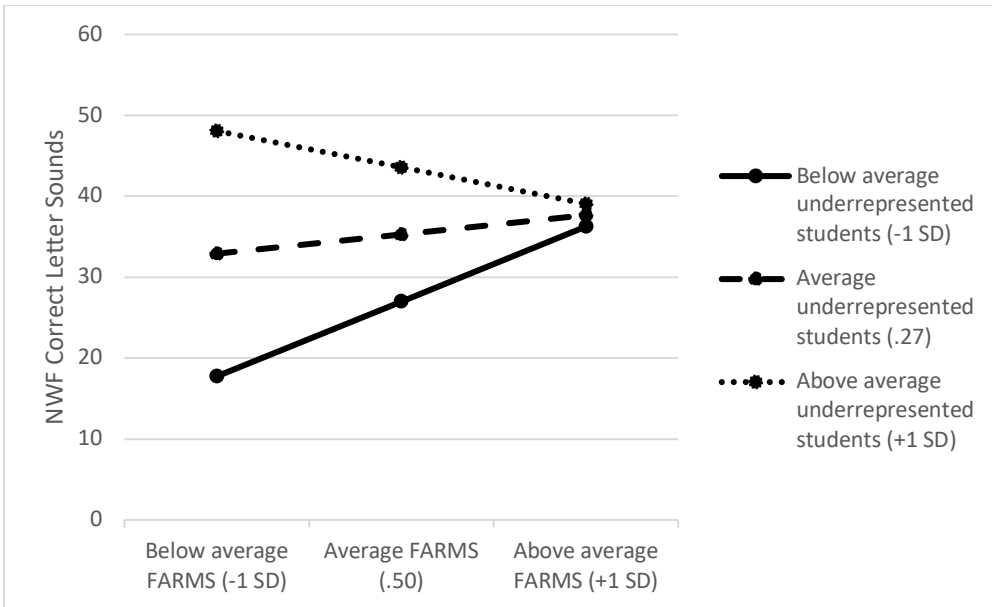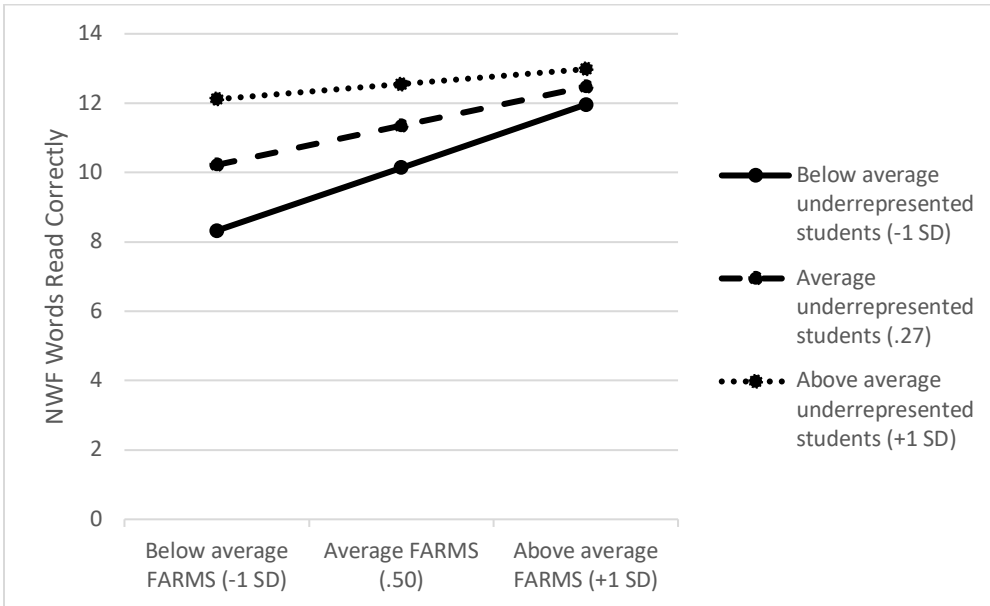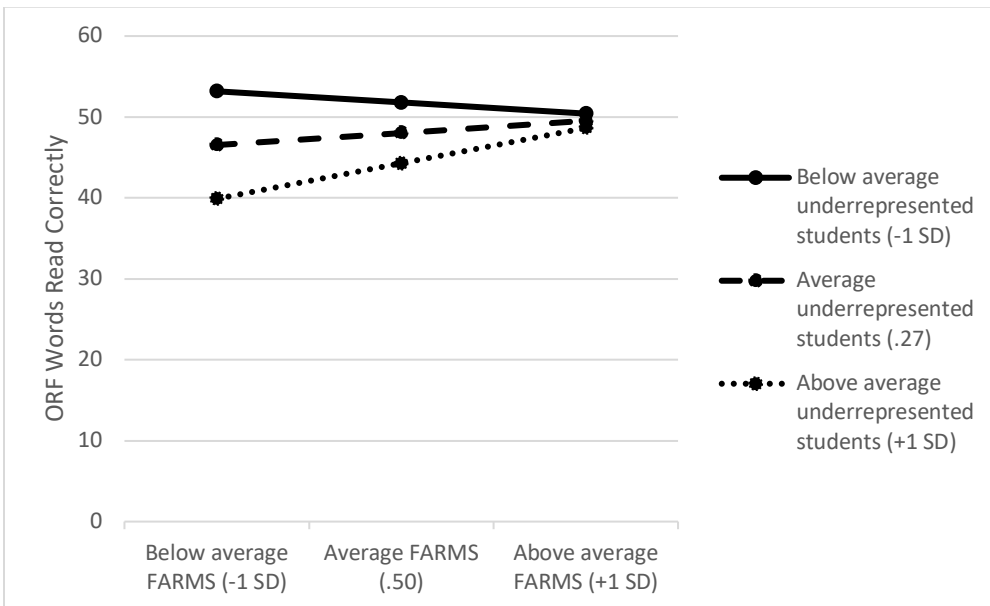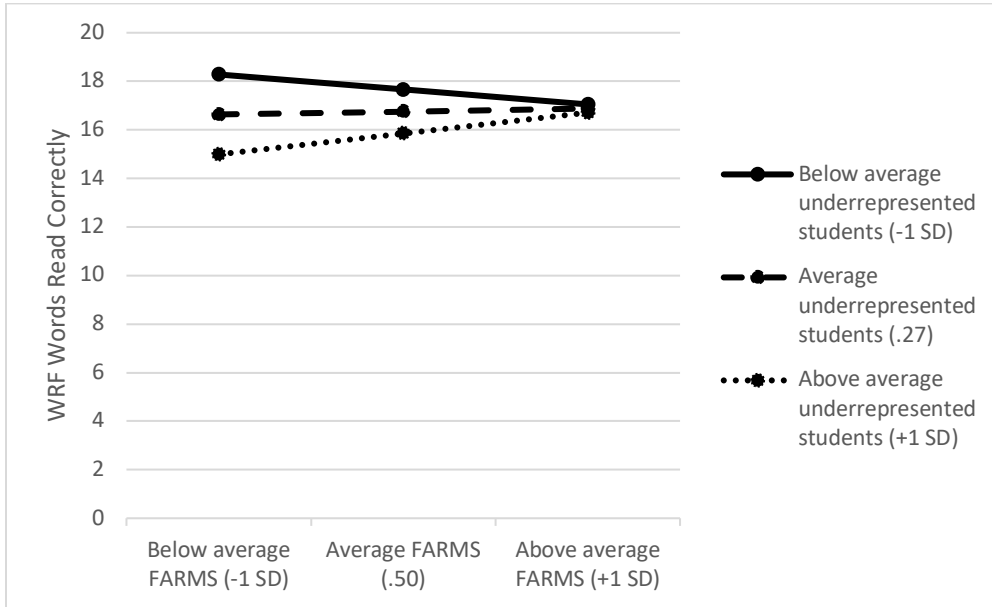# REFERENCES CITED

Alexander, K.L. Entwisle D.R. & Olson L.S. (2007). Lasting consequences of the summer learning gap. American Sociological Review, 72(2), 167- 180

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Beck, C. T. (2008). Reading Fluency as a Predictor of Reading Proficiency in Low-Performing, High-Poverty Schools. *School Psychology Review*, *37*(1), 18–37.

Bast, J., & Reitsma, P. (1997). Mathew Effects in Reading: A Comparison of Latent Growth Curve Models and Simplex Models with Structured Means. *Mulitvariate Behavioral Research*, *32*(2), 135–167.

Bast, J., & Reitsma, P. (1998). Analyzing the Development of Individual Differences in Terms of Matthew Effects in Reading: Results From a Dutch Longitudinal Study. *Developmental Psychology*, *34*(5), 1373–1399.

Bauman, K. (2017). *School Enrollment of the Hispanic Population: Two Decades of Growth*. The United States Census Bureau. https://www.census.gov/newsroom/blogs/random- samplings/2017/08/school_enrollmentof.html

Baumert, J., Nagy, G., & Lehmann, R. (2012). Cumulative Advantages and the Emergence of Social and Ethnic Inequality: Matthew Effects in Reading and Mathematics Development Within Elementary Schools?: Cumulative Advantages in Reading and Math? *Child Development*, *83*(4), 1347–1367. https://doi.org/10.1111/j.1467- 8624.2012.01779.x

Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why poor children are more likely to become poor readers: The school years. *Australian Journal of Education*, *57*(3), 190–213. https://doi.org/10.1177/0004944113495500

Butler, S. R., Marsh, H. W., Sheppard, M. J., & Sheppard, J. L. (1985). Seven-Year Longitudinal Study of the Early Prediction of Reading Achievement. *Journal of Educational Psychology*, *77*(3), 349–361.

Cain, K., & Oakhill, J. (2011). Matthew Effects in Young Readers: Reading Comprehension and Reading Experience Aid Vocabulary Development. *Journal of Learning Disabilities*, *44*(5), 431–443. https://doi.org/10.1177/0022219411410042

Celano, D. and S.B.Neuman (2009). When Schools Close, the Knowledge Gap Grows. Phi Delta Kappan. Vol. 90, No. 04, December 2008. Pp. 256- 262.

Center on Teaching and Learning. (2020a). *UO DIBELS Data System: About Us*.
https://dibels.uoregon.edu/about#history

Center on Teaching and Learning. (2020b). *UO DIBELS Data System Features*.
https://dibels.uoregon.edu/features/

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early
reading achievement: Evidence from the Early Childhood Longitudinal Study
(ECLS) Kindergarten to first grade sample. *Journal of Educational Psychology*,
98(3), 489–507. https://doi.org/10.1037/0022-0663.98.3.489

Chmielewski, A. K., & Reardon, S. F. (2016). Patterns of Cross-National Variation in the
Association Between Income and Academic Achievement. *AERA Open*, *2*(3),
233285841664959. https://doi.org/10.1177/2332858416649593

Common Core of Data (CCD) (2019). ED Public Data Frequently Asked Questions.
National Center for Education Statistics (NCES).
https://nces.ed.gov/ccd/doc/ED_Public_Data_FAQs.docx

Cunningham, A., & Stanovich, K. E. (1997). Early reading acquisition and its relation to
reading experience and ability 10 years later. *Develomental Psychology*, *33*(6),
934–945.

Cunningham, A., & Stanovich, K. E. (1998). What Reading Does for the Mind. *American
Educator*. https://www.aft.org/sites/default/files/periodicals/cunningham.pdf

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using Curriculum-based
Measurement to Establish Growth Standards for Students with Learning
Disabilities. *School Psychology Review*, *30*(4), 19.

DIBELS. (2019a). *8th Edition of the Dynamic Indicators of Basic Early Literacy Skills
(DIBELS): Administration and Scoring Guide*. https://dibels.uoregon.edu

DIBELS-8 Technical Manuel. (2020). *DIBELS-8 Technical Manuel*. Center on Teaching and
Learning; University of Oregon.

Fiester, L. (2013). *Early Warning Confirmed: A research update on 3rd grade reading* (p.
35). Annie E. Casey Foundation.

Fontenelle-Tereshchuk, D. Mental Health and the COVID-19 Crisis: The Hopes and
Concerns for Children as Schools Re-open. *Interchange* 52**,** 1–16 (2021).
https://doi.org/10.1007/s10780-020-09413-1

Foster, W. A., & Miller, M. (2007). Development of the Literacy Achievement Gap: A Longitudinal Study of Kindergarten Through Third Grade. *Language, Speech, and Hearing Services in Schools*, *38*(3), 173–181. https://doi.org/10.1044/0161-1461(2007/018)

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, *88*, 3–77.

Frey, B. B. (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc. https://doi.org/10.4135/9781506326139

Fuchs, L. S., & Fuchs, D. (1993). Formative Evaluation of Academic Progress: How Much Growth Can We Expect? *School Psychology Review*, *22*(1).

Fuchs, L. S., & Fuchs, D. (2011). Using CBM for Progress Monitoring in Reading. National Center on Student Progress Monitoring. Retrieved from: https://files.eric.ed.gov/fulltext/ED519252.pdf

Garcia & Weiss (2020). COVID-19 and student performance, equity, and U.S. education policy. Economic Policy Institute. Retrieved from: https://www.epi.org/publication/the-consequences-of-the-covid-19-pandemic-for-education-performance-and-equity-in-the-united-states-what-can-we-learn-from-pre-pandemic-research-to-inform-relief-recovery-and-rebuilding/

Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., and Tilly, W.D. (2008). Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide. (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Goffreda, C. T., Diperna, J. C., & Pedersen, J. A. (2009). Preventive screening for early readers: Predictive validity of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). *Psychology in the Schools*, *46*(6), 539–552. https://doi.org/10.1002/pits.20396

Good, R., Simmons, D., & Smith, S. (1998). Effective Academic Interventions in the United States Evaluating and Enhancing the Acquisition of Early Reading Skills. *School Psychology Review*, *21*(1), 45–56.

Gottardo, A., Chiappe, P., Siegel, L. S., & Stanovich, K. E. (1999). Patterns of word and nonword processing in skilled and less-skilled readers. *Reading and Writing*, *11*, 465–487.

Graney, S. B., Missall, K. N., Martínez, R. S., & Bergstrom, M. (2009). A preliminary investigation of within-year growth patterns in reading and mathematics curriculum-based measures. *Journal of School Psychology*, *47*(2), 121–142. https://doi.org/10.1016/j.jsp.2008.12.001

Hanushek, & Rivkin. (2006). Teacher Quality. In *Handbook of the Economics of Education* (Vol. 2, pp. 1052–1078).

Hasbrouck, J., & Tindal, G. (2017). *An update to compiled ORF norms* (Technical Report No. 1702). University of Oregon.

Hemphill, F. C., Vanneman, A., & Rahman, T. (2011). *Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress*. National Center for Education Statistics: National Assessment of Educational Progress. https://nces.ed.gov/nationsreportcard/pubs/studies/2011459.aspx

Holmes-Smith, P. (2006). *School Socioeconomic Density and Its Effect on School Performance* (School Research Evaluation and Measurement Services). New South Wales Department of Education and Training.

Illinois State Board of Education. (n.d.). *Targeted School Assistance Programs*. https://www.isbe.net/Pages/TargetedAssistanceSchoolwideProgram.aspx

Juel, C. (1988). Learning to Read and Write: A Longitudinal Study of 54 Children From First Through Fourth Grades. *Journal of Educational Psychology, 80*(4), 437–447.

Kaminski, R., Cummings, K. D., Powell-Smith, K. A., & Good. (2008). Best Practices in Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) for Formative Assessment and Evaluation. In *Best practices in school psychology V* (Vol. 4, p. 25). National Association of School Psychologists.

Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology*, *54*(1), 1–20. https://doi.org/10.1348/000711001159357

Ladson-Billings, G. (2006). From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools. *Educational Researcher*, *35*(7), 3–12. https://doi.org/10.3102/0013189X035007003

Logan, J. A. R., & Petscher, Y. (2010). School profiles of at-risk student concentration: Differential growth in oral reading fluency. *Journal of School Psychology*, *48*(2), 163–186. https://doi.org/10.1016/j.jsp.2009.12.002

Lyon, G. R. (1995). Toward a Definition of Dyslexia. *Annals of Dyslexia*, *45*, 1–27.

McNamara, J. K., Scissons, M., & Dahleu, J. (2005). A Longitudinal Study of Early Identification Markers for Children At-Risk for Reading Disabilities: The Matthew Effect and the Challenge of Over Identification. *Reading Improvement*, *42*(2), 80–97.

Moats, L. C. (1999). Teaching reading is rocket science: What expert teachers of reading should know and be able to do. *American Federation of Teachers*.

Morgan, L., Parkas, G., & Hibel, J. (2008). Matthew Effects for Whom? *Learning Disability Quarterly*, *31*(4), 187–198.

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten Children's Growth Trajectories in Reading and Mathematics. *Journal of Learning Disabilities*, *44*(5), 472–488.

Murrar, & Brauer. (2018). *Mixed Model Analysis of Variance*. SAGE Encyclopedia of Education Research, Measurement, and Evaluation.

National Center for Educational Statistics. (n.d.). *NCES Locale Classifications and Criteria*. U.S. Department of Education. https://nces.ed.gov/programs/edge/docs/LOCALE_CLASSIFICATIONS.pdf

National Center for Educational Statistics. (2015). *School Composition and the Black-White Achievement Gap*. U.S Department of Education. https://nces.ed.gov/nationsreportcard/pubs/studies/2015018.aspx

National Center for Educational Statistics. (2019c). *Table 214.10. Number of public school districts and public and private elementary and secondary schools: Selected years, 1869-70 through 2017-18.* U.S. Department of Education. https://nces.ed.gov/programs/digest/d19/tables/dt19_214.10.asp

National Center for Educational Statistics. (2019a). *Digest of Education Statistics: Table 216.30. Number and percentage distribution of public elementary and secondary students and schools, by traditional or charter school status and selected characteristics: Selected years, 1999-2000 through 2017-18.* U.S. Department of Education. https://nces.ed.gov/programs/digest/d19/tables/dt19_216.30.asp

National Center for Educational Statistics. (2019b). *NAEP Report Card: Reading*. U.S Department of Education. https://www.nationsreportcard.gov/reading/nation/groups/?grade=4

National Center for Educational Statistics. (2020a). *Racial/Ethnic Enrollment in Public Schools*. U.S Department of Education. https://nces.ed.gov/programs/coe/indicator_cge.asp

National Center for Educational Statistics. (2020b). Public Charter School Enrollment. U.S. Department of Education. https://nces.ed.gov/programs/coe/indicator_cgb.asp

National Reading Panel. (2001). *Report of the National Reading Panel: Teaching Children to Read | NICHD – Eunice Kennedy Shriver National Institute of Child Health and Human Development*. https://www.nichd.nih.gov/publications/pubs/nrp/smallbook

Northrop, L. (2017). Breaking the Cycle: Cumulative Disadvantage in Literacy. *Reading Research Quarterly*, *52*(4), 391–396.

Organisation for Economic Cooperation and Development. (2010). *PISA 2009 results: Overcoming social background – equity in learning opportunities and outcomes. (Volume II)*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264091504-en.

Palardy, G. (2008). Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, *19*(1), 21–49.

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of Individual Differences in Reading: Results From Longitudinal Studies in English and Finnish. *Journal of Educational Psychology*, *97*(3), 21.

Pfost, M., Dörfler, T., & Artelt, C. (2012). Reading competence development of poor readers in a German elementary school sample: An empirical examination of the Matthew effect model: reading competence development. *Journal of Research in Reading*, *35*(4), 411–426. https://doi.org/10.1111/j.1467-9817.2010.01478.x

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual Differences in Reading Development. *American Educational Research Journal*, *84*(2), 203–244.

Phillips, L. M., Norris, S. P., Maynard, A. M., & Osmond, W. C. (2002). Relative Reading Achievement: A Longitudinal Study of 187 Children From First Through Sixth Grades. *Journal of Educational Psychology*, *94*(1), 3–13.

Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2011). Matthew Effects in Reading Comprehension. *Journal of Learning Disabilities*, *44*(5), 402–420.

Rathburn, A., West, J., & Germino-Hausken, E. (2004). *From Kindergarten Through Third Grade: Children's Beginning School Experiences* (p. 85). U.S Department of Education Statistics: Institute of Education Sciences.

Reardon, Weathers, Fahle, Jang, & Kalogrides. (2019). *Is Separate Still Unequal? New Evidence on School Segregation and Racial Academic Achievement Gaps*. Stanford Center for Education Policy Analysis.

Reschly, A. L. (2010). Reading and School Completion: Critical Connections and Matthew Effects. *Reading and Writing Quarterly*, *26*(1), 67–90.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-Based Measurement Oral Reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, *47*(6), 427–469. https://doi.org/10.1016/j.jsp.2009.07.001

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*, *42*(4), 546–567. https://doi.org/10.1598/RRQ.42.4.5

Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. Teachers College Record, 107(9), 1999–2005.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, *48*.

Scarborough, H. S., & Parker, J. D. (2003). Matthew effects in children with learning disabilities: Development of reading, IQ, and psychosocial problems from grade 2 to grade 8. *Annals of Dyslexia*, *53*(1), 47–71. https://doi.org/10.1007/s11881-003-0004-6

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew Effect for IQ but Not for Reading: Results from a Longitudinal Study. *Reading Research Quarterly*, *30*(4), 894. https://doi.org/10.2307/748203

Silberglitt, B., & Hintze, J. M. (2007). How Much Growth Can We Expect? A Conditional Analysis of R—CBM Growth Rates by Level of Performance. *Exceptional Children*, *74*(1), 71–84. https://doi.org/10.1177/001440290707400104

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children* (Committee on the Prevention of Reading Difficulties in Young Children, Ed.). National Academy Press.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–407.

Stanovich, K. E. (1996). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–407.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. Guilford Press.

Stanovich, K. E., Cunningham, A., & Feeman, D. (1984). Relation between early reading acquisition and word decoding with and without context_ A longitudinal study of first-grade children_.pdf. *Journal of Educational Psychology*, *76*(4), 668–677.

Stanovich, K. E., Nathan, R. G., & Zolman, J. E. (1998). The Developmental Lag Hypothesis in Reading: Longitudinal and Matched Reading-Level Comparisons. *Child Development*, *59*(1), 71–86.

Stanton-Chapman, T., Chapman, D., & Scott, K. (2001). Identification of Early Risk Factors for Learning Disabilities. *Journal of Early Intervention*, *24*(3), 193–206.

Stiefel, Schwartz, & Ellen. (2006). Disentangling the racial test score gap: Probing the evidence in a large urban. *Journal of Policy Analysis and Management*, *26*(1).

Thomson, & De Bortolli. (2010). *Contextual factors that influence the achievement of Australia's Indigenous students* (Contextual Factors That Influence the Achievement of Australia's Indigenous Students: Results from PISA 2000–2006.). Australian Council for Educational Research (ACER).

Tindal, G., Nese, J. F. T., Stevens, J. J., & Alonzo, J. (2016). Growth on Oral Reading Fluency Measures as a Function of Special Education and Measurement Sufficiency. *Remedial and Special Education*, *37*(1), 28–40. https://doi.org/10.1177/0741932515590234

U.S. Department of Agriculture (2020). Child Nutrition Programs: National School Lunch Program. Retrieved from https://www.ers.usda.gov/topics/food-nutrition-assistance/child-nutrition-programs/national-school-lunch-program/

U.S. Department of Agriculture (2021). Press Release: USDA Issues Pandemic Flexibilities for Schools and Day Care Facilities through June 2022 to Support Safe Reopening and Healthy, Nutritious Meals. Retrieved from: https://www.usda.gov/media/press-releases/2021/04/20/usda-issues-pandemic-flexibilities-schools-and-day-care-facilities

U.S. Department of Education. (2016). *Supporting School Reform By Leveraging Federal Funds in a Schoolwide Program*. https://www2.ed.gov/policy/elsec/leg/essa/essaswpguidance9192016.pdf

U.S. Department of Education. (2018). *Improving Basic Programs Operated by Local Educational Agencies (Title I, Part A)*. https://www2.ed.gov/programs/titleiparta/index.html

Walburg, & Tsai. (1983). Matthew Effects in Education. *American Educational Research Journal*, *20*(3), 359–373.

Yeo, S., Fearrington, J., & Christ, T. J. (2011). An investigation of gender, income, and special education status bias on curriculum-based measurement slope in reading. *School Psychology Quarterly*, *26*(2), 119–130. https://doi.org/10.1037/a0023021