



# International Journal of Educational Methodology

Volume 8, Issue 2, 259 - 273.

ISSN: 2469-9632

<http://www.ijem.com/>

## Reading Trajectories in Elementary Grades: A Longitudinal Analysis

**Malkeet Singh\***   
Meta Inc, USA

**Hugh Dunn**  
University of Hawaii at Manoa, USA

**Hella Bel Hadj Amor**  
Education Northwest, USA

*Received: January 19, 2022 • Revised: March 2, 2022 • Accepted: April 27, 2022*

**Abstract:** Research shows that children's reading proficiency levels in the early grades positively correlate with students' future academic achievement. This study provides the first-ever analysis of reading achievement trajectories for a cohort of students in grades 3 to 5 in 2014–17 in Hawaii schools. Hawaii serves a diverse student population whose characteristics differ in ways often overlooked by standard US racial and ethnic classifications. Our analysis shows that Native Hawaiian and Pacific Islander students not only start at a lower reading proficiency than their peers in Grade 3, but the achievement gap widens as they move from Grade 3 to Grade 5. Moreover, we find a strong association between students' third-grade performance and reading achievement growth rate above and beyond all other factors in our longitudinal model. The difference in performance patterns between student subgroups across the elementary grades can serve as an accurate baseline for yearly monitoring. In light of our findings, we discuss implications for policy and practice.

**Keywords:** *Elementary grades, growth modeling, reading trajectories, sociodemographic factors.*

**To cite this article:** Singh, M., Dunn, H., & Bel Hadj Amor, H. (2022). Reading trajectories in elementary grades: A longitudinal analysis. *International Journal of Educational Methodology*, 8(2), 259-273. <https://doi.org/10.12973/ijem.8.2.259>

### Introduction

The facility with which we learn, how academically successful we are, the extent to which we behave or misbehave, whether we need special education services, how easily and well we learn English, whether we persist in school, graduate high school, enroll and succeed in college, how much we earn and accomplish professionally are among the multiple outcomes that positively correlate with reading proficiency as early as the third grade (Fiester, 2010; Hein et al., 2013; Hernandez, 2011; Lesnick et al., 2010; Yoshikawa et al., 2013). An achievement gap in Grade 3 is consequently a significant source of concern.

Examining the state standardized assessment results of Native Hawaiian (NH) students and their higher-performing peers reveals a statistically significant and increasing achievement gap starting in Grade 3 (Singh, 2011; Singh et al., 2014). NH students experience circumstances that profoundly hinder their educational, economic, and social progress, resulting in performance that lags behind White students, even those with the same socioeconomic or special education status.

Nevertheless, the cornerstone piece of legislation that governs education in this nation demands "critical protections for America's disadvantaged and high-need students, "teaching" to high academic standards that will prepare them to succeed in college and careers," and "positive change in our lowest-performing schools, where groups of students are not making progress" (Every Student Succeeds Act, 2015).

The global pandemic has exacerbated achievement gaps between traditionally underserved students and their peers. The federal government has earmarked funding to address these issues urgently (Azevedo et al., 2021; Kuhfeld et al., 2021; Office of Elementary & Secondary Education, n.d.).

Boosting the life chances of students who are not achieving reading proficiency at Grade 3 and beyond is imperative. Researchers can contribute by building evidence of the effectiveness of strategies and interventions for teaching children how to read (Foorman et al., 2016; Gersten et al., 2008; National Reading Panel, 2000; Shanahan et al., 2010). They can also identify and apply methods that track student achievement trajectories' longitudinally to identify disadvantages, inform instruction, and improve academic outcomes for all students. Identifying disadvantages "requires moving beyond

\* **Corresponding author:**

Malkeet Singh, Meta Inc, USA. ✉ [jmalkeet@gmail.com](mailto:jmalkeet@gmail.com)



average-based differences between groups and implementing fine-grained comparative analyses of how students traverse through the achievement distribution over time" (Quintana & Correnti, 2020, p. 1630). Cross-sectional comparisons alone cannot adequately explain differences in student performance and growth over time and the contribution of the resources available to them (Pfof et al., 2014).

This paper demonstrates how school districts can effectively and realistically use state-level longitudinal data to understand how student academic growth trajectories differ by demographic characteristics and initial performance (Morgan et al., 2011). Districts can leverage these analyses to identify student challenges and their underlying root causes and then select appropriate evidence-based strategies to meet the needs of every student and close achievement gaps. Investigating children's early educational experience when they are "maximally sensitive to home, and school influences" is an opportunity for timely intervention, and students' "cognitive growth rates (are) higher in the first few grades than they are later on" (Entwisle & Alexander, 1992, p. 73).

### *Research Questions*

Our study answers two research questions. First, how do students' reading trajectories vary during the elementary school years? Second, how are growth rates related to student demographic characteristics and their initial reading status at Grade 3? The findings will equip practitioners with a research-based method to support tailored programmatic decisions based on longitudinal, student-level reading performance.

1. What is the difference in the average trajectory of actual change associated with student demographic characteristics in Hawaii's elementary grades?
2. Accounting for student demographic characteristics, is there a relationship between a student's initial status (Grade 3 performance) and growth rate?

## **Literature Review**

### *Theoretical Framework*

Research suggests differences in children's reading growth development account for individual cognitive development variation (Lerkkanen et al., 2004). In particular, Pfof et al. (2014) synthesized 25 years of research that collectively suggests three general developmental patterns characterizing inter-individual differences in children's reading development. They refer to the first developmental pattern as the cumulative model that illustrates increasing differences in student performance across time (Bast & Reitsma, 1998; Bodovski & Farkas, 2007; Kempe et al., 2011; Leppänen et al., 2004; Onatsu-Arivilommi & Nurmi, 2000; Pfof et al., 2014; Stanovich, 2009; Walberg & Tsai, 1983; Williamson et al., 1991). Termed the Matthew effect (Walberg & Tsai, 1983), this trend in children's reading performance over time characterizes an increasingly wider gap between advantaged and disadvantaged children (Morgan et al., 2008; Salaschek et al., 2014). In the context of reading outcomes, "over time, better readers get even better, and poorer readers become relatively poorer" (Bast & Reitsma, 1998, p. 1373).

Beyond observing a positive correlation between children's initial and ensuing performance, Rigney (2010) posited that the Matthew effect could be absolute or relative. An absolute Matthew effect reveals developmental patterns demonstrating high levels of student performance at the outset increase over time. In contrast, initially, low levels of student performance will decrease over time (Rigney, 2010). For example, a longitudinal study follows students from Grade 1 to Grade 5; Entwisle et al. (1997) determined that an achievement disparity between disadvantaged and advantaged students can increase from less than a year in Grade 1 to nearly three years at the end of Grade 5. Likewise, a relative Matthew effect assumes that students who demonstrate high initial reading skills will improve over time. However, it also suggests that students who demonstrate low initial reading skills will make little or no gains (Rigney, 2010).

The second type of developmental pattern refers to either the compensatory effect (Aarnoutse & van Leeuwe, 2000; Parrila et al., 2005; Phillips et al., 2002) or the developmental-lag model (Francis et al., 1996; Parrila et al., 2005; Rourke, 1976). This model argues that children who enter school with lower reading proficiency levels demonstrate faster growth rates than students who start school with higher reading proficiency levels (Morgan et al., 2011; Salaschek et al., 2014). Hence, students' initial inter-individual differences will decrease over time rather than broaden or remain constant. Namely, students with low pretest results will demonstrate more significant gains over time than their peers who had higher pretest results. The third model suggests that the performance gap and growth rates will remain stable over time among initially low- and high-performing children (Pfof et al., 2014).

### *Closing Achievement Gaps*

Modeling and examining children's reading progress can yield insights that can guide policy and educational practice, especially in closing achievement gaps (Kim et al., 2010; Shanley, 2016). The literature on closing achievement gaps suggests two prevailing approaches. The preventative view assumes that children enter school on a level playing field

and argues that it is incumbent on schools to ensure that disadvantaged children keep pace with their peers (Johnson, 2002).

Based on evidence supporting school readiness practices, the reparative perspective contends that disadvantaged children enter school with lower proficiency than their more advantaged peers. Thus, schools are responsible for providing underprivileged students with the necessary support to accelerate learning and close the learning gaps. (Shin et al., 2013). In either case, the achievement gap between student groups is an enduring issue in American education. Nevertheless, all efforts to close the gaps should be informed by methods designed to monitor students' performance outcomes and longitudinally analyze their developmental trajectories across multiple years.

### *Importance and Relevance*

Transitioning from elementary to middle school is critical in a student's educational pathway. Middle-school teachers expect incoming students to be grade-level proficient in reading and prepared to meet the academic demands of middle school. However, there are racial or ethnic differences in reading performance during the elementary grades (Camera, 2016; Hemphill & Vanneman, 2011; The Nation's Report Card, 2019; Vanneman et al., 2009). These differences are prominent starting from Grade 3 (Singh et al., 2014).

To investigate the growth rates of different groups of students, we can employ longitudinal statistical procedures to explicitly model student performance change over time. While this has occurred for some racial or ethnic groups (Chatterji, 2006; Lesaux et al., 2007; Lubienski & Crane, 2010), no single longitudinal study has investigated the relationships between predictor variables and student growth in Hawaii's public school system. The likely reason is that those previous assessments, such as the Hawaii State Assessment (HSA) I, HSA II, and HSA III, were not vertically linked. Vertically scaled assessments, which relate to student growth over time, require comparative information about results from assessments of different difficulty levels. The challenges of vertical scaling of assessments have existed since using standards-based assessments to measure individual growth (Patz, 2007).

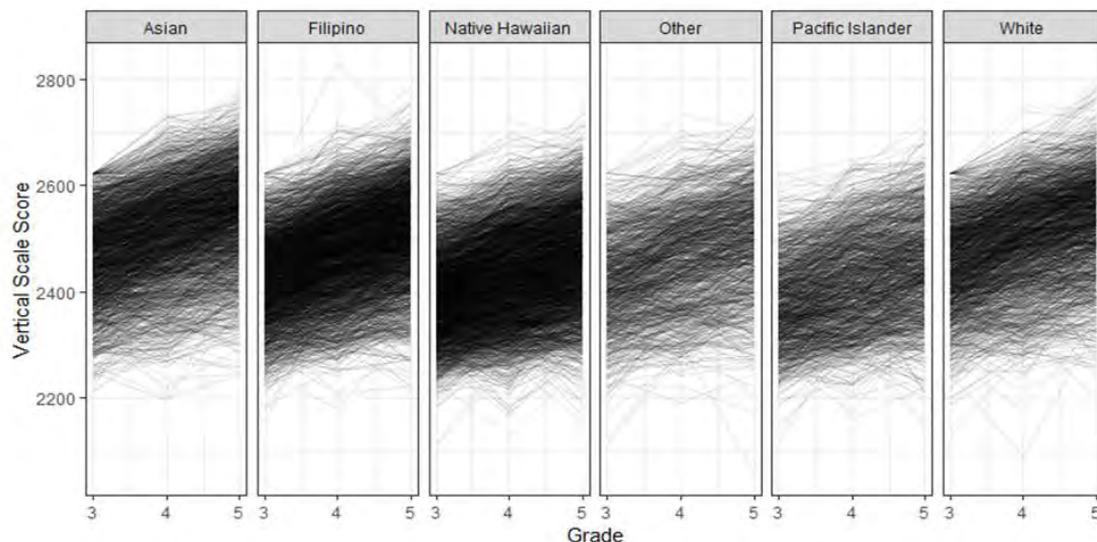
In 2013, the Hawaii Department of Education (HIDOE) implemented the English language arts common core state standards (CCSS) based on explicit specifications for the reading domain's learning continuum. Moreover, newly developed assessments aligned to the cross-grade CCSS. Thus, the new assessments created a research-based reference frame for individuals' and groups' locations on a learning continuum. Therefore, it allowed for the proper use of its test scores as indicators of student growth.

## **Methodology**

### *Research Design*

The new assessment implemented in 2013 allowed us to employ individual growth curve modeling techniques to analyze our longitudinal data with three data waves, corresponding to Grades 3, 4, and 5. The scores range from 2,000 to 3,000 spanning sufficient content to indicate increasing performance effectively. We kept the outcome measure scale the same as Hawaii's state reporting protocol.

We followed Singer and Willett's (2003) recommendation for a linear-change individual growth model within a multilevel framework with three data waves. A visual inspection of the data also supports our choice of a linear-change individual growth model as the population model most likely generates our sample data (Figure 1).



*Figure 1. Spaghetti plots by ethnicity. There is variation in initial and growth patterns among different racial/ethnic student groups.*

The individual growth model included time-invariant demographic predictor variables: students' gender, male, coded zero if female or one if male; categorized as free or reduced-price lunch (FRPL), an indicator to represent poverty (zero, is not classified as FRPL, one categorized as FRPL in any of the elementary grades); special education status (SPED) (zero, does not have an individualized educational plan (IEP), and one, has an IEP in any of the elementary grades).

We grouped the pupils into six major racial and ethnic groups: Asian (Chinese, Korean, and Japanese students), Filipino, Hawaiian, Pacific Islander (PI) (Micronesian, Samoan, and Tongan students), White, and "Other" (for example, American Indian, Black, Guamanian, Hispanic, Portuguese, and multiple race/ethnicities). Moreover, we used the White student group as a baseline; therefore, we created five dummy-coded indicators for the student groups. In total, there were eight individual-level predictors in the model.

### Sample and Data Collection

The analytic sample is the 2014-15 third-grade cohort of students with reading scores in all three grades, around 85% of the population. In the complete-case analysis, we removed students who did not have scores in Grades 4 or 5 or both. We also removed students who had repeated a grade ( $n = 28$ ), resulting in a sample size of 12 139 students.

### Analyzing of Data

We included the time variable (grade levels). We centered it at Grade 3 (initial status) to explicitly model the within-individual change via the time slope parameter. We included only the time-invariant predictors' effects to explain any changes at the intercept and time slope outcomes. We added cross-level interactions between the first level time predictor and the second level student predictors, which modifies the linear time slope of student predictors.

School fixed effects were included as indicator variables to adjust for any clustering effect. We excluded English language status as a predictor because almost half of the PI students were English language learners. Therefore, adding the variable leads to a multicollinearity issue between the two variables. We used the student growth model where the rate is a function of time ( $Grade_{ij}$ ), and we write the first level model as the following.

#### First-level model

$$Y_{ij} = [\pi_{0i} + \pi_{1i} * (Grade_{ij} - 3)] + [\epsilon_{ij}]$$

$i$  and  $j$  represent pupils and time (i.e., waves).

$i$  goes from values of 1 through 12,139 (the sample size), and  $j$  runs through values from 1 through 3 (for the three waves).  $Y_{ij}$  represents the reading scale score for student  $i$  at time  $j$ . Moreover,  $\pi_{0i}$  and  $\pi_{1i}$  are the individual rate parameters of the  $i^{\text{th}}$  student in the population. By centering the time variable as  $Grade_{ij} - 3$ , the intercept  $\pi_{0i}$  shows student  $i$ 's "true" reading performance at Grade 3.  $\pi_{1i}$  means student  $i$ 's actual annual (or per grade) rate of change in reading score. We model a straight line adequately representing each student's "true" change over grades. We assume that any changes from linearity arise from random error ( $\epsilon_{ij}$ ).

The second-level model represents the differences in student growth parameters across students and their time-invariant characteristics. The second-level model features one equation for each first-level growth parameter, which defines the association between the rate and the other factors as follows:

#### Second-level model

$$\begin{aligned} \pi_{0i} &= \beta_{00} + \beta_{01} \text{Male} + \beta_{02} \text{FRPL} + \beta_{03} \text{SPED} + \beta_{04} \text{Asian} + \beta_{05} \text{Filipino} + \\ &\beta_{06} \text{NH} + \beta_{07} \text{Other} + \beta_{08} \text{PI} + \sum_k \beta_k \text{Sch}_k + r_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11} \text{Male} + \beta_{12} \text{FRPL} + \beta_{13} \text{SPED} + \beta_{14} \text{Asian} + \beta_{15} \text{Filipino} + \\ &\beta_{16} \text{NH} + \beta_{17} \text{Other} + \beta_{18} \text{Islander} + \sum_k \beta_k \text{Sch}_k + r_{1i} \end{aligned}$$

The linear-change individual growth model estimates each student's change trajectory by specifying a rate of change with two components: one constant across students and another allowed to differ across them. Thus, we will enable the intercept and slope to vary among students and covary. The final specified model is as follows:

#### Final model

$$\begin{aligned} Y_{ti} &= \beta_{00} + \beta_{01} \text{Male} + \beta_{02} \text{FRPL} + \beta_{03} \text{SPED} + \beta_{04} \text{Asian} + \beta_{05} \text{Filipino} + \\ &\beta_{06} \text{NH} + \beta_{07} \text{Other} + \beta_{08} \text{PI} + \sum_k \beta_k \text{Sch}_k + \\ &(\beta_{10} + \beta_{11} \text{Male} + \beta_{12} \text{FRPL} + \beta_{13} \text{SPED} + \beta_{14} \text{Asian} + \beta_{15} \text{Filipino} + \\ &\beta_{16} \text{NH} + \beta_{17} \text{Other} + \beta_{18} \text{PI} + \sum_k \beta_k \text{Sch}_k) * (Grade_{ij} - 3) + \\ &r_{0i} + r_{1i} * (Grade_{ij} - 3) + \epsilon_{ij} \end{aligned}$$

We used the free R software and packages (Bates et al., 2015; R Core Team, 2017) for all analyses. We report unbiased estimates of the variance and covariance parameters using the restricted maximum likelihood procedure.

We also explored other specifications by including interactions among the student-level predictors. However, these did not result in convergence or non-statistically significant results. Therefore, we settled on the more economical model. Only the time variable term interacts with the student-level predictors to study the different student groups' differential growth patterns. The model allows for a straightforward interpretation of the results.

### Sensitivity Analyses

For sensitivity tests, we reanalyzed the data in two ways. In the first test, we kept all students with incomplete scores in subsequent waves (Grade 4 or 5) as the multilevel framework allows for incomplete or unbalanced data related to time. In the second one, we imputed missing data using multiple imputations with chained equations. The procedure, developed by Enders et al. (2018), employs a conditional specification and a Bayesian framework to obtain the posterior distributions (Gelfand, 2000). The method allows missing data to be imputed using a specification identical to the final analysis model. We created 20 multiple imputed samples. The results of the numerous imputed samples are then combined to provide unbiased results (Tables 1A).

### Findings / Results

We report descriptive statistics for the analytic sample and present the results disaggregated by racial/ethnic student groups. On average, students' scores will be higher in later grades in vertically scaled assessments than in earlier grades. The variance in scores also becomes slightly more extensive in later grades, as shown in the boxplots below (Figure 2). We also report additional descriptive results in the appendix section for the sample with incomplete waves of data. We report the pooled estimates from the 20 multiple-imputed samples.

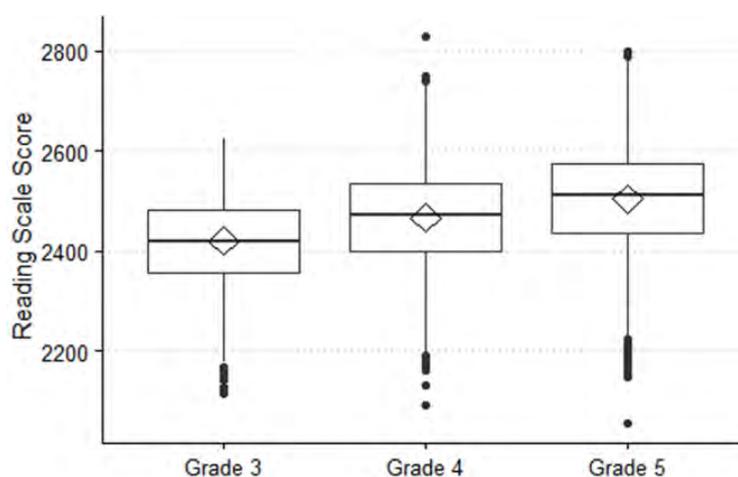


Figure 2. Reading scaled score distribution in grades 3, 4, and 5. As expected with vertically scaled assessments, scores tend to be higher as students move up the grades, thus justifying the growth model's use. The diamond represents the mean score, while the box's midline represents the median.

The disaggregated performance by student group shows that Asian students, on average, scored the highest in all three elementary grades. In contrast, NH and PI students scored the lowest (Table 1). The descriptive results only illustrate the differences among student groups and do not consider confounding factors. We can only ascertain the unique relationship between a predictor and the outcome through multivariate statistical analyses.

Table 1. Descriptive statistics for grades 3 to 5 disaggregated by racial/ethnic groups. As expected with vertically aligned assessments, mean scores for student groups increase from one grade to the next.

| Demographics | N     | Grade 3 |       | Grade 4 |       | Grade 5 |       |
|--------------|-------|---------|-------|---------|-------|---------|-------|
|              |       | M       | SD    | M       | SD    | M       | SD    |
| Asian        | 2099  | 2464    | 83.77 | 2512    | 88.91 | 2550    | 91.17 |
| Filipino     | 2936  | 2424    | 79.45 | 2474    | 85.12 | 2514    | 88.78 |
| NH           | 3135  | 2392    | 79.24 | 2432    | 89.11 | 2467    | 92.95 |
| Other        | 1034  | 2416    | 84.67 | 2463    | 90.46 | 2499    | 96.41 |
| PI           | 1213  | 2376    | 79.08 | 2420    | 87.84 | 2453    | 94.76 |
| White        | 1722  | 2450    | 83.98 | 2498    | 90.16 | 2538    | 92.69 |
| All          | 12139 | 2421    | 86.40 | 2467    | 93.91 | 2504    | 98.19 |

*Growth Modeling Results*

By comparing the model with and without predictors in the level-2 specification, we estimated the residual variance reduction in the random coefficients (intercept and linear-time slope). The specified model predictors account for 43.7% (6184.8 – 3480.51/6184.8) of the residual variance in the random intercepts and 51.1% (159.8 – 78.19/159.8) residual variance in the random time-slope coefficients (Table 2).

The intercept represents the Grade 3 reading scale score of a White female who is not eligible for FRPL and does not have an IEP. This score is estimated to be 2482.15 (95% Confidence Level (CL): 2465.86, 2498.44). On average, this same student is predicted to increase her performance by 53.05 points at each subsequent grade (95% CL: 46.36, 59.74).

Statistically controlling for other factors in the model, male students, on average, had an adjusted score of 16.25 points below their female peers in third grade (95% CL: -18.72, -13.78). As their growth rate in reading lagged behind girls' by 2.35 points, this disadvantage increased as boys progressed to Grades 4 and 5. In other words, the model estimates that, on average, boys underperformed girls by 16.25 points in Grade 3, and by the end of Grade 5, this gap increased to 20.95 points.

With other factors held constant, students eligible for FRPL start with an adjusted score of 32.79 points lower than students who are not eligible (95% CL: -35.72, -29.86). The FRPL-eligible students had an average growth rate of 1.31 points (95% CL: -2.51, -0.11) lower than non-eligible students. FRPL-eligible students would score 33.10 points lower in Grade 4 and 34.41 points lower in Grade 5 than their non-eligible peers, increasing the gap from one grade to the next.

The scores of students with IEPs were estimated to be 92.55 points (95% CL: -96.87, -88.23) lower than their counterparts without IEPs in Grade 3, holding other factors constant. This gap increased in later grades. The growth rate was estimated to be 7.31 points (95% CL: -9.09, -5.54) lower than the gap for students without an IEP leading to a gap expected to widen from 92.55 in third, 99.86 in fourth, and 107.17 in fifth grades.

Keeping other demographic factors constant and using White students as the reference group, the model showed that Asian and White students exhibited no statistically significant differences in their Grade 3 adjusted performance and growth rates. Filipino students, however, had an adjusted score of 14.45 points (95% CL: -18.97, -9.93) lower than their White peers in Grade 3. This gap seems to have remained stable in Grades 4 and 5 (as the difference in growth rates between them and their White peers is not statistically significant).

Statistically, there was no significant difference between the adjusted mean between the Other and White students. The Other category had an adjusted mean of 19.50 lower than White students in Grade 3 (95% CL: -24.92, -14.07), and this gap appeared to widen to 22.00 points in Grade 4 and 24.50 points in Grade 5 as their adjusted rate was 2.50 points (95% CL: -4.73, -0.27) lower than White students. However, their rates were not statistically different (95% CL: -4.16, 0.27).

NH students had an adjusted average score of 30.99 points (95% CL: -35.40, -26.58) lower than their White peers in Grade 3 and a slower growth rate than their White peers' by 5.25 points (95% CL: -7.06, -3.44) per grade/year. Therefore, their gap had increased to 36.24 points in Grade 4 and 41.49 points in Grade 5. PI students' adjusted score was 44.25 points (95% CL: -49.80, -38.70) lower than White students in Grade 3, and because of a lower growth rate of 5.70 points (95% CL: -7.98, -3.42) than their White peers', this gap widened to 50.00 points in Grade 4, and 55.75 points in Grade 5. To illustrate this visually, we plot the adjusted average trends for male students who are not eligible for FRPL or have an IEP for these three racial/ethnic groups (Figure 3).

Table 2. Estimated parameters of the individual growth model

|                  | Coefficient  | Estimate   | 95% Confidence Intervals |         |
|------------------|--------------|------------|--------------------------|---------|
|                  |              |            | Lower                    | Upper   |
| <b>Intercept</b> | $\beta_{00}$ | 2482.15*** | 2465.86                  | 2498.44 |
| <b>Male</b>      | $\beta_{01}$ | -16.25***  | -18.72                   | -13.78  |
| <b>FRPL</b>      | $\beta_{02}$ | -32.79***  | -35.72                   | -29.86  |
| <b>SPED</b>      | $\beta_{03}$ | -92.55***  | -96.87                   | -88.23  |
| <b>Asian</b>     | $\beta_{04}$ | -2.24n.s.  | -7.02                    | 2.55    |
| <b>Filipino</b>  | $\beta_{05}$ | -14.45***  | -18.97                   | -9.93   |
| <b>NH</b>        | $\beta_{06}$ | -30.99***  | -35.40                   | -26.58  |
| <b>Other</b>     | $\beta_{07}$ | -19.50***  | -24.92                   | -14.07  |
| <b>PI</b>        | $\beta_{08}$ | -44.25***  | -49.80                   | -38.70  |

Table 2. Continued

|                                  | Coefficient    | Estimate   | 95% Confidence Intervals |       |
|----------------------------------|----------------|------------|--------------------------|-------|
|                                  |                |            | Lower                    | Upper |
| <b>Time</b>                      | $\beta_{10}$   | 53.05***   | 46.36                    | 59.74 |
| <b>Male x Time</b>               | $\beta_{11}$   | -2.35***   | -3.37                    | -1.34 |
| <b>FRPL x Time</b>               | $\beta_{12}$   | -1.31*     | -2.51                    | -0.11 |
| <b>SPED x Time</b>               | $\beta_{13}$   | -7.31***   | -9.09                    | -5.54 |
| <b>Asian x Time</b>              | $\beta_{14}$   | -0.64n.s.  | -2.60                    | 1.32  |
| <b>Filipino x Time</b>           | $\beta_{15}$   | -0.19n.s.  | -2.04                    | 1.67  |
| <b>NH x Time</b>                 | $\beta_{16}$   | -5.25***   | -7.06                    | -3.44 |
| <b>Other x Time</b>              | $\beta_{17}$   | -2.50*     | -4.73                    | -0.27 |
| <b>PI x Time</b>                 | $\beta_{18}$   | -5.70***   | -7.98                    | -3.42 |
| <b>Intercept Variance</b>        | $\tau^2_{U0}$  | 3480.51*** |                          |       |
| <b>Intercept-Time Covariance</b> | $\tau^2_{U01}$ | 302.98***  |                          |       |
| <b>Time Variance</b>             | $\tau^2_{U1}$  | 78.19***   |                          |       |
| <b>Residual Variance</b>         | $\sigma^2$     | 1415.07*** |                          |       |

Notes: Complete cases only i.e., students with results all three grade levels ( $n = 12, 139$ ). n.s. – not significant,  $p > 0.05$ ; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

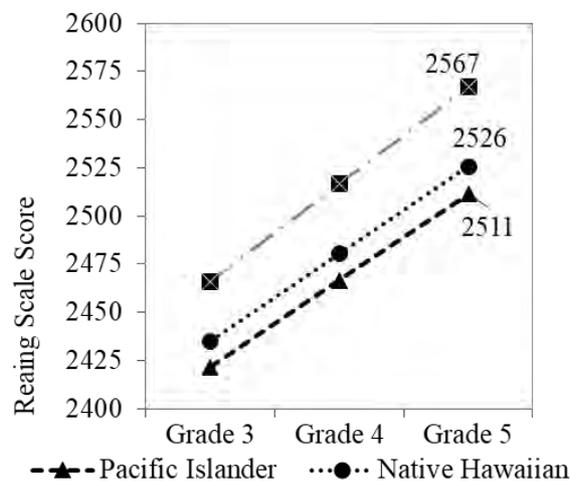


Figure 3. Estimated adjusted trends for White, NH, and PI male students. We expect gaps to become more significant between Whites and NH or PI groups as they progress from third to fifth grades.

To assess the appropriateness of the model, we graphed the residuals against fitted values. The plot showed no obvious pattern of residual heteroscedasticity (Figure 4, left side). The quantile-quantile plot showed mostly a homogeneous residual distribution except at the lower quartile, which may imply a heavier tail than usual (Figure 4, right side). Overall, the residuals do not violate the homoscedasticity and normality assumptions.

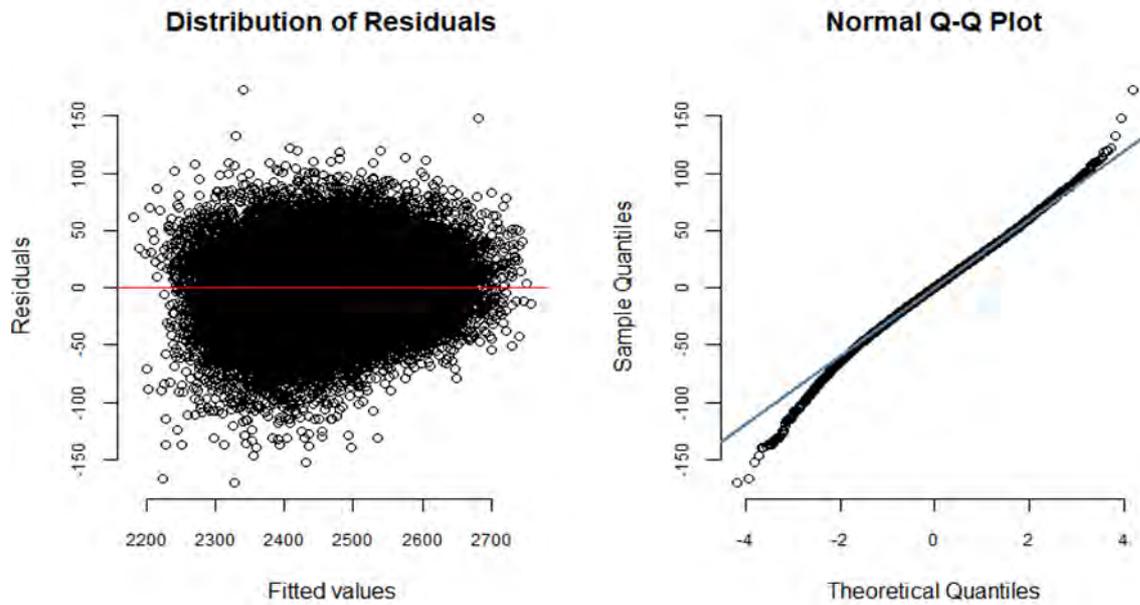


Figure 4. There is no striking pattern and generally gaussian overall except for a slight deviation at the data's tail ends.

Sensitivity Analyses

The sensitivity test analyses show that the whole sample's estimates with incomplete outcome data or multiple imputations were almost identical to those obtained from the complete-case analysis. The different analyses did not affect our conclusion except for the Other versus White comparison groups (-2.50,  $p < 0.05$  versus -1.94,  $p = 0.09$ , see Tables 3 and 4). Typically, the pooled estimates' confidence intervals are wider in the multiple imputation method because of accounting for the additional variation among the 20 imputed samples.

Table 3. Estimated parameters of the individual growth model (larger sample)

|                                  | Coefficient    | Estimate   | 95% Confidence Intervals |         |
|----------------------------------|----------------|------------|--------------------------|---------|
|                                  |                |            | Lower                    | Upper   |
| <b>Intercept</b>                 | $\beta_{00}$   | 2486.96*** | 2471.89                  | 2502.03 |
| <b>Male</b>                      | $\beta_{01}$   | -16.33***  | -18.62                   | -14.03  |
| <b>FRPL</b>                      | $\beta_{02}$   | -30.93***  | -33.64                   | -28.21  |
| <b>SPED</b>                      | $\beta_{03}$   | -94.28***  | -98.24                   | -90.32  |
| <b>Asian</b>                     | $\beta_{04}$   | -3.63n.s.  | -7.96                    | 0.69    |
| <b>Filipino</b>                  | $\beta_{05}$   | -15.24***  | -19.30                   | -11.19  |
| <b>NH</b>                        | $\beta_{06}$   | -32.14***  | -36.11                   | -28.17  |
| <b>Other</b>                     | $\beta_{07}$   | -21.62***  | -26.22                   | -17.02  |
| <b>PI</b>                        | $\beta_{08}$   | -46.75***  | -51.76                   | -41.74  |
| <b>Time</b>                      | $\beta_{10}$   | 52.22***   | 45.64                    | 58.80   |
| <b>Male x Time</b>               | $\beta_{11}$   | -2.47***   | -3.47                    | -1.48   |
| <b>FRPL x Time</b>               | $\beta_{12}$   | -1.32*     | -2.50                    | -0.14   |
| <b>SPED x Time</b>               | $\beta_{13}$   | -7.39***   | -9.12                    | -5.66   |
| <b>Asian x Time</b>              | $\beta_{14}$   | -0.48n.s.  | -2.40                    | 1.44    |
| <b>Filipino x Time</b>           | $\beta_{15}$   | -0.13n.s.  | -1.94                    | 1.67    |
| <b>NH x Time</b>                 | $\beta_{16}$   | -5.36***   | -7.12                    | -3.60   |
| <b>Other x Time</b>              | $\beta_{17}$   | -2.29*     | -4.44                    | -0.14   |
| <b>PI x Time</b>                 | $\beta_{18}$   | -5.35***   | -7.57                    | -3.13   |
| <b>Intercept Variance</b>        | $\tau^2_{U0}$  | 3571.68*** |                          |         |
| <b>Intercept-Time Covariance</b> | $\tau^2_{U01}$ | 301.43***  |                          |         |
| <b>Time Variance</b>             | $\tau^2_{U1}$  | 80.03***   |                          |         |
| <b>Residual Variance</b>         | $\sigma^2$     | 1409.46*** |                          |         |

Notes: Sample includes incomplete waves for the 2013 cohort i.e., all third-grade students, whether they had grade 4 or grade 5 reading scores excluding any repeaters ( $n = 28$ ). n.s. – not significant,  $p > 0.05$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Table 4: Estimated parameters of the individual growth model using multiple imputations for missing data

|                                  | Coefficient    | Estimate   | 95% Confidence Intervals |         |
|----------------------------------|----------------|------------|--------------------------|---------|
|                                  |                |            | Lower                    | Upper   |
| <b>Intercept</b>                 | $\beta_{00}$   | 2486.70*** | 2471.82                  | 2501.92 |
| <b>Male</b>                      | $\beta_{01}$   | -16.33***  | -18.62                   | -14.03  |
| <b>FRPL</b>                      | $\beta_{02}$   | -30.94***  | -33.66                   | -28.23  |
| <b>SPED</b>                      | $\beta_{03}$   | -94.24***  | -98.19                   | -90.29  |
| <b>Asian</b>                     | $\beta_{04}$   | -3.72n.s.  | -8.04                    | 0.59    |
| <b>Filipino</b>                  | $\beta_{05}$   | -15.39***  | -19.43                   | -11.33  |
| <b>NH</b>                        | $\beta_{06}$   | -32.24***  | -36.21                   | -28.28  |
| <b>Other</b>                     | $\beta_{07}$   | -21.61***  | -26.22                   | -17.02  |
| <b>PI</b>                        | $\beta_{08}$   | -46.80***  | -51.80                   | -41.80  |
| <b>Time</b>                      | $\beta_{10}$   | 51.98***   | 45.16                    | 58.80   |
| <b>Male x Time</b>               | $\beta_{11}$   | -2.48***   | -3.50                    | -1.46   |
| <b>FRPL x Time</b>               | $\beta_{12}$   | -1.34*     | -2.55                    | -0.14   |
| <b>SPED x Time</b>               | $\beta_{13}$   | -7.43***   | -9.23                    | -5.63   |
| <b>Asian x Time</b>              | $\beta_{14}$   | -0.27n.s.  | -2.24                    | 1.70    |
| <b>Filipino x Time</b>           | $\beta_{15}$   | -0.15n.s.  | -1.81                    | 2.12    |
| <b>NH x Time</b>                 | $\beta_{16}$   | -5.10***   | -7.04                    | -3.16   |
| <b>Other x Time</b>              | $\beta_{17}$   | -1.94n.s.  | -4.16                    | 0.27    |
| <b>PI x Time</b>                 | $\beta_{18}$   | -5.11***   | -7.46                    | -2.76   |
| <b>Intercept Variance</b>        | $\tau^2_{U0}$  | 3530.82*** |                          |         |
| <b>Intercept-Time Covariance</b> | $\tau^2_{U01}$ | 322.87***  |                          |         |
| <b>Time Variance</b>             | $\tau^2_{U1}$  | 98.69***   |                          |         |
| <b>Residual Variance</b>         | $\sigma^2$     | 1436.64*** |                          |         |

Notes: Full sample (excluding 28 repeaters). n.s. – not significant,  $p > 0.05$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

The linear association between the intercept and time-slope residual deviances showed a correlation of 0.58. Therefore, Grade 3 reading (initial status) remains a strong predictor of a student's trajectory even after accounting for demographic characteristics and school fixed effects. Specifically, Grade 3 reading performance accounts for 33.7 percent of the variance in student growth differences above and beyond other demographic and school-level factors. Thus, the finding is consistent with the literature that Grade 3 reading proficiency is a critical milestone predictive of children's future success. The result also conforms to the Matthew effect, characterized by a strong association between a student's beginning reading achievement and subsequent learning rates.

## Discussion

Our analysis and findings are an essential first step in truly understanding various student groups' comparative academic growth patterns. We illustrate it with data from the Hawaii public schools, and school districts can replicate it with their data.

Our study confirms the need to intervene early, starting in elementary grades (Hernandez, 2011; Singh, 2011; Singh et al., 2014) and possibly even earlier than Grade 3 (Adams, 1994; Coyne et al., 2004; Fiester, 2013; Foster & Miller, 2007; Fuchs et al., 2004; Mastropieri et al., 1999; McCoach et al., 2006; McNamara et al., 2011). A promising foundation for future success depends on making high-quality early childhood education broadly accessible. Expanding offerings to include health and readiness, including cognitive and social-emotional development, can contribute to system-level change (Fiester, 2010, 2013; Hernandez, 2011). We also need to build a seamless transition from early to later grade. This systemic vision goes well beyond the school system and calls for an integrated and well-funded cross-sector approach.

Any system-level change will occur in the neighborhoods where students and their families live and their circumstances, including poverty, mobility, insecurity, and access to healthcare (Fiester, 2013). Student achievement is also closely connected to families, communities, and peers (Coleman, 1966). Closing achievement gaps is the responsibility of a wide range of stakeholders, including but certainly not limited to those who serve in the education field. Moreover, engaging caregivers and equipping them to support their children's learning and well-being is critical (Fiester, 2010; Lonigan & Shanahan, 2009). Leveraging local culture and asset framing is another opportunity. Over time, the educational experiences of NH children have contributed to their low academic performance. They have increasingly been educated in centers away from their families and communities since Protestant missionaries arrived in the state in the early 1800s with little attempts to leverage and integrate the local culture, beliefs, and priorities (Grace & Serna, 2012). Exploring culturally-appropriate early childhood (and beyond) educational offerings and instructional strategies may support

children's and youth's cognitive and social-emotional development and chip at the gap between NH students and their peers (Benner, 2011; Wurdeman-Thurston & Kaomea, 2015).

Ensuring student reading proficiency in Grade 3 and beyond will benefit from a diverse body of well-prepared and well-supported teachers and leaders who can effectively implement the differentiated, culturally responsive instructional practices that will help students learn and demonstrate their learning (Foorman et al., 2016; Gersten et al., 2008; Kamil et al., 2008; Lonigan & Shanahan, 2009; Rodriguez et al., 2004; Shanahan et al., 2010; Tomlinson, 2017).

School districts can implement interventions to increase attendance so students take full advantage of the effective educators who serve them (Fiester, 2010). Additional strategies to address other occasions for interrupted learning, such as summer loss and the consequences of the schooling during COVID-19, which disproportionately affected traditionally underserved students, are warranted (Fiester, 2010; Kuhfeld et al., 2021; Kwakye & Kibort-Corcker, 2021).

This body of connected efforts will help level the playing field between traditionally underserved children and students and their peers. Continuing to collect data will enable further research to deepen our understanding of the challenges, measure progress, and support accountability (Hein et al., 2013).

### Conclusion

Our study shows that the Grade 3 reading performance disparity between NH students and their peers appears in the early elementary grades. The gap widens in later elementary grades. Therefore, this early gap is a crucial policy issue in education in the United States (Slavin et al., 2009) and may be a unique opportunity for improvement. Investigations at the macro (district) and micro (schools and classrooms) levels would yield invaluable insights into students' inequitable outcomes and allow researchers, educators, and policymakers to improve these outcomes in the long- and short-term. Our findings support the need for further research, including qualitative studies. We must critically investigate factors such as classroom cultures and students' experiences that could be affecting the learning trajectories of students of color and other traditionally underserved students during and before elementary grades.

### Recommendations

While improving literacy achievements in the primary grades holds promise for future student outcomes, it is not an "inoculation to future literacy success" (Lai et al., 2014, p. 305). Student performance in middle school is another predictor of their future academic success (Balfanz et al., 2007; Kieffer & Marinell, 2012). Addressing low performance and gaps in middle school may reduce the need to do so in high school, by which time patterns of low performance are entrenched. Hence, the research community has called for substantial interventions to prevent high school dropout outcomes (Mac Iver, 2010). We intend to extend the longitudinal analyses from this study to the middle grades when the data becomes available.

Researchers have essential questions to understand NH students' circumstances to serve them better: Do NH families have the same access to appropriate early childhood education and healthcare? To what extent do they use those services, and why? Are the early health outcomes of NH children different from those of their non-N.H. peers? Do they live in equally safe neighborhoods? Are they more mobile? What resources do they have available in school, at home, and in the community? How is their learning supported at home? What of their social-emotional well-being? What do their schools need? What do their teachers need? What are the expectations for them of the adults in their life? Answers to these questions may unearth rich opportunities for success and equitable outcomes for NH children.

### Limitations

Although our findings support the Matthew effect, the data does not allow us to make causal statements. Other variables, such as parental education, may have increased the model's precision, but HODOE does not collect such data. Our results only generalize to the student population in Hawaii. However, other states and districts can use similar methods to check for similar patterns in their data.

### References

- Aarnoutse, C., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation*, 6(3), 251-278, [https://doi.org/10.1076/1380-3611\(200009\)6:3:1-A:FT251](https://doi.org/10.1076/1380-3611(200009)6:3:1-A:FT251)
- Adams, M. J. (1994). *Beginning to read: Thinking and learning about print*. MIT Press.
- Azevedo, J. P., Rogers, F. H., Ahlgren, S. E., Cloutier, M. H., Chakroun, B., Changl, G. C., Mizunoya, S., Reuge, J. N., Brossard, M., & Lynn, J. (2021). *The state of the global education crisis: A path to recovery (Vol. 2): Executive summary (English)*. World Bank Group. <https://bit.ly/3vQgPCq>

- Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, 42(4), 223-235. <https://doi.org/10.1080/00461520701621079>
- 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(6), 1373. <https://doi.org/10.1037/0012-1649.34.6.1373>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Benner, A. D. (2011). The transition to high school: Current knowledge, future directions. *Educational Psychology Review*, 23(29). <https://doi.org/10.1007/s10648-011-9152-0>
- Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal*, 108(2), 115-130. <https://doi.org/10.1086/525550>
- Camera, L. (2016, January 16). *Achievement gap between white and black students still gaping*. US News & World Report. <https://bit.ly/3LmX7Vj>
- 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>
- Coleman, J. S. (1966). *Equality of educational opportunity*. National Center for Education Statistics. <https://bit.ly/30K2no4>
- Coyne, M. D., Kame'enui, E. J., Simmons, D. C., & Harn, B. A. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities*, 37(2), 90-104. <https://doi.org/10.1177/00222194040370020101>
- Enders, C. K., Du, H., & Keller, B. T. (2018). Substantive model-compatible imputation for multilevel regression models with random coefficients, interaction effects, and other non-linear terms. *Psychological Methods*.
- Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review*, 57(1), 72-84. <https://doi.org/10.2307/2096145>
- Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality*. Westview.
- Every Student Succeeds Act, 20 USC § 6301 (2015). <https://bit.ly/3vK00aI>
- Fiester, L. (2010). *Early warning: Why reading by the end of third grade matters*. Kids Count special report. Annie E. Casey Foundation. <https://bit.ly/3ES1Ye0>
- Fiester, L. (2013). *Early warning confirmed: A research update on third-grade reading*. The Annie E. Casey Foundation. <https://bit.ly/3MzbBSj>
- Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd Grade* (NCEE 2016-4008). National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences. <https://bit.ly/3EWJom4>
- 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\(2007/018\)](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(1), 3-17. <https://doi.org/10.1037/0022-0663.88.1.3>
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first Grade: Word identification fluency versus nonsense word fluency. *Exceptional Children*, 71(1), 7-21. <https://doi.org/10.1177/001440290407100101>
- Gelfand, A. E. (2000). Gibbs sampling. *Journal of the American Statistical Association*, 95(452), 1300-1304. <https://doi.org/10.1080/01621459.2000.10474335>
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & 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). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. <https://bit.ly/3vKAcgl>

- Grace, D. J., & Serna, A. K. (2012). Early childhood education and care for Native Hawaiian children in Hawai'i: A brief history. *Early Child Development and Care*, 183(2), 308-320. <https://doi.org/10.1080/03004430.2012.673487>
- Hein, V., Smerdon, B., & Sambolt, M. (2013). *Predictors of postsecondary success*. College & Career Readiness & Success Center at American Institutes for Research. <https://eric.ed.gov/?id=ED555671>
- Hemphill, F. C., & Vanneman, A. (2011). *Achievement gaps: How Hispanic and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress* (NCES 2011-459). National Center for Education Statistics, Institute of Education Sciences, US Department of Education. <https://bit.ly/3vkMn4i>
- Hernandez, D. J. (2011). *How third-grade reading skills and poverty influence high school graduation*. Annie E. Casey Foundation.
- Johnson, R. S. (2002). *Using data to close the achievement gap: How to measure equity in our schools*. Corwin Press.
- Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A Practice Guide* (NCEE #2008-4027). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. <https://bit.ly/3OAAtep>
- Kempe, C., Eriksson-Gustavsson, A. L., & Samuelsson, S. (2011). Are there any Matthew effects in literacy and cognitive development? *Scandinavian Journal of Educational Research*, 55(2), 181-196. <https://doi.org/10.1080/00313831.2011.554699>
- Kieffer, M. J., & Marinell, W. H. (2012). *Navigating the middle grades: Evidence from New York City*. The Research Alliance for New York City Schools.
- Kim, Y.-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology*, 102(3), 652-667. <https://doi.org/10.1037/a0019643>
- Kuhfeld, M., Ruzek, E., Lewis, K., Soland, J., Johnson, A., Tarasawa, B., & Dworkin, L. (2021). *Understanding the initial educational impacts of COVID-19 on communities of color*. NWEA. <https://bit.ly/3KqwmY2>
- Kwakye, I., & Kibort-Corcker, E. (2021). *Facing learning disruption: Examining the effects of the COVID-19 pandemic on K-12 students*. Washington Student Achievement Council. <https://eric.ed.gov/?id=ED613296>
- Lai, M. K., Wilson, A., McNaughton, S., & Hsiao, S. (2014). Improving achievement in secondary schools: Impact of a literacy project on reading comprehension and secondary school qualifications. *Reading Research Quarterly*, 49(3), 305-334. <https://doi.org/10.1002/rrq.73>
- Leppänen, U., Niemi, P., Aunola, K., & Nurmi, J. E. (2004). Development of reading skills among preschool and primary school pupils. *Reading Research Quarterly*, 39(1), 72-93. <https://doi.org/10.1598/RRQ.39.1.5>
- Lerkkanen, M. K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J. E. (2004). Reading performance and its developmental trajectories during the first and the second grade. *Learning and Instruction*, 14(2), 111-130. <https://doi.org/10.1016/j.learninstruc.2004.01.006>
- Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology*, 99(4), 821-834. <https://doi.org/10.1037/0022-0663.99.4.821>
- Lesnick, J., Gorge, R. M., Smithgall, C., & Gwynne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college enrollment? A longitudinal analysis of third-grade students in Chicago in 1996-97 and their educational outcomes. A report to the Annie E. Casey Foundation*. Chapin Hall at the University of Chicago. <https://eric.ed.gov/?id=ED517805>
- Lonigan, C. J., & Shanahan, T. (2009). *Developing early literacy: Report of the National Early Literacy Panel. Executive summary. A scientific synthesis of early literacy development and implications for intervention*. National Institute for Literacy. <https://bit.ly/3LrAkaY>
- Lubienski, S., & Crane, C. C. (2010). Beyond free lunch: Which family background measures matter? *Education Policy Analysis Archives*, 18(11), 1-43. <https://doi.org/10.14507/epaa.v18n11.2010>
- Mac Iver, M. A. (2010). *Gradual disengagement: A portrait of the 2008-09 dropouts in the Baltimore City Schools* (Research report). Baltimore Education Research Consortium. <https://bit.ly/3rXpG4f>
- Mastropieri, M. A., Leinart, A., & Scruggs, T. E. (1999). Strategies to increase reading fluency. *Intervention in school and clinic*, 34(5), 278-283. <https://doi.org/10.1177/105345129903400504>
- McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of educational psychology*, 98(1), 14-28. <https://doi.org/10.1037/0022-0663.98.1.14>

- McNamara, J. K., Scissons, M., & Gutknecht, N. (2011). A longitudinal study of kindergarten children at risk for reading disabilities: The poor really are getting poorer. *Journal of learning disabilities*, 44(5), 421-430. <https://doi.org/10.1177/0022219411410040>
- Morgan, P. L., Farkas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly*, 31(4), 187-198. <https://doi.org/10.2307/25474651>
- Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities*, 44(5), 472-488. <https://doi.org/10.1177/0022219411414010>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Author. <https://bit.ly/37K2oYO>
- Office of Elementary & Secondary Education. (n.d.). *Elementary and secondary school emergency relief fund*. <https://bit.ly/3ETQWG2>
- Onatsu-Arvilommi, T., & Nurmi, J. E. (2000). The role of task-avoidant and task-focused behaviors in the development of reading and mathematical skills during the first school year: A cross-lagged longitudinal study. *Journal of educational psychology*, 92(3), 478-491. <https://doi.org/10.1037/0022-0663.92.3.478>
- 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), 299-319. <https://doi.org/10.1037/0022-0663.97.3.299>
- Patz, R. J. (2007). *Vertical scaling in standards-based educational assessment and accountability systems*. Council of Chief State School Officers.
- Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research*, 84(2), 203-244. <https://doi.org/10.3102/0034654313509492>
- Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology*, 94(1), 3-13. <https://doi.org/10.1037/0022-0663.94.1.3>
- Quintana, R., & Correnti, R. (2020). The concept of academic mobility: Normative and methodological considerations. *American Education Research Journal*, 57(4), 1625-1664. <https://doi.org/10.3102/0002831219876935>
- R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rigney, D. (2010). *The Matthew effect: How advantage begets further advantage*. Columbia University Press.
- Rodriguez, J. L., Jones, E. B., Pang, V. O., & Park, C. D. (2004). Promoting academic achievement and identity development among diverse high school students. *High School Journal*, 87(3), 44-53. <https://doi.org/10.1353/hsj.2004.0002>
- Rourke, B. P. (1976). Reading retardation in children: developmental lag or deficit? In R. M. Knights & D. J. Bakker (Eds.), *The neuropsychology of learning disorders: theoretical approaches* (pp. 125-137). University Park Press.
- Salaschek, M., Zeuch, N., & Souvignier, E. (2014). Mathematics growth trajectories in first grade: Cumulative vs. compensatory patterns and the role of number sense. *Learning and Individual Differences*, 35, 103-112. <https://doi.org/10.1016/j.lindif.2014.06.009>
- Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd Grade: A practice guide (NCEE 2010-4038)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. <https://bit.ly/3LE28cB>
- Shanley, L. (2016). Evaluating longitudinal mathematics achievement growth: Modeling and measurement considerations for assessing academic progress. *Educational Researcher*, 45(6), 347-357. <https://doi.org/10.3102/0013189X16662461>
- Shin, T., Davison, M. L., Long, J. D., Chan, C. K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, 23, 92-100. <https://doi.org/10.1016/j.lindif.2012.10.002>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195152968.001.0001>
- Singh, J. M. (2011). *Long-term effects of Native Hawaiian students' early academic achievement under the No Child Left Behind legislation: A multilevel cohort analysis* (Publication No. 3475730) [Doctoral dissertation, University of Hawai'i at Manoa]. ProQuest Dissertations and Theses Global.

- Singh, M., Bel Hadj Amor, H., & Zhang, S. (2014). *Native Hawaiian students' achievement gap in reading: A longitudinal study from Hawai'i*. Education Northwest.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). *Effective beginning reading programs: A best-evidence synthesis*. Center for Data-Driven Reform in Education.
- Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of education*, 189(1-2), 23-55. <https://doi.org/10.1177/0022057409189001-204>
- The Nation's Report Card. (2019). *Results from the 2019 mathematics and reading assessments*. Author. <https://bit.ly/3Ls7Y0y>
- Tomlinson, C. A. (2017). *How to differentiate instruction in academically diverse classrooms* (3rd ed.). ASCD.
- Vanneman, A., Hamilton, L., Baldwin Anderson, J., & Rahman, T. (2009). *Achievement gaps: How black and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress* (NCES 2009-455). National Center for Education Statistics, Institute of Education Sciences, US Department of Education. <https://bit.ly/3y5wDUu>
- Walberg, H. J., & Tsai, S. L. (1983). Matthew effects in education. *American Educational Research Journal*, 20(3), 359-373. <https://doi.org/10.3102/00028312020003359>
- Williamson, G. L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analyses of academic achievement. *Journal of Educational Measurement*, 28(1), 61-76. <https://doi.org/10.1111/j.1745-3984.1991.tb00344.x>
- Wurdeman-Thurston, K., & Kaomea, J. (2015). Fostering culturally relevant literacy instruction: Lessons from a Native Hawaiian classroom. *Language Arts*, 92(6), 424-435. <https://www.jstor.org/stable/24577534>
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Society for Research on Child Development and Foundation for Child Development. <https://bit.ly/3vRfKdy>

**Appendix***Table A1. Descriptive statistics across grades 3 to 5 disaggregated by racial/ethnic student groups. With a vertically aligned assessment, mean scores increase for student groups from one grade to another.*

| <b>Demographics</b> | <b>Grade 3</b> |          |           | <b>Grade 4</b> |          |           | <b>Grade 4</b> |          |           |
|---------------------|----------------|----------|-----------|----------------|----------|-----------|----------------|----------|-----------|
|                     | <i>N</i>       | <i>M</i> | <i>SD</i> | <i>N</i>       | <i>M</i> | <i>SD</i> | <i>N</i>       | <i>M</i> | <i>SD</i> |
| Asian               | 2298           | 2463     | 84.56     | 2177           | 2511     | 89.28     | 2109           | 2550     | 91.09     |
| Filipino            | 3150           | 2425     | 80.44     | 3040           | 2474     | 85.30     | 2956           | 2514     | 88.89     |
| NH                  | 3404           | 2392     | 80.58     | 3237           | 2432     | 89.12     | 3181           | 2467     | 93.48     |
| Other               | 1415           | 2417     | 84.81     | 11198          | 2463     | 89.83     | 1047           | 2499     | 96.50     |
| PI                  | 1418           | 2375     | 80.34     | 1299           | 2420     | 88.57     | 1232           | 2453     | 95.13     |
| White               | 2627           | 2452     | 83.36     | 2111           | 2498     | 89.09     | 1755           | 2538     | 92.94     |