

# Readiness for Entering Kindergarten: The Impact on Future Academic Achievement

Lynn D. Fielding, Jay Maidment,  
and Christian N. K. Anderson

Overview: This paper<sup>1</sup> describes how children’s entering kindergarten scores in reading and math predict their subsequent achievement each year with surprising accuracy through Grade 5 but also, though less accurately, through Grade 8. Part I summarizes our 13 major conclusions and observations in a 6 pages tightly written format shaped for a fairly nontechnical readership but based on some statistical correlations and technical explanations. If it is too technical, skip to Part II which covers the same points in a longer but more lay-person/parent-friendly descriptions with graphs showing the relationship of children’s scores when they enter kindergarten compared to the odds of where they will be five years later. Part III assumes statistical expertise in its supporting tables, graphs, and extended descriptions of the mathematical models. Appendices A-D provide supplementary data.

## PART I

**DATASET INFORMATION:** The Northwest Evaluation Association (NWEA) located in Portland, Oregon,<sup>2</sup> has, since 1976, tested scores of millions of elementary school children in the United States. In 2015 it provided us with test results for five cohorts of students (about 380,000), beginning with students who started kindergarten in the fall of 2006 and were tracked through the fall of 2010, and ending with students who began in the fall of 2010, tracked through spring 2015.<sup>3</sup> This paper uses two groups of students. The reading set (n=61,691) consists of students for whom 12 consecutive reading scores from fall-K to spring-Grade 5 are available in addition to these same students’ math scores beginning and end of school year from fall kindergarten

---

<sup>1</sup> This paper is an amalgamation of the work product of Lynn Fielding, Jay Maidment, and Christian Anderson from March 2015 to the present. Jay Maidment prepared a 26-page stand-alone work dated February 16, 2017, describing the development of a model predicting reading test scores in the fall of Grade 5 from kindergarten fall scores. The Maidment work describes the chronological process of the model’s development, its final state, and its mathematical nuances in greater detail than this paper.

<sup>2</sup> See Appendix A for an extended description of NWEA.

<sup>3</sup> See Appendix B for an extended description of this dataset, including some science scores.

through Grade 5. The math set (n=51,781) consists of students for whom 12 consecutive math scores are available from fall-K through spring-Grade 5 in addition to these same students' fall kindergarten reading score. Thus, we have scores for 12 tests for every student analyzed in this study.<sup>4</sup>

- 1. Range at kindergarten.** National organizations report a five-year range in language-and literacy-related skills (reading) of entering kindergartners. "Some kindergartners may have skills characteristic of the typical 3-year-old, while others might be functioning at the level of the typical 8-year-old."<sup>5</sup> Our analysis of NWEA data is not inconsistent with this finding. A four-to-five-year gap would also explain why it is so difficult for children who are behind and whose scores lie in the 1-20<sup>th</sup> percentiles have such difficulty catching up to the level of students who start ahead and whose scores lie in the 80<sup>th</sup>-99<sup>th</sup> percentiles.
- 2. When do students move to a higher or lower quintile in reading and/or math?** A common assumption is that students who attend school regularly, study, and generally take education seriously (or have parents who do) will improve steadily over time, climbing into the next highest quintile with similarly motivated and/or talented classmates, while those who are less motivated or productive will drift down steadily into the next lower quintile. This is true for a limited amount of students. However, our data suggest a surprisingly different picture for a majority of students: e.g., that while individual students do indeed move a limited distance (either up or down) from their beginning reading and math quintiles each year, on average, kindergartners will have moved about 85% as far by the fall of Grade 1 as they will by the fall of Grade 8. This 85% holds true regardless of what yardstick the model uses: percentile, decile, standard deviation, class rank, or RIT points above average (73% during kindergarten and 90% by Grade 3). Each of these yardsticks confirms that on average students move about as far in the first year as they do in nine years. This cap on net movement explains why we can predict Grade 8 test scores almost as accurately as the scores for Grades 3 or 5. This is *net movement*. Between each grade, students do indeed rise and fall a great deal as measured by any of the yardsticks listed, but on balance most are drawn back towards their starting position. (See Part III for statistical patterns shown in a variety of ways.)

---

<sup>4</sup> Missing scores were interpolated for reading (10.5%) and math (11.7%) when a score was available for the previous and subsequent year.

<sup>5</sup> This description can be understood as the current stance taken by the nation's professional educators. See *Learning to Read and Write: Developmentally Appropriate Practices for Young Children*. It is the position statement prepared jointly by the International Reading Association and the National Association for the Education of Young Children, adopted May 1998 and still, after 20 years, considered the most solid professional position. The statement's authors are quoting Dr. Jeni Riley of the University of London. This joint NAEYC/IRA position statement is endorsed by the following organizations: American Speech Language-Hearing Association, Association for Childhood Education International, Association of Teacher Educators, Council for Early Childhood Professional Recognition, Division for Early Childhood/Council for Exceptional Children, National Association of Early Childhood Specialists in State Departments of Education, National Association of Early Childhood Teacher Educators, National Association of Elementary School Principals, National Association of State Directors of Special Education, National Council of Teachers of English, Zero to Three/National Center for Infants, Toddlers, & Families. The concepts in this joint position statement are supported by the following organizations: American Academy of Pediatrics, American Association of School Administrators, American Educational Research Association, and the National Head Start Association.

- 3. Relationship of summer reading gain and loss to entering kindergarten scores.** A common assumption held by educators (and usually parents) is that the summer break means that momentum falters, students no longer exercise their learning skills in a disciplined way, and their academic achievement across those vacation months suffers as a result. We found two patterns when we analyzed summer gain and loss, focused on the correlation between spring kindergarten scores and those of the fall of Grade 1.
- a. Gain: Most (64.4%) students, and all percentile groups, gain a few RIT points between the end of kindergarten and the beginning of Grade 1. Thereafter, only the top three percentiles consistently gain points over summer vacation from Grades 1 through 5. The 97 lower percentiles consistently lose a few RIT points.
  - b. Cumulative gain or loss by percentile group in Grades K-5: Across all five vacations from the end of kindergarten to the beginning of Grade 5, the lowest 50 percentile groups lose 4-5 RIT points. The percentile groups above 50 lose progressively fewer points until the 80<sup>th</sup> percentile group. The 80<sup>th</sup> to 100<sup>th</sup> percentile groups continue to gain (maximum 4-5 RIT points). The result is a cumulative difference of 8-10 RIT points between the top and bottom percentiles, equivalent to a little more than a year of normal growth.<sup>6</sup>
  - c. Despite these clear trends in *group* scores, the impact of fall kindergarten reading scores on summer gain and loss on the *individual* level is masked by the extreme variability in individual summer gain and loss. As a result, the correlations between these two variables measured on an *individual* level (fall-K reading and total summer gain through Grade 5) is small ( $r^2 = 0.015$ ).
- 4. Impact of initial starting points in reading using a simple correlation. But first, a short review about correlations.** When two variables correlate perfectly (rare), the correlation is 1.0, and when individual data values of such variables are graphed, one variable on the x axis and the second on the y axis, the resulting dots form a perfectly straight line. When two variables are not correlated at all, the correlation is 0.0, and when the individual values of those two variables are graphed, the individual values look like a shotgun pattern. Most correlations fall somewhere between these extremes. According to a rule-of-thumb known as Cohen’s standard, correlations between 0.10 and 0.29 represent a small association, coefficients between 0.30 and 0.49 represent a medium association, and coefficients of 0.50 and above represent a large and strong association or relationship.

Correlations are known as “r” values. When “r” values are squared ( $r^2$ ), the results are commonly stated as percentages from 0% to 100% and explain the percent of the relationship between two variables attributable solely to the two variables.

Now back to the impact of initial starting point in reading.

We first calculated the impact of incoming kindergarten scores on subsequent academic performance:

---

<sup>6</sup> “Normal growth” is here is the difference in RIT points between a given percentile at the Grades 4 and 5, which at the 20th, 50th, 70th, and 90th percentiles—respectively is 7, 8, 7, and 7 RIT points. NWEA 2011 RIT Scale Norms Study, rev. January 2012, Table B-1, Fall, Reading.

- a. *By Individuals:* The degree of association between incoming language and literacy scores (reading) of 61,691 entering kindergartners and their scores in the fall of each subsequent year is 0.691 (Grade 1), 0.581 (Grade 2), 0.549 (Grade 2), 0.539 (Grade 4), and 0.531 with an  $R^2$  of 0.282 at Grade 5, ( $0.531 \times 0.531 = 0.282$ ). In other words, entering kindergarten reading scores account for 28% of the range we see by Grade 5.
- b. *By percentile groups:* We sorted the fall kindergarten scores of the 61,691 students from high to low. We then divided them into 100 groups of 617 students each (percentile groups) and compared each group's average kindergarten score with that group's average score in each subsequent year. The correlations are 0.993 (Grade 1), 0.989 (Grade 2), and 0.993 for each grade thereafter, with an  $R^2$  of 0.989 at Grade 5. The *average* Grade 5 reading score can be predicted within 5/9ths of a RIT point<sup>7</sup> from average kindergarten fall reading scores for percentile groups.<sup>8</sup>

**5. Impact of initial starting points in reading using math as an additional variable in a two-variable reading prediction model.**

- a. *By group:* In order to investigate whether entry-level math skills had any effect on reading ability five years after kindergarten, we further divided the 100 reading percentile groups of 617 students into quintiles based on their fall kindergarten math scores. This step yielded 500 groups of 123 students each. It is always more difficult to predict results for small groups than large groups. Hence, correlations between kindergarten reading and later grade reading scores are lower for the 500 small groups than for the 100 large groups (0.924 for the 500 small groups vs. 0.993 for the 100 large groups in Grade 5). However, including kindergartners' math scores in the predictive model boosted these correlations back up to high levels (0.986 in Grade 5).
- b. *By individual:* These results suggested to us that entry-level math skills are associated with later reading skills. Predicting later *individual* reading scores using both entry-level math and reading does, in fact, improve individual correlations by about 0.04 at each grade level (See Part III).

**6. Comparison to poverty correlations.** Educators often think of poverty as a main driver of academic success. Actually, we found that kindergarten reading readiness was almost twice as strong in predicting future academic achievement.

- a. According to Snow, Burns, and Griffin's influential work, when achievement scores and SES are measured individually for all children in a large sample, the strength of the

---

<sup>7</sup> The NWEA has created a large bank of test question in language usage, reading, mathematics and science with each question assigned a difficulty level along an interval scale which runs vertically through all grades. The scale was developed using a model devised by a Danish mathematician Georg Rasch (1901-1980) who created a series of social science measures adaptable to education assessment. The units on the NWEA interval scale are called *Rasch Units* or RITs.

<sup>8</sup> The correlation between the 100 average group scores of 617 students in kindergarten and Grade 5 is high, although the scatter or range of *individual* scores within each group of 617 students is wide. The Law of Large Numbers states that the range of average scores should be ~25 times smaller than the range of individual scores for groups of 617 students.

association ranges between 0.22 and 0.31. The strength of association when the unit of measure is a school or district in a large sample is in the range of 0.68.<sup>9</sup>

- b. By contrast, and as shown in 1b, the correlation between reading scores in fall K and future academic performance is almost twice as high: 0.531 for individual students (or 0.569 when including math) and 0.986 to 0.996 for groups, depending on the size of the group.

Obviously, poverty and incoming scores are cross-correlated but our data set did not have free or reduced lunch counts for either the individual students or for the schools so we could not determine the degree of their association. It is difficult to affect poverty as a variable but incoming scores can be changed. Given the greater size of the correlation between kindergarten and Grade 5 scores, it is likely that if parents and school districts can increase low incoming scores, it is likely to erase much of the impact traditionally attributed to poverty.

7. **Impact of initial starting points in math using a simple correlation.** Replicating the same analysis above but switching math scores for reading scores, we calculated the impact of initial math scores at kindergarten scores on subsequent academic math performance on an individual and group level as follows:
  - a. *By individuals:* The degree of association of math scores in the fall of kindergarten of 51,781 students to the fall math scores of each subsequent year is 0.707 (Grade 1), 0.612 (Grade 2), 0.606 (Grade 2), 0.600 (Grade 4), and 0.594 (Grade 5).
  - b. *By percentile groups:* The degree of association of average fall kindergarten and fall Grade 5 math scores is 0.995. Our database was 51,781 students, sorted by their kindergarten scores, divided into 100 groups of 517 students, and compared to the average scores of these same student groups in the fall of Grade 5.<sup>10</sup>
8. **Impact of initial starting points in math also using reading in a two-variable reading prediction model.** The degree of association of average fall kindergarten and average fall Grade 5 math scores of the 51,781 students is 0.989 with an  $r^2$  of 0.978. We derived this conclusion using the same process as for reading. That is, we sorted the 51,781 students into percentiles (that is, 100 groups of 517 each), based on their fall kindergarten *math scores*, then subdivided each of those 100 groups into quintiles based on their *reading scores* (thus generating 500 groups containing 103 students each). As before, we found that entry-level reading skills are associated with later math skills. Using both reading and math at the individual level modestly improved model predictions (from 0.594 to 0.608 in Grade 5; see Sections 1a and 1b).
9. **“Free” increases in math scores when reading scores increase.** One of the consistent unpublished findings in previous research has been that, when reading scores improve, math scores also improve even though teachers may not be putting any additional focus on

---

<sup>9</sup>Catherine E. Snow, M. Susan Burns, and Peg Griffin, eds., *Preventing Reading Difficulty in Young Children* (Washington, DC: National Research Council/National Academy Press, 1998), 126-127.

<sup>10</sup> The correlations slightly stronger for math and reading, the drop is greater in reading (drops by 0.16) than math (drops 0.11). For both subjects, most of the drop comes during kindergarten for reading (69%) and math (84%).

math.<sup>11</sup> Relatively few students have kindergarten-Grade 5 scores (n=61,700), so we investigated this trend for Grades 1-5; 310,000 students had complete reading scores for all of grade school. We found a strong correlation between gain in reading and gain in math between Grades 1 and 5 (n=310,000,  $r=0.469$ , slope = 0.45). (See Part III.)

- 10. Science correlations.** At our request, the NWEA provided spring scores at Grade 5 (n=49,785) and Grade 8 (n=16,064) in science. Using these data, we created models for Grade 5 science scores which explain:
- 28.7% of the fall Grade 5 score variation ( $r^2$ ) using fall kindergarten reading and fall kindergarten math scores. A further analysis of this model reveals that 13.2% of the predictability comes from reading and 15.5% from math. In other words, we can predict Grade 5 science scores from kindergarten scores as accurately as we can predict Grade 5 reading scores.
  - 35% percent of the variation ( $r^2$ ) using Grade 1 test scores: fall reading scores (15.5%) and fall math scores (18.5%).
  - 59% of the variation ( $r^2$ ) using Grade 5 fall reading and math scores, ethnicity, and gender. It is significant that, while fall reading explains 30% of the variation and fall math 27%, the two demographic factors (gender and ethnicity) explain just 2% together. In harmony with points 10a and 10b above, this finding indicates that starting point is a far better predictor of subsequent performance than demography
- 11. Relative impact studies.** Because making optimal use of limited educational resources is so important, some researchers<sup>12</sup> have identified and ranked factors that have the greatest impact on student achievement. However, we know of no study that has included entering kindergarten literacy and math skills as a variable. We argue that the level of literacy and math skills possessed by entering kindergartners not only significantly influences subsequent achievement but may be the strongest single factor.
- 12. Testing for other factors.** We used two demographic factors (gender and ethnicity) to examine their influence compared to kindergarten reading and math scores. The effects were slight—more than ten times weaker than entry-level math and reading ability when they were detectable at all.
- 13. Causation.** This single study of correlations of initial starting points and subsequent achievement does not rise to the level of statistical causation.

---

<sup>11</sup> This “free” improvement in math, independent of special instruction or emphasis, emerged in the 2005-06 analysis by Maidment of student scores in the entire Washington State, followed by analysis of Kennewick and Moses Lake NWEA scores. This conclusion was also seen in a large project funded by the Elgin Foundation in the Appalachian areas of southern Kentucky and southern Virginia managed by Lynn Fielding, Linda Vinson, and Dave Montague in 2011-2013. The present study of kindergarten impact also checked specifically for this “free math” improvement. All four studies found the same correlation between reading improvement and math improvement.

<sup>12</sup> Researchers who have investigated and ranked influences include William Sanders “value added” work in the late 80s and 90s, and John Hattie’s *252 Influences and Effect Sizes Relating to Student Achievement* <https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement> (2018).

# PART II

Note: This section uses the same headings/numbering in Part I but strives for more family-friendly language, graphs and charts to communicate with those who have no specialized training in statistics. There is a quick refresher on correlations in Part I paragraph 4 above.

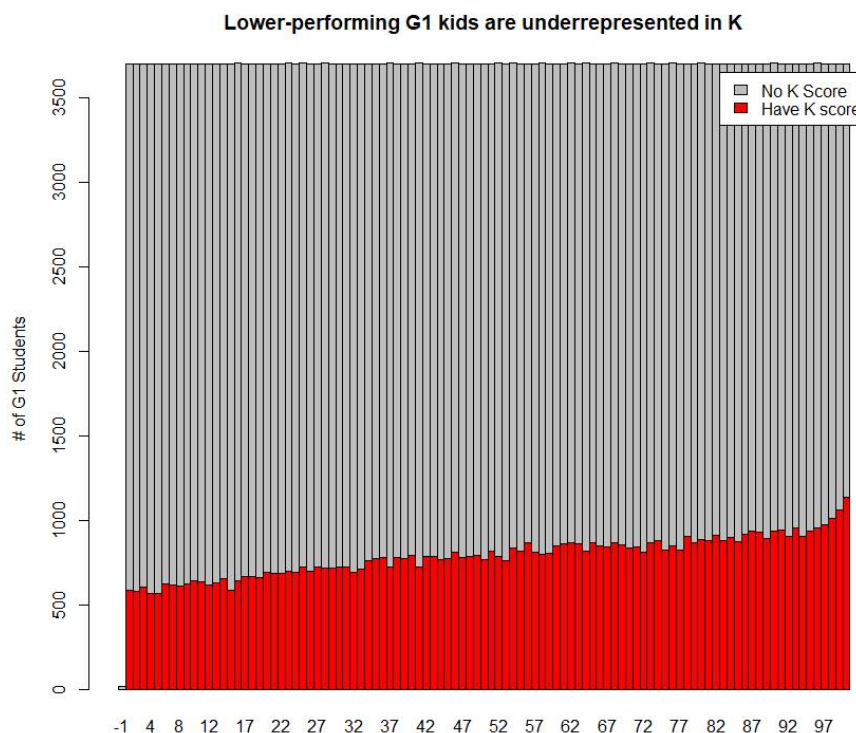
**Data set:** The kindergarten reading data set of 61,691, when compared against the Grade 1 data set of the 380,000 students shows an underrepresentation of students in the lower percentiles.

The direction of the skew suggests that our findings are slightly more conservatively stated than they would be without the skew.

**1. Range at kindergarten.** As discussed in Part I.1, a consensus has solidified among the nation's educators that children enter kindergarten with a challenging range of skills—up to five years. Expecting teachers to spend those first crucial nine months catching up

children who lag far behind their peers and, simultaneously, continue to stimulate children who are reading independently and exploring math concepts is not realistic. Parents and care-giving institutions can do a great deal toward ensuring that children are ready for kindergarten and, as a result, prepared to experience and enjoy academic success in later grades.

**2. When are optimum periods of improvement in reading and math?** According to our analysis, the period of greatest academic differentiation occurs between birth and kindergarten. Parents who want to smooth their children's academic paths will have the greatest influence on achievement by focusing on those crucial first five years. The second period of greatest differentiation occurs between kindergarten's fall and spring, followed by the nine months of Grade 1. The take-away message for elementary school principals is to



take advantage of this second period by placing their most effective reading teachers and aides in kindergarten and Grade 1. During Grades 2-8, students who have a firm foundation in K-Grade 1 will continue to surge ahead in relation to their peers. Time on task and the quality of teaching will certainly influence the elementary years both positively and negatively; but our analysis suggests that later educational achievement will have difficulty overcoming the impact of this early influence.

Phrased as educational odds, students who test in the bottom quintile<sup>13</sup> in the fall of kindergarten have about a 50% chance of escaping that quintile by the fall of Grade 1 (6,039 of 12,338 = 48.95%). If they are still in the bottom quintile when they start Grade 1, they have less than a 40% chance of moving up out of that ranking during the next five years (2,482 of 6,399 = 39.4%) and only a 47% chance over the next eight years (399 of 851 = 46.9%). Encouragingly, students whose kindergarten achievement brings them out of the bottom quintile by Grade 1 have a good chance of staying above it; only about 25% will have fallen back to it by Grade 5 (1,593 of 6,299 = 25.3%) or even by Grade 8 (223 of 807 = 27.6%).

Students who start kindergarten in the top quintile (80<sup>th</sup> percentile or higher) have more than a 60% chance of holding that ranking until the fall of Grade 1 (7,522 of 12,338 = 60.97%), a 50% chance until Grade 5 (6,309 of 12,518 = 50.4%), and a 48% chance until Grade 8 (841 of 1,744=48.2%). Of those 60% of students who remain in the top quintile from kindergarten to Grade 1, a full two-thirds will also be in the top quintile in Grade 5 (4,964 of 7,522 = 65.99%), while only about a quarter of students who fall below the top quintile climb back to it by Grade 5 (1,345 of 4,816 = 27.9%).

The news from the center quintile or grade level (41<sup>st</sup> to 60<sup>th</sup> percentile) and the quintiles on either side are more hopeful and can be seen from the chart on page 12 “Movement of 61,691 Students from Initial Reading Quintiles to the Fall of Grade 5.

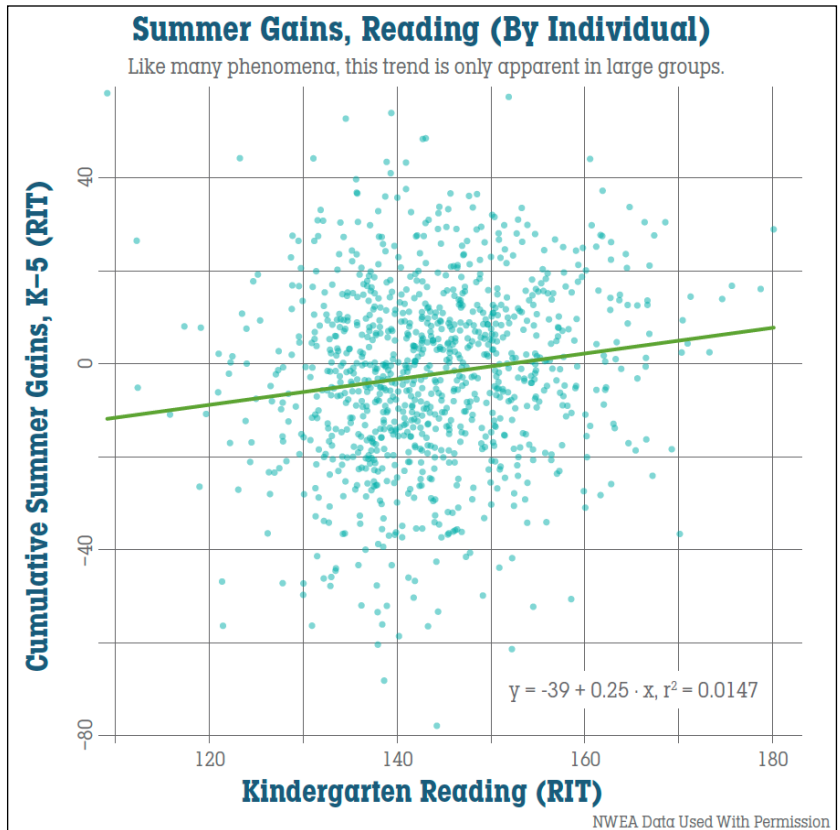
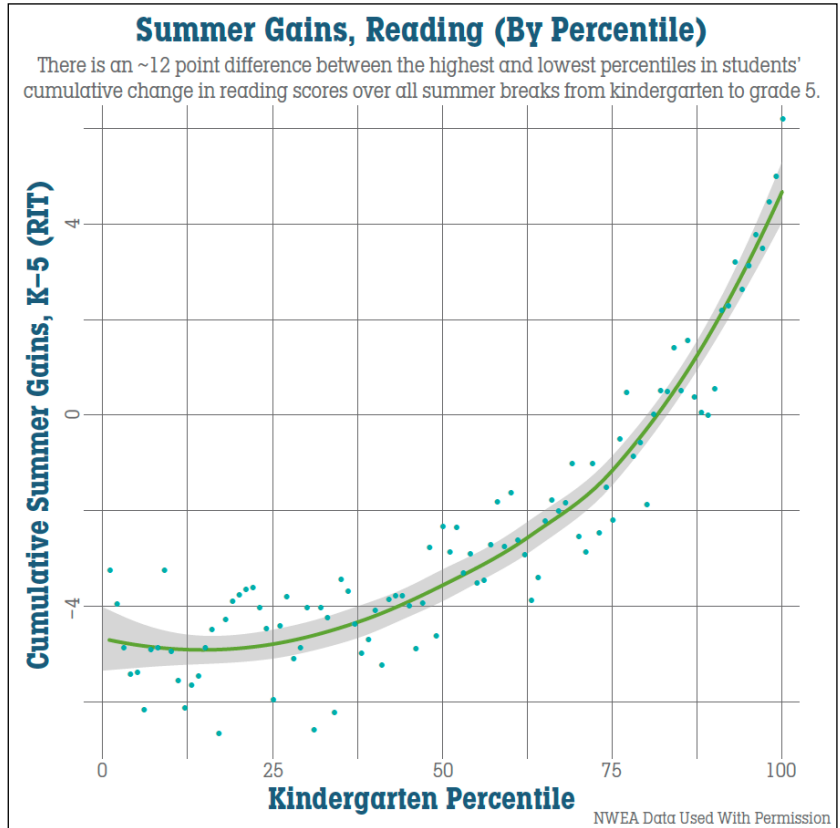
- 3. Relationship of summer gain and loss to entering kindergarten scores.** Popular wisdom suggests that students will slip backward in academic achievement over summer breaks. This is both true and not true. Individuals vary so much *within* groups that it’s very difficult to make any predictions except *for* groups. To be precise, between leaving kindergarten in the spring and entering Grade 1 in the fall, students actually gain a few RIT points in reading ability. For each of their summer breaks thereafter, they will lose a few RIT points. If they entered kindergarten with high reading scores, summer vacations will have less impact on their reading ability than they do on students with low reading ability. However, this pattern shows up only for groups, not for individuals. (See the next two graphs, plotted first for individuals, then for groups.)

---

<sup>13</sup> “Quintile” means that the total population we are studying is divided into five equal parts.

Summer gain and loss numbers are typically calculated using large groups, which effectively masks much of the variability that occurs between individuals. This scatterplot shows cumulative summer gain / loss for individuals K-Grade 5. The trendline predicts a 12-point difference between low-entry and high-entry kindergartners, quite a small amount compared to the nearly 250-RIT points of difference among individuals.

In contrast, the scatterplot shows the same data, but with individuals grouped into 100 percentiles, each containing of 617 students with comparable scores. The 12-RIT point difference between high and low entry-point students explained only 1.5% of the difference among individuals, but it now captures about 90% of the variation between percentile groups.



**4. Impact of initial starting points in reading shown as simple odds and charts with the percentile groups illustrated with three graphs.** It takes a disproportionate amount of space (five and a half pages) to fully illustrate the relationship between reading scores in kindergarten and Grades 1 through 5, but the point is important enough to justify it. First, we take all of the students together, and ask:

Of the 12,338 students who started kindergarten with reading skills in the **top 20%** (or top quintile), where did these same 12,338 students score in fall of Grade 5?

|       |   |   |
|-------|---|---|
| 6,311 | (51%) were still in the top 20%                                   | <i>Reading approximately at:</i><br>2 years above grade level |
| 3,102 | (25%) were in the 60 <sup>th</sup> to 80 <sup>th</sup> percentile | 1 year above grade level                                      |
| 1,669 | (14%) were in the 40 <sup>th</sup> to 60 <sup>th</sup> percentile | at grade level  |
| 864   | ( 7%) were in the 20 <sup>th</sup> to 40 <sup>th</sup> percentile | 1 year below grade level                                      |
| 393   | ( 3%) were in the 1 <sup>st</sup> to 19 <sup>th</sup> percentile  | 2-3 years below grade level                                   |

Now let's look at the lowest quintile, or the 12,338 students who started kindergarten with reading skills in the **bottom 20%**. Where did these same 12,338 students score in the fall of Grade 5?

|       |   |   |
|-------|---|---|
| 421   | ( 3%) reached the top 20 percent                                  | <i>Reading approximately at:</i><br>2 years above grade level |
| 1,128 | ( 9%) were in the 60 <sup>th</sup> to 80 <sup>th</sup> percentile | 1 year above grade level                                      |
| 2,063 | (17%) were in 40 <sup>th</sup> to 60 <sup>th</sup> percentile     | at grade level  |
| 3,316 | (27%) were in the 20 <sup>th</sup> to 40 <sup>th</sup> percentile | 1 year grade level  |
| 5,410 | (44%) were in the 1 <sup>st</sup> to 19 <sup>th</sup> percentile  | 2-3 years below grade level                                   |

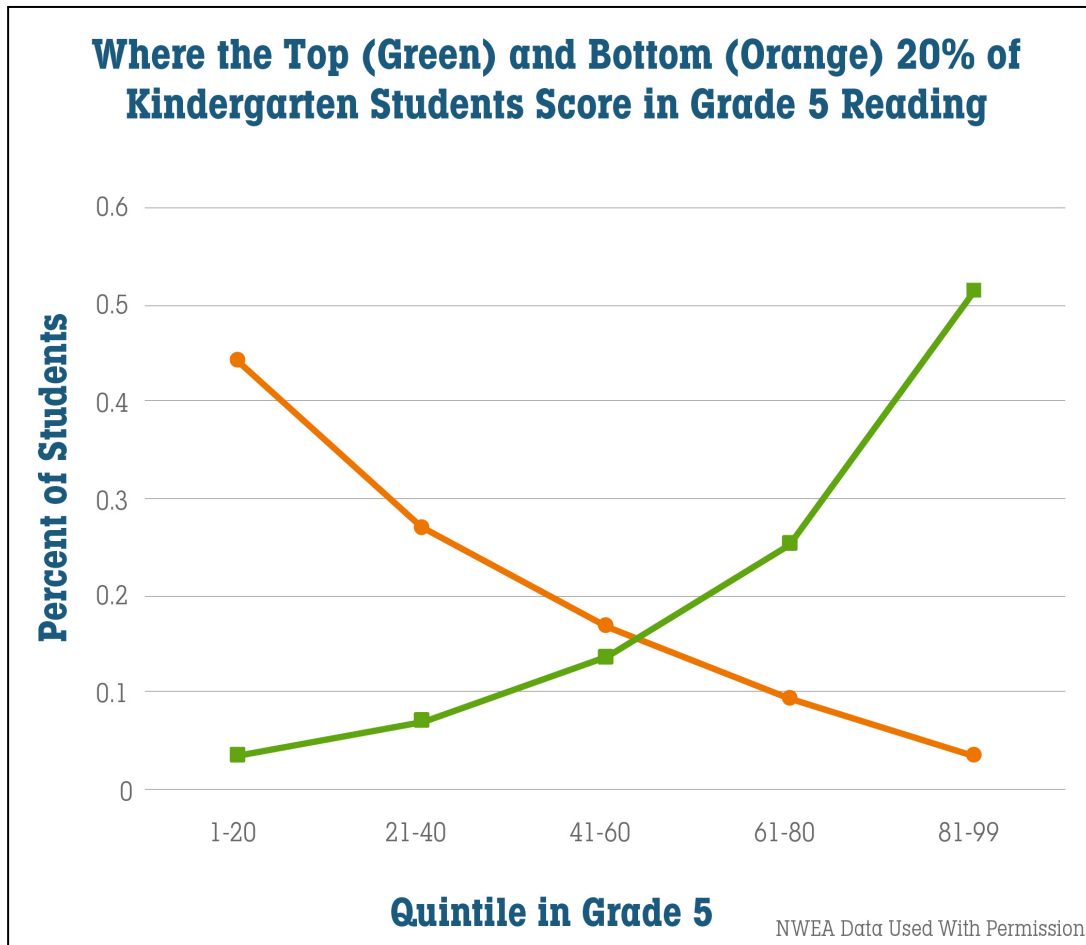
The graph below has a red line representing where students starting in the top 20 percent in kindergarten score in fall of Grade 5, with the blue line which representing those who starting kindergarten in the lowest 20 percent score in Grade 5.

The surprises come on both ends. Many students make significant change. Twenty-nine percent (29%) reach grade level or above. But most students (71%), who start one, two, or three years behind when they enter kindergarten behind are likely to stay behind. The percentage of students who move into a higher quintile gets progressively smaller.

Only 3% of the children who start in the lowest quintile will achieve the highest quintile by the time they reach Grade 5, and only 12.5% will even read above grade level (above the 60<sup>th</sup> percentile) by the fall of Grade 5.<sup>14</sup>

---

<sup>14</sup> The odds are oppressive. A student who enters kindergarten reading 15 percentile points below average is 3.5 times more likely than all other students to still be at least 15 percentile points below average by Grade 5, and 69.2 times more likely to remain behind students who enter 15 points above average.



Clearly, the 3% who come from behind to find placement in the highest quintile are exceptionally talented, exceptionally hard-working, have exceptional teachers and parents, or all of the above. But they *are* exceptions. Most of their classmates will make annual, but not exceptional growth. **Without specific diagnostic intervention, most students who start behind are likely to stay behind.** It's also true that some of the children who start ahead will drift down into the two lower quintiles but, in an almost perfect parallel, only 10% of students who start in the top quintile (above the 80<sup>th</sup> percentile) will have the bad luck (or bad study habits) to end up reading below grade level (that is, below the 40<sup>th</sup> percentile) by Grade 5.

To explain it another way, students who start reading in the top 20 percentiles (or top quintile) are 15 times more likely to still be reading in the top 20 percentiles in Grade 5 than students who start kindergarten in the bottom 20 percentiles (lowest quintile).<sup>15</sup>

What could be the long-term effects of this statistical pattern? The most important is what it means about continued educational opportunities. Of course we recognize that college isn't for everyone. A nation needs plumbers as well as philosophers, mechanics as well as

<sup>15</sup> Calculated to three places, 3.4% of the lowest 20% of entering kindergarten students make it to the top quintile by Grade 5 while 51.2 who started there are still there (3.4% x 15 = 51.0%).

mathematicians. On-the-job training and community college associate degrees are extremely important. But it's tragic if a talented child finds doors to a desired future closing instead of opening. Parents may recognize that their child has remarkable capabilities and assume, "Well, of course he (or she) will go to college." Not necessarily. The United States has only one seat at a four-year university for every four graduating high school seniors. The top 10% have the best odds of being admitted to a four-year college. Few seats go to students who rank in the 50th percentile.

The chart below shows all 61,691 students sorted by their reading scores from highest to lowest, divided into five groups (quintiles) of 12,338 each. (The right-hand panel contains the percentages you saw in the first chart.) A useful element of this chart is that it shows the raw numbers as well as the percentages for each kindergarten group (in left-to-right columns) in addition to the up-and-down rows that show the quintile of the students' scores in the fall of Grade 5.

| Movement of 61,691 Students from Initial Kindergarten Reading Quintiles to the Fall of Grade 5 |                     |                   |                                   |               |               |                |                   |                                   |       |       |                |
|--|---------------------|-------------------|-----------------------------------|---------------|---------------|----------------|-------------------|-----------------------------------|-------|-------|----------------|
|  | Grade 5 Percentile: | Raw numbers       |                                   |               |               |                | Percentages       |                                   |       |       |                |
| Fall of Grade 5  | 81-99               | 421               | 959                               | 1,714         | 2,934         | 6,311          | 3%                | 8%                                | 14%   | 24%   | 51%            |
|  | 61-80               | 1,128             | 1,991                             | 2,779         | 3,338         | 3,102          | 9%                | 16%                               | 23%   | 27%   | 25%            |
|  | 41-60               | 2,063             | 2,742                             | 2,997         | 2,866         | 1,669          | 17%               | 22%                               | 24%   | 23%   | 14%            |
|  | 21-40               | 3,316             | 3,328                             | 2,809         | 2,022         | 864            | 27%               | 27%                               | 23%   | 16%   | 7%             |
|  | 0-20                | 5,410             | 3,318                             | 2,040         | 1,177         | 393            | 44%               | 27%                               | 17%   | 10%   | 3%             |
|  | <b>K</b>            | <b>12339</b>      | <b>12,338</b>                     | <b>12,338</b> | <b>12,338</b> | <b>12,338</b>  | 100%              | 100%                              | 100%  | 100%  | 100%           |
|  | K Percentile:       | 1-20              | 21-40                             | 41-60         | 61-80         | 81-100         | 1-20              | 21-40                             | 41-60 | 61-80 | 81-100         |
|  |                     | <b>Bottom 20%</b> | <b>Where students scored in K</b> |               |               | <b>Top 20%</b> | <b>Bottom 20%</b> | <b>Where students scored in K</b> |       |       | <b>Top 20%</b> |

Of the 12,338 students who start kindergarten at grade level or between the 41<sup>st</sup> to 60<sup>th</sup> percentile, the outcomes are fairly balanced. Twenty-four percent (24%) or 2,997 of the students are still at grade level, with 37% moving upward into the fourth quintile (23%) and the top quintile (14%). A near equal number of students move downward to the second quintile (23%) and the bottom one (17%).

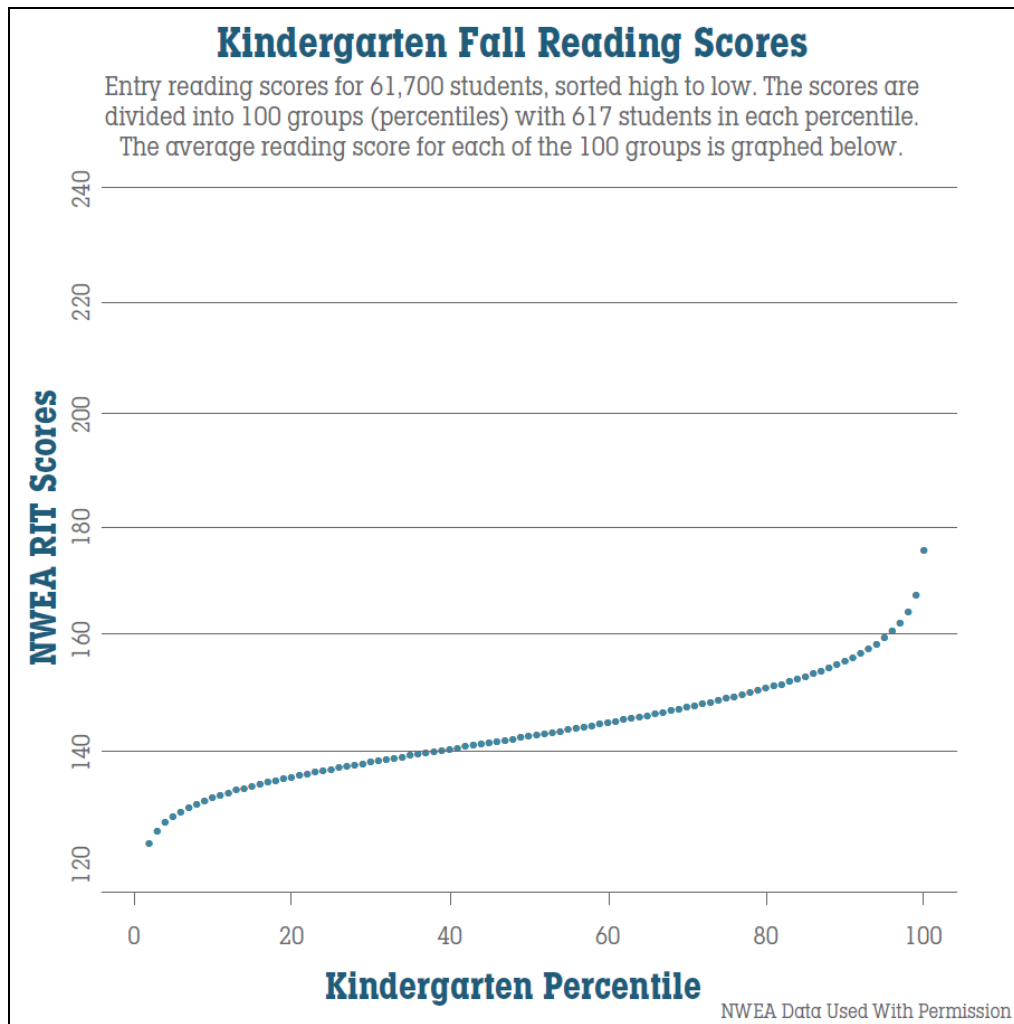
The patterns describe student movement at the national level. In many schools, student movement is consistently upward with greater than national normed annual growth. That upward movement is offset on consistent downward movement in other schools. John Cronin and Branin Bowe have written a particular insightful study showing the ranking of schools by the percent of students making national annual growth by subject.<sup>16</sup>

<sup>16</sup>See John Cronin and Branin Bowe, "Study of Growth Index Performance by School," NWEA research report (January 2006), relevant tables starting at page 8

The following three graphs show how the reading scores of entering kindergartners impact later scores. The statistical and mathematical data of the comparison for Figure 1-5 are shown in Appendix D.

**Figure 1** illustrates the sorting of the 61,691 fall reading kindergarten scores from high to low, their division into 100 groups of 617 students each, averaging their scores (at this point the range in the scores within each group is very small) and then graphing the average of each percentile. These 100 groups consisting of 617 each are called the percentile groups.

**Figure 1**

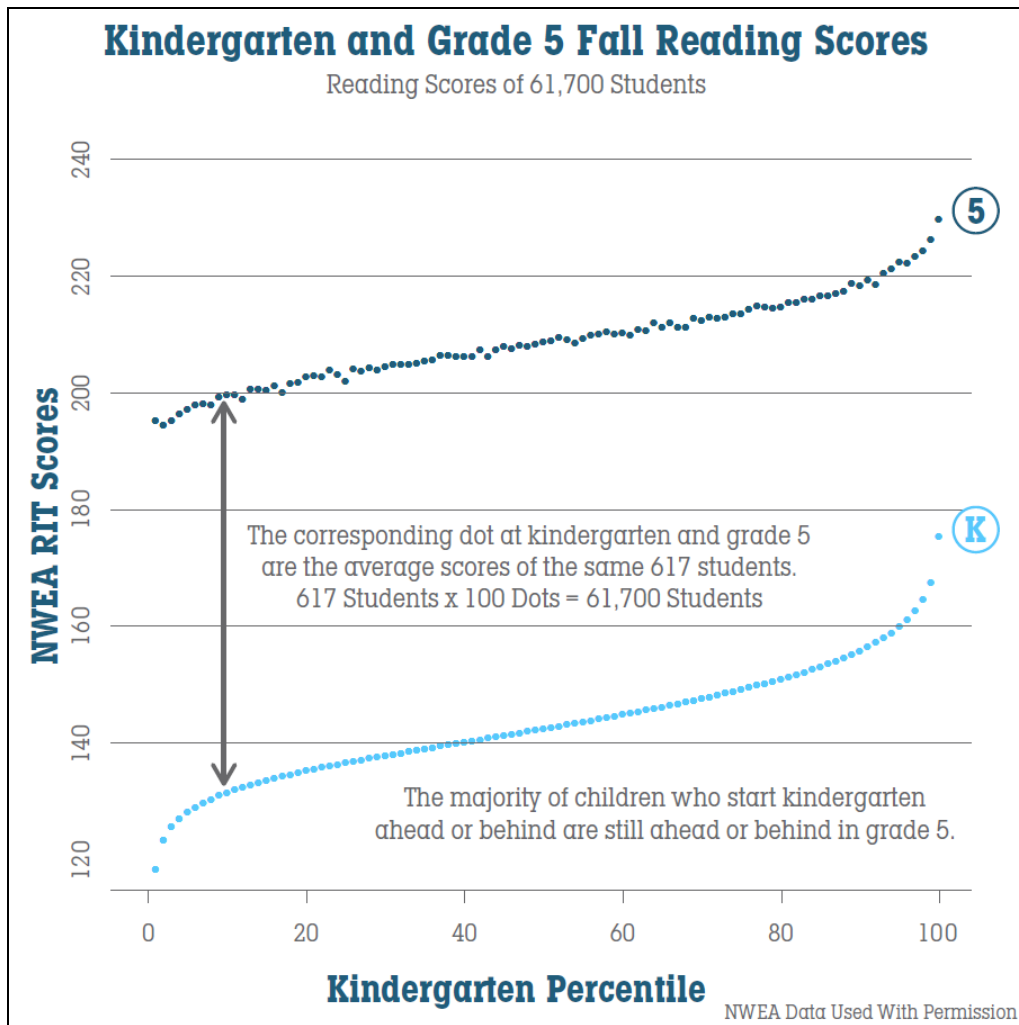


**Discussion:** Each of the dots represents one of the hundred percentile groups of 617 students. Ignoring the top and bottom percentiles, the rest form a fairly straight line with an upward slope of about one-third of a RIT point per percentile. The lower four points of the tail drop about nine RIT points, the upper four points rise about 13 RIT points. The range

between the very bottom students in RIT points (y-axis, 118.7) and the very top students (175.6) is slightly less than 70.

**Figure 2** compares the fall kindergarten scores we have previously seen in Figure 1 to the fall test scores in Grade 5. The light blue dots are exactly the same as the Figure 1 dots. Each dark blue dot represents the average reading score in Grade 5 of the same students who comprised the light blue dots just five years later.

**Figure 2**



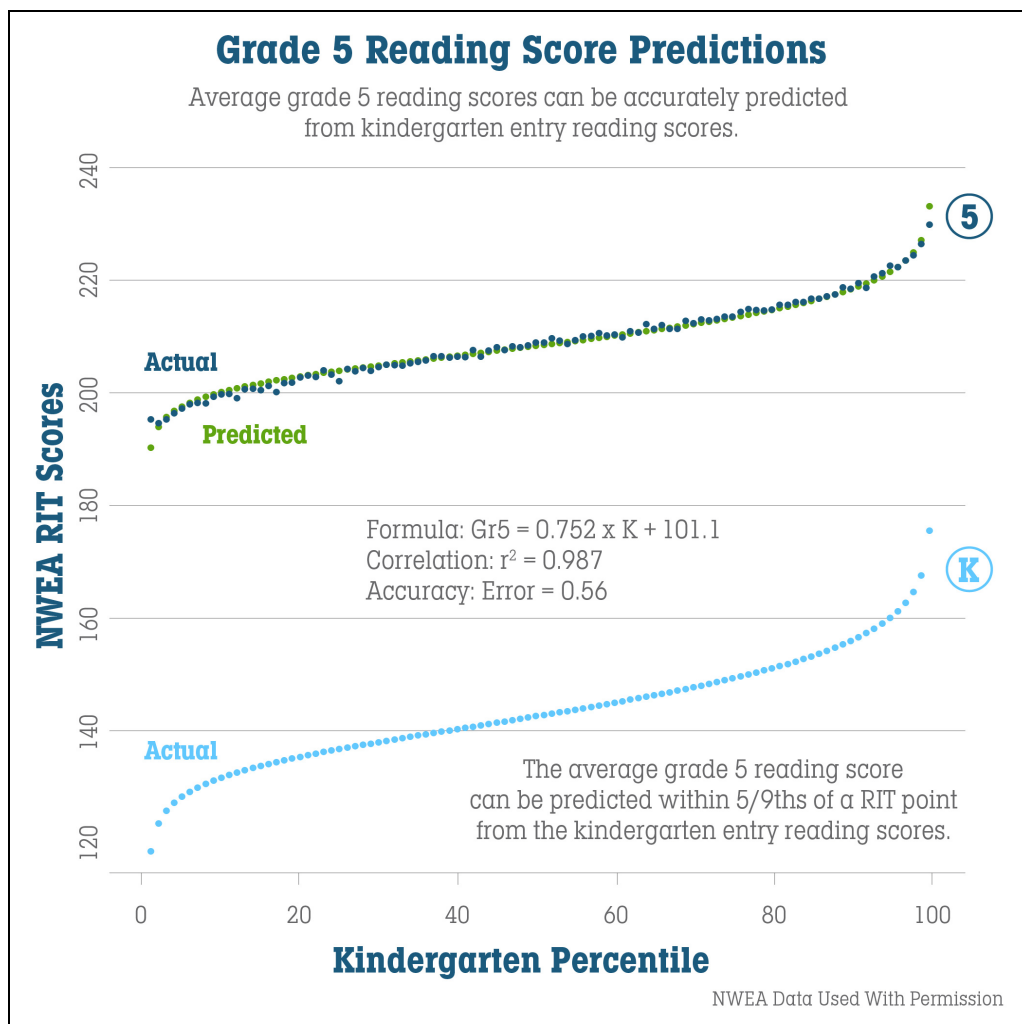
**Two Observations:**

- The parallel between grades is so close that the Grade 5 percentile groups look like a copy of the kindergarten percentile groups, just 65 RIT points higher. You have to look hard to see the few differences.
- One of these differences is that the Grade 5 line doesn't have the same "tails" as the kindergarten line. The bottom four points of the kindergarten line have moved up and the top four points have moved down, making a straighter-looking curve. This means

that the very lowest-scoring students have made significant progress, while students at the very top, though continuing to hold their high ranking, are making slower progress in relation to the general “line” formed by the percentile groups.

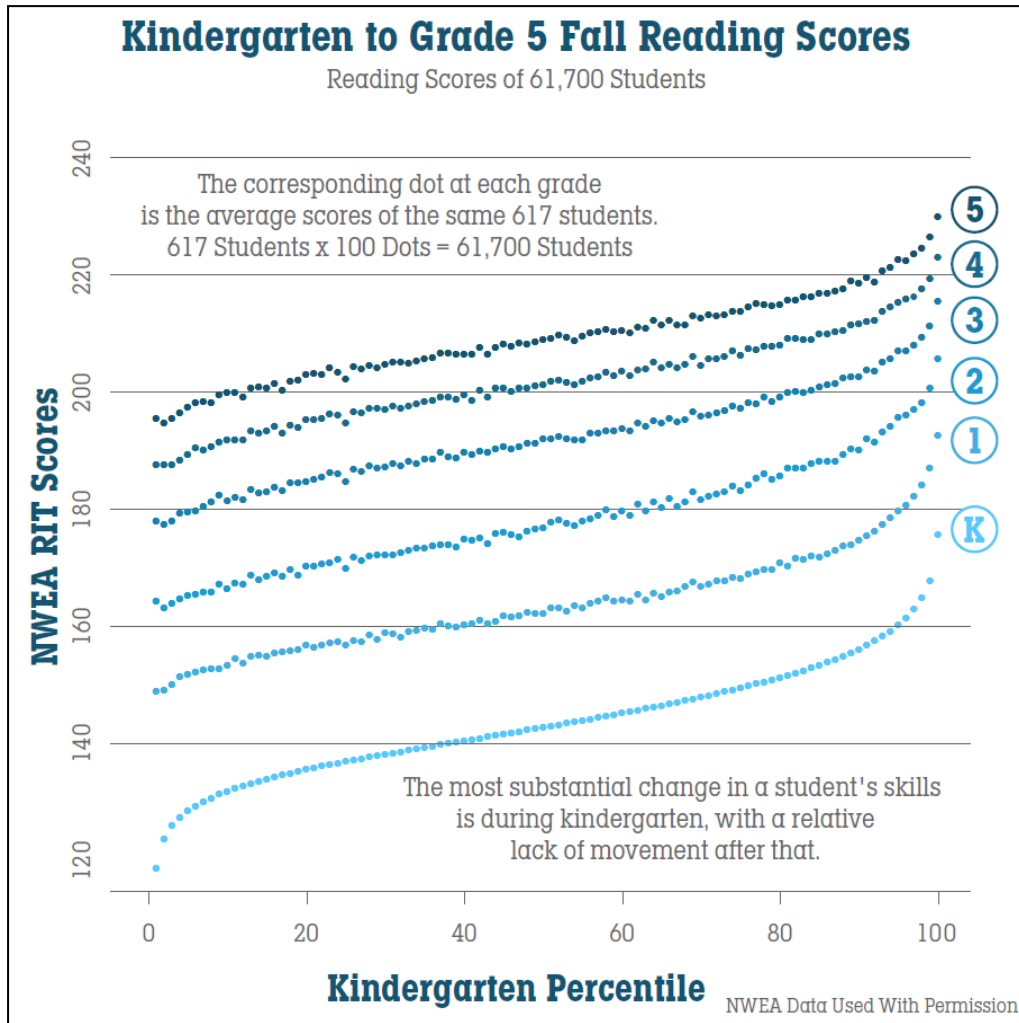
**Figure 3** contrasts the actual (red) versus the predicted (hollow) scores of tests administered in the fall of Grade 5. These predictions were made using a relatively simple statistical procedure called “linear regression”. The predictions have a 0.993 correlation with the actual test scores, and the figure gives you a good sense of what such a high correlation looks like: Most of the red data circles actually lie within the hollow prediction circles. In fact, the average score of most percentile groups can be predicted to within 5/9ths of a RIT point.

**Figure 3**



**Figure 4** shows the percentile groups' reading scores from the fall of kindergarten through each grade level up to the fall of Grade 5. The grade level of each band is labeled with a number in a circle just to the right of the last dot of each group.

**Figure 4**



**Three Observations:**

- There is a bigger gap between the kindergarten line and the Grade 1 than any of the other lines. This gap highlights that more growth occurs during kindergarten than during any other grade, especially at the lower percentiles. This pattern is, in part, due to the fact that “annual” or “normal growth” is not measured equally across the years. The NWEA assigns 18 RIT points of growth during kindergarten and gradually decreases annual growth to 8 points at Grade 5.<sup>17</sup>

<sup>17</sup> The RIT points for average growth at the 50th percentile are K-1 = 18; G1-2 = 16; G2-3 = 14; G3-4 = 9; G4-5 = 8; G5-6 = 5; G6-7 = 4; G7-8 = 3; G8-9 = 2; G9-10 = 2; G10-11 = 0.

- There is a slight relative increase in growth during Grade 1 between the 50<sup>th</sup> and 90<sup>th</sup> percentiles that is offset by a decrease in growth during Grade 2.
- By covering up the four to five percentiles at the tails (the four or five dots at the far left and far right), we get lines that are almost perfect matches with each other through the years.

**5. Impact of initial starting points in *reading* using math as an additional variable in a two-step reading prediction model.** The largest variation between the percentile groups appears in Figure 4 between the fall of kindergarten and fall of Grade 1. In fact, by merely using the spring of kindergarten (instead of the entering fall scores) to predict later reading scores, the correlations improve substantially at each grade level.

| Grade Predicted              | 1      | 2      | 3      | 4      | 5      |
|------------------------------|--------|--------|--------|--------|--------|
| Using Fall of Kindergarten   | 0.6913 | 0.5807 | 0.5494 | 0.5388 | 0.5307 |
| Using Spring of Kindergarten | 0.8219 | 0.6706 | 0.6332 | 0.6134 | 0.6042 |

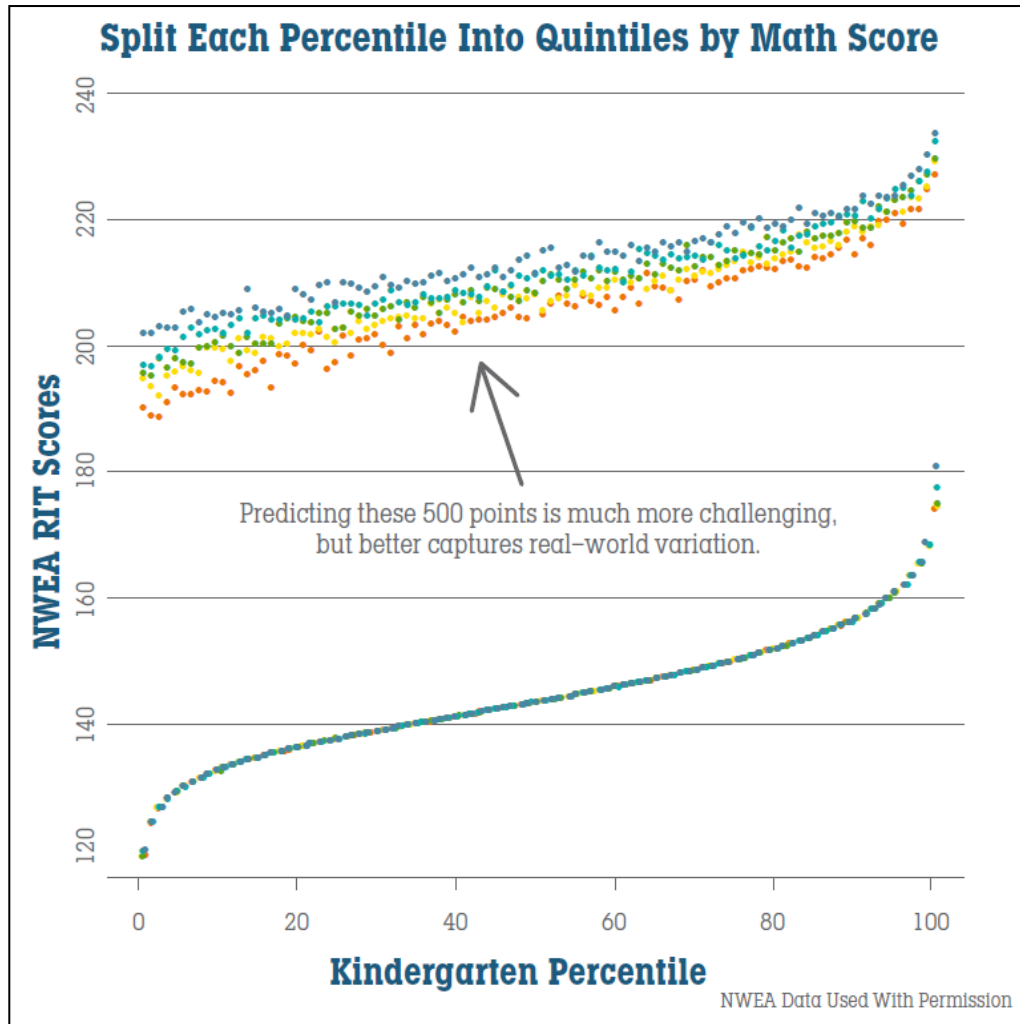
The increased correlation runs from 0.13 in Grade 1, to about 0.07 across the later grades. The disadvantage of using spring data to project Grade 5 scores is the requirement of waiting through the school year before the tests are administered and the scores become available. However, we found that if we used fall kindergarten math scores as a second variable with kindergarten fall reading scores, as described in Part 1.5 the correlation on the individual level increased from almost as much--from 0.531 to 0.571. In other words, we could get nearly the same level of predictability, without the nine-month wait.

Figure 5 shows how this happens at the group level. First, we took the 100 dots from kindergarten (see Figure 1), each representing 617 students with about the same level of reading ability. Second, we subdivided them further into five groups based on their math ability, for a total of 500 groups of 123 students each. On the graph below, the 100 “highest math ability” subgroups are color-coded blue, the 100 “high math ability” groups are teal, the 100 “average math ability” subgroups are green, the 100 “low math ability” groups are yellow, and the 100 lowest math ability groups are orange. In kindergarten these dots are almost indistinguishable from each other because their reading skills are so close. Not so in Grade 5, however.

You can see that by Grade 5, these formerly overlapping groups have spread out into distinct bands, with the “lowest math ability” group (orange dots) consistently lagging behind the other four, and the “highest math ability” group (blue dots) coming out on top. This is something you can see that by Grade 5, these formerly overlapping groups have spread out into distinct bands, with the orange “lowest math ability” group consistently lagging behind the other four, and the blue “highest math ability” group coming out on top.

This is something of a surprise, since there's no obvious reason to think that, if we take two kindergartners who read at the same level, the one who is better at math will also be better at reading five years later. And yet, so say the data.

Figure 5



Because all the five math groups had the same kindergarten reading score, attempting to predict their Grade 5 reading score was harder because the group size was smaller (now 123). Using only kindergarten reading scores, the correlation fell from 0.993 to 0.924. However, when we predict Grade 5 reading using both reading *and* math scores, the correlation increases to 0.986. The statistical and mathematical data of the comparison for Figure 5 are shown in Appendix D.

## Observations:

- Using this enhanced dataset, the prediction of reading scores in Grade 5 becomes even more accurate. Within every percentile, the five subgroups in kindergarten have, with very few exceptions, separated themselves in a clear pattern by Grade 5. The blue (highest math) dots are always on top and the orange (lowest math) dots are always on the bottom in every percentile at Grade 5, even though they started out in the same place. For the most part, the middle three quintiles have also taken their place in order between the blue and orange. When we compare all 500 points, only 42 quintiles (8.4%) do not follow this pattern. For example, we took the students in the 21st percentile of kindergarten reading (RIT = 135.8) and split them into five groups by kindergarten math. The top group averaged +13.3 points better on math than reading (RIT=149.1); the bottom group averaged -7.3 points worse (RIT=128.5). By Grade 5, the second quintile had pulled slightly ahead of the third quintile in reading (RIT 203.8 vs. 203.4). This quintile's activity was one of very few exceptions, however; and even here, the top quintile had a reading score of 207.6 by Grade 5, and the bottom quintile had reached only 199.

- 6. Comparison to poverty correlations.** As we discussed in Part I, correlations between individual students in poverty and their test scores typically lie in the range of 0.22 and 0.31.<sup>18</sup> Correlations in this range are not considered particularly strong. By contrast, correlations between initial starting points in reading and math with later performance are strong. Even years later, in Grade 5, the correlation is almost twice as high (0.571) when both kindergarten reading and math scores are used than it is with poverty.

In this study, we used groups as small as 123 and still found a correlation between starting point and Grade 5 achievement in the range of 0.98. Group studies of poverty and low academic achievement commonly cite a correlation of 0.66-0.68. However, these studies involve large groups in schools, districts, or states, groups far larger than the 123 or even 617 student groups used in our correlations. These correlations are repeatedly produced because free-or-reduced lunch data (considered good measurements of the poverty in a particular group) are readily available publicly for each of these groups, while entering kindergarten scores have not been available until the last ten years and are still rarely available to the public or researchers.

---

<sup>18</sup> When achievement scores and SES are measured *individually* for all children in a large sample is 0.23 in White's (1982) meta-analysis of 174 samples, 0.22 in a sample of 1,459 9-year-olds (Walberg and Tsai, 1985), 0.31 in another meta-analysis of prediction studies (Horn and O'Donnell (1984). The strength of the association ranges between .22 and .31 whereas the strength of association when the unit of measure is a school or district in a large sample, the strength of association is in the range of 0.68. Snow, Burns, Griffin, eds., *Preventing Reading Difficulty in Young Children* (1998), 125-127.

We are less interested in poverty correlations than earlier researchers because such correlations direct attention to the socio-economic status of students' families which is beyond the effect control of schools or districts. By contrast, focusing on the skills of entering kindergartners generally alerts families and educators to the factors of time-on-task, quality instruction and quality curriculum that drive both pre- and post-K learning.

Clearly poverty is associated with lower academic skills, but its weaker correlation on both the individual and group levels suggests that policymakers should pay attention to the time that children approaching kindergarten should spend on math and literacy skills. In the future, school districts could improve the scores of K through Grade 5 students more by encouraging parents and child-care providers to spend more time on language and math skill acquisition before kindergarten.

- 7. Impact of kindergarten scores in *math* expressed as odds of subsequent achievement, charts, and graphs.** In this section, we explore alternative ways to view the associations between math scores in kindergarten and Grade 5. First, we took the 10,356 students who entered kindergarten scoring in the **top 20%** of math skills and linked them to their math scores in the fall of Grade 5.

*K math skills at approximately:*

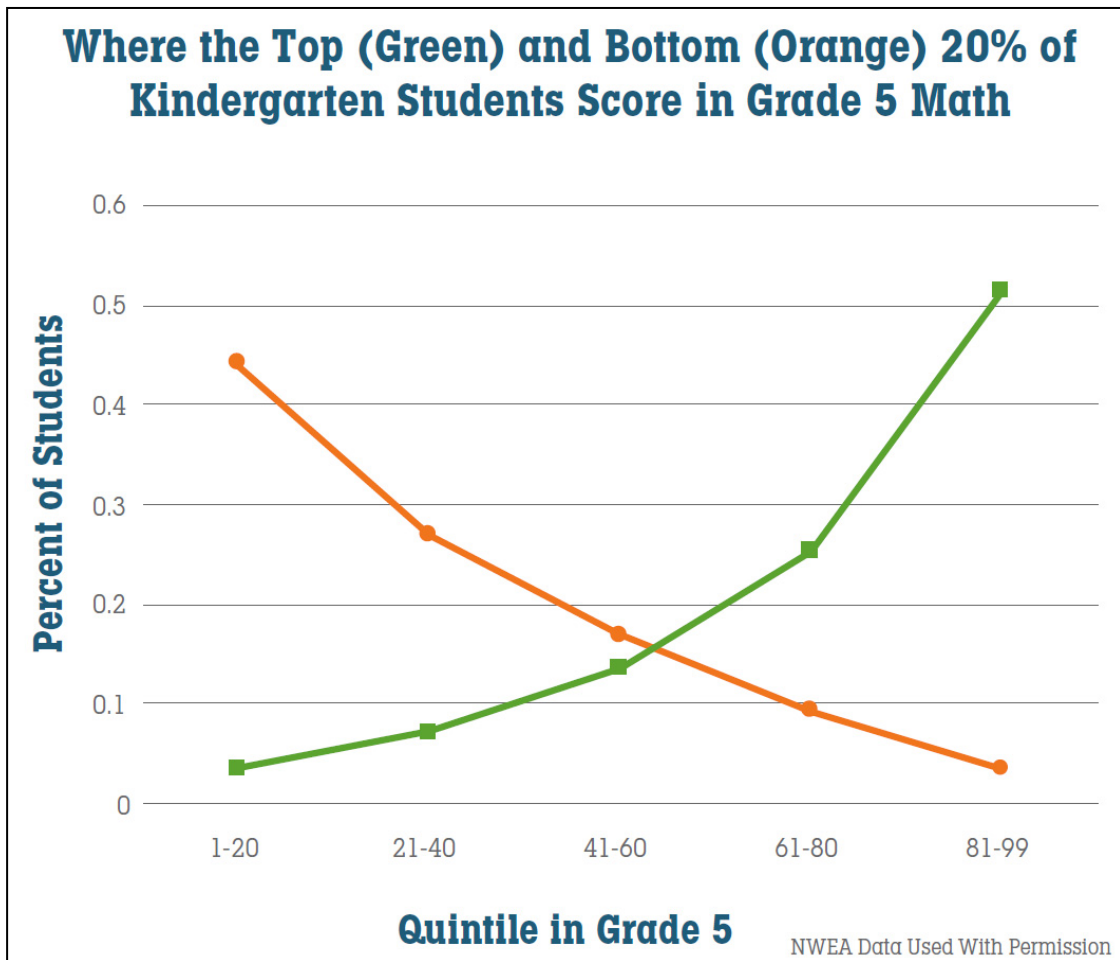
|       |   |                             |
|-------|---|-----------------------------|
| 5,491 | (53%) were still in the top 20%                                   | 2 years above grade level   |
| 2,635 | (25%) were in the 60 <sup>th</sup> to 80 <sup>th</sup> percentile | 1 year above grade level    |
| 1,376 | (13%) were in the 40 <sup>th</sup> to 60 <sup>th</sup> percentile | at grade level              |
| 650   | ( 6%) were in the 20 <sup>th</sup> to 40 <sup>th</sup> percentile | 1 year below grade level    |
| 202   | ( 2%) were in the 1 <sup>st</sup> to 19 <sup>th</sup> percentile  | 2-3 years below grade level |

What about the 10,356 students who started kindergarten with math skills in the **bottom 20%**? How were they faring in in the fall of Grade 5?

*K math skills at approximately:*

|       |   |                             |
|-------|---|-----------------------------|
| 381   | ( 4%) reached the top 20 percent                                  | 2 years above grade level   |
| 848   | ( 8%) were in the 60 <sup>th</sup> to 80 <sup>th</sup> percentile | 1 year above grade level    |
| 1,511 | (15%) were in 40 <sup>th</sup> to 60 <sup>th</sup> percentile     | at grade level              |
| 2,635 | (25%) were in the 20 <sup>th</sup> to 40 <sup>th</sup> percentile | 1 year below grade level    |
| 4,981 | (48%) were in the 1 <sup>st</sup> to 19 <sup>th</sup> percentile  | 2-3 years below grade level |

The pattern in math is very similar to the pattern of reading scores. The top and bottom quintiles (20%) are virtually a mirror (inverse) image of each other.



A few students (4%) will climb from the lowest into the highest quintile, but only 11.9% of students who start in the bottom quintile (1-20<sup>th</sup> percentiles) will read above grade level (above the 60<sup>th</sup> percentile) by the fall of Grade 5.<sup>19</sup>

In other words, most students who start behind stay behind. The percentage of students climbing up to the next quintile (or beyond) diminishes with each quintile. Some of the children who start ahead will drift down into lower quintiles but only 8% of students who start in the top quintile (80-99<sup>th</sup> percentile) will read below grade level (below the 40<sup>th</sup> percentile) by Grade 5. Students who start reading in the top 20 percentiles (the top quintile) are a little over 16 times more likely to still be reading in the top 20 percentiles in

<sup>19</sup> A student who enters kindergarten reading 15 RIT points below average is 3.5 times more likely than all other students to still be at least 15 RIT points below average by Grade 5, and 69.2 times more likely to remain behind students who enter 15 points above average. Similarly, students who score 15 or more RIT points below average on the kindergarten math test are 3.6 times more likely than all other students to still be 15 or more points below average in Grade 5, and 72.3 times more likely than those who entered kindergarten 15 RIT points or more above average.

Grade 5 than students who start kindergarten in the bottom 20 percentiles (the bottom quintile).<sup>20</sup>

Like the chart in Part II-4 that shows students sorted by reading scores, we sorted all 51,781 students in the total group by their math scores from highest to lowest. Then we divided them into five groups of 10,356 each and tracked them through to the fall of Grade 5.

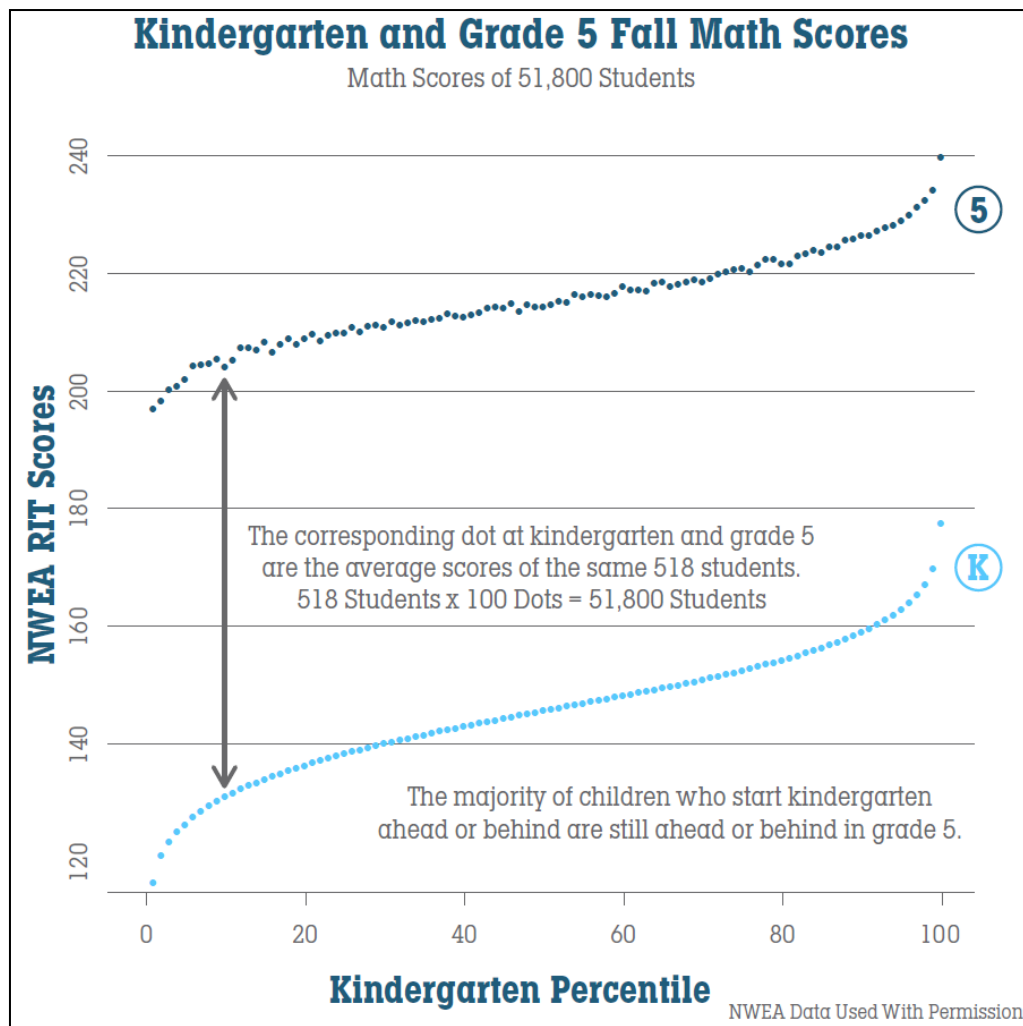
**Movement of 51,781 Students from Initial Kindergarten Math Quintiles to Fall of Grade 5 in Raw Numbers**

|                                  |       | Raw Numbers |                                       |       |       |         | Percentages |                                       |       |       |         |
|----------------------------------|-------|-------------|---------------------------------------|-------|-------|---------|-------------|---------------------------------------|-------|-------|---------|
|                                  |       | 1-20        | 21-40                                 | 41-60 | 61-80 | 81-100  | 1-20        | 21-40                                 | 41-60 | 61-80 | 81-100  |
| Where students scored in Grade 5 | 81-99 | 381         | 710                                   | 1,370 | 2,405 | 5,491   | 4%          | 7%                                    | 13%   | 23%   | 53%     |
|                                  | 61-80 | 848         | 1,596                                 | 2,264 | 3,013 | 2,635   | 8%          | 15%                                   | 22%   | 29%   | 25%     |
|                                  | 41-60 | 1,511       | 2,407                                 | 2,627 | 2,432 | 1,379   | 15%         | 23%                                   | 25%   | 23%   | 13%     |
|                                  | 21-40 | 2,635       | 2,877                                 | 2,465 | 1,729 | 650     | 25%         | 28%                                   | 24%   | 17%   | 6%      |
|                                  | 0-20  | 4,981       | 2,766                                 | 1,630 | 777   | 202     | 48%         | 27%                                   | 16%   | 8%    | 2%      |
|                                  |       | 1-20        | 21-40                                 | 41-60 | 61-80 | 81-100  | 1-20        | 21-40                                 | 41-60 | 61-80 | 81-100  |
|                                  |       | Bottom 20%  | Where students scored in kindergarten |       |       | Top 20% | Bottom 20%  | Where students scored in kindergarten |       |       | Top 20% |

- Impact of initial starting points in math using reading as an additional variable.** In this section, we use the scores of kindergarten math *and reading* tests to predict Subsequent math scores. The results are more modest. Math-to-math correlations are typically 0.03-0.06 points higher than reading-to-reading correlations. In other words, the use of reading as a second variable has a far weaker impact on math correlations than using math as a second variable had on reading correlations. Small group predictability improved from 0.969 to 0.989, while individual predictability improved from 0.594 to 0.608.

The graph below for math percentile groups looks surprisingly like the graph above for reading percentile groups. We divided the 51,781 entering kindergartners into 100 groups of 517 or 518 students with similar ability in math. Tracking these groups' average math performance through Grade 5 shows that, just like our reading analysis, students who start behind tend to stay behind, while students who start ahead tend to stay ahead.

<sup>20</sup> About 3.2% of the lowest 20% of entering kindergarten students make it to the top quintile by Grade 5 while 51.2 who started there are still there, for a ratio of  $0.5115 / .0319 = 16.0$ .



The reading/math relationship. To a substantial degree, the students at the ends of both the reading and math continuums are the same students. Most children (88%) who enter kindergarten reading in the highest 20 percent also score in the highest (64%) or the second highest (24%) 20 percent in math. The majority of children (85%) who read in the lowest 20 percent also score in the lowest (60%) or the second lowest (25%) 20 percent in math.

9. **Free increases in math scores when reading scores increase.** An interesting fact turned up in Jay Maidment’s (2005-06) non-published study of the Washington State student scores was corroborated his and Fielding’s their joint work on the Elgin project (2011-13). Improvement in reading scores during Grades 1-5 was accompanied by “free” improvements in math scores, even though teachers were giving math no additional instructional time or emphasis. While we don’t have information on which students received special intervention in reading or math in this data set, we did find that for every

two points students improved in reading, they improved, on average, one point in math. (See Part III.9.)

**10. Science correlations.** Science scores show only a marginal amount of differentiation between Grade 5 and Grade 8. Of the 6,949 students who took both tests, there was a correlation of 0.749 between the two scores. A linear model indicates that students could expect to improve 45.3 RIT points between tests and keep 5/6ths of every point they were above/below the mean in Grade 5. Performance on the science score in Grade 5 test can be forecast from Grade 1 with 35% accuracy ( $r^2$ ) using the test scores from both reading (15.5%) and math (18.5%) in Grade 1. Backing up a year to kindergarten test scores lets us predict science performance with only slightly less accuracy ( $r^2=28.7%$ ); 13.2% of the predictability comes from entering kindergartners' reading ability and 15.5% from math.

Gender and ethnicity had only a slight effect on science test scores, but they were overwhelmed by math and reading ability. We used four variables to predict science scores in Grade 5: math and reading in Grade 5, ethnicity (eight factors), and gender (discussed above). This model showed that science scores were ~60% predictable ( $r^2=.590$ ): 50.8% comes from math, 45.8% from reading, and a mere 3.4% from ethnicity and gender combined.

**11. Relative impact studies.** Although different researchers and educators make a case for which factors are the "most important" in ensuring reading success, our study, we feel, makes the strongest case for one simple factor: entering kindergarten reading at or above grade level.

**12. Testing for other factors.** We also used the larger dataset ( $n=310,000$ ) to examine the possible influence of five other factors that are considered important demographic influences.

**a.** Gender: Gender was not a relevant variable once we factored out females' head start.

**b.** The average female student in the group of 310,000 was 0.42 RIT points ahead of the average male by Grade 5. However, once we took into account the impact of their entering kindergarten scores, the increase dropped to 0.01 RIT points. Ethnicity: Ethnicity is a variable between Ethnicity 1 and 2 but as with gender, these achievement gaps were substantially a result of head starts, rather than performance in school. Again, a simple average suggested that ethnicity accounted for dis/advantages for students in Grade 5. Taking the extremes, Ethnicity 1 (African) lagged 16 points behind Ethnicity 2 (Asian) in reading and 18 points in math. However, the explanatory power of this finding was low ( $r^2<0.03$ ) due to large variations within each ethnicity. Simply removing the achievement gap in Grade 1 from the achievement gap in Grade 5 reduced the Grade 5 differences from 16 to less than 3.5 RIT points in both reading and math. Comparing the eight ethnicities to each other (1 vs. 2, 1 vs. 3, 1 vs. 4, etc., for a total of 28 comparisons) in two subjects ( $28 \times 2 = 56$  comparisons) without the head start showed only trivial differences in 55 cases; but the 56th comparison-- the widening gap in math between Ethnicity 1 and 2-- is extremely significant: Grade 1 scores are approximately 13 times more influential in predicting Grade 5 scores than ethnicity; 93%

of the explanatory power lies in Grade 1 scores—more specifically, 92.6% for reading and 93.0% for math.

- c. Socio-economic status (SES): The NWEA data does not provide socio-economic data.
- d. School data: The variation between schools in annual growth was very small compared to the variation in annual growth between individual students in the same school.
- e. State: With the exception of Kentucky, state and region were not strong predictor variables. Lynn Fielding, at the Elgin Foundation’s invitation, provided extensive intervention to the school districts of interest in Kentucky.<sup>21</sup>

**13. Causation:** This single study of correlations of initial starting points in kindergarten in reading and math and subsequent achievement at grades 1-5 does not rise to the level of statistical causation.

**That being said,** educators deal with starting points all the time in K-12. Starting points are merely test scores and one teachers ending points are the next teachers starting points. Test scores in the fall of kindergarten are no different. They merely measure the amount of growth in two academic areas since birth.

No educator believes that testing creates *growth in learning* but beginning academic status affects outcomes consistently and over time. A student two years below grade level in math (and another two years behind the top 20% of the class) will have to make 200% of normal growth for two years be at grade level and 200% of normal growth for another two years to reach the top quintile. That same student will have to make 150% of normal growth for eight years to reach the top quintile. And a huge focus on elementary teachers for the last decade has been to create that accelerated growth for behind-students, especially in reading. At the same time, a student starting in the top quintile can remain there indefinitely merely by making normal annual growth each year.

The response to these finding is obvious and perhaps some of the foremost researchers in the area of the impact of early academic status on subsequent academic achievement, said it best. **“The most powerful pre-school avenue for boosting fifth-grade achievement appears to be improving the basic academic skills of low-achieving children prior to kindergarten entry.”**<sup>22</sup>

---

<sup>21</sup> The Elgin Foundation funded significant programs: the adoption of a 90% third grade reading goal, extensive community outreach for children from birth to age 5, required NWEA testing at fall, winter and spring for Grades K-3, a uniform curriculum, uniform increases in instructional time for groups of children who were equally below grade level, and extensive academic coaching for superintendents and school boards, for principals, and for teachers

<sup>22</sup> Amy Claessens, Greg Duncan, Mimi Engel, “Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K” Economics of Education Review Volume 28, Issue 4, August 2009, Pages 415-427

# PART III

## NWEA DATA, STATISTICAL ANALYSIS, AND MATHEMATICAL MODELS

Note: This section uses the same topics, numbered in the same order, as Part I and Part II.

1. **Range at kindergarten.**
  
2. **When are optimum periods of improvement in reading and math?** To answer this question, for each of the students in the reading sample, we calculated how many RIT points above/below average they were in kindergarten and compared that result to how many RIT points above/below average they were in Grade 8. The difference between the two was considered “total movement”. For example, a student who was 5.3 RIT points above average in kindergarten, but 9.3 RIT points above average in Grade 8 would have a total movement of 4.0 RIT points. Next, we considered how far the student had moved in each grade between kindergarten and Grade 8 as a proportion of the total movement. If this student had scored 4.3 RIT points above the mean in Grade 1, it would count as a movement of 1.0 RIT point since kindergarten, and 25% of the total movement. Once these proportions were calculated for every student in the sample, we average them over all the students in the grade to get an average movement for the entire grade.

To validate these results, we used four additional metrics to measure movement: class rank, percentile, decile, and standard deviations (z-scores). Results were quantitatively similar.

| <b>Mean Proportion of Reading Movement between Kindergarten and Grade 8 Accomplished at Each Grade</b> |        |        |        |        |        |        |        |   |
|--|--------|--------|--------|--------|--------|--------|--------|---|
| <b>GRADE</b>   | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8 |
| <b>RIT above avg</b>   | 0.7309 | 0.7524 | 0.8980 | 1.1020 | 0.9765 | 1.0422 | 0.9589 | 1 |
| <b>Rank</b>  | 0.8590 | 0.8405 | 0.9256 | 0.9747 | 0.9886 | 0.9929 | 0.9987 | 1 |
| <b>Percentile</b>  | 0.8590 | 0.8404 | 0.9256 | 0.9746 | 0.9886 | 0.9928 | 0.9987 | 1 |
| <b>Decile</b>  | 0.8596 | 0.8404 | 0.9254 | 0.9748 | 0.9899 | 0.9939 | 0.9985 | 1 |
| <b>SDs</b>   | 0.8308 | 0.8169 | 0.9171 | 0.9703 | 0.9890 | 0.9959 | 0.9966 | 1 |

Note that, because the standard deviation of RIT scores is larger in Grade 4 than Grade 8, an average student has actually moved more RIT points in Grades 4 and 6 than Grade 8.

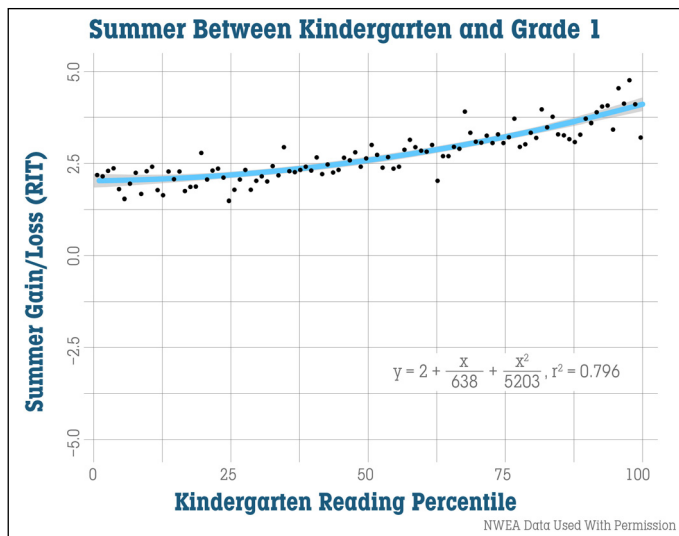
The analysis was also performed on the math sample, with similar results.

| Mean Proportion of Math Movement between Kindergarten and Grade 8 Accomplished at Each Grade |       |       |       |       |       |       |       |   |
|--|-------|-------|-------|-------|-------|-------|-------|---|
| GRADE  | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8 |
| RIT above avg  | 0.910 | 0.898 | 0.921 | 0.989 | 0.980 | 0.966 | 0.995 | 1 |
| Rank   | 0.872 | 0.851 | 0.908 | 0.983 | 1.004 | 0.985 | 1.013 | 1 |
| Percentile   | 0.872 | 0.851 | 0.908 | 0.983 | 1.004 | 0.985 | 1.013 | 1 |
| Decile   | 0.873 | 0.851 | 0.909 | 0.983 | 1.003 | 0.985 | 1.013 | 1 |
| SDs  | 0.875 | 0.854 | 0.914 | 0.989 | 1.008 | 0.990 | 1.011 | 1 |

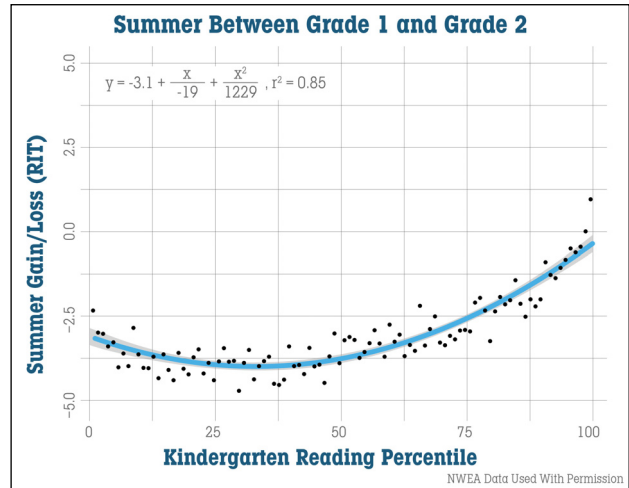
3. **Summer gain and loss.** As indicated in Parts 1.6 and 2.6, reading scores generally decline over the summer. Except for kindergartners moving into Grade 1, the fall test scores are, on average, a few points lower than the previous year’s spring test score. However, students who entered kindergarten with few reading skills tend to decrease more than peers who entered kindergarten with well-developed reading skills. The exact shape of this distribution varies from year to year, but the net effect (calculated as the sum of all summer gains/losses) is that students who entered kindergarten below the 50th percentile lose 4-6 RIT points over the five years of summer vacations, while students who entered kindergarten above the 80th percentile gain from 0-6 RIT points during their five years of summer vacations, depending on how much higher than the 80th percentile they started.

In the figures below, averages are calculated for each percentile (100 groups of 617 students). It should be noted that there is an exceptionally large spread around these averages, so there is very little correlation between individuals’ vacation gains and their entering reading score ( $r^2 = 0.015$ ).

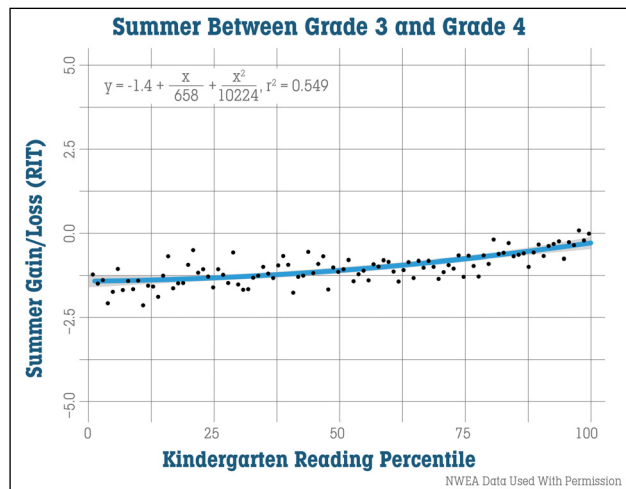
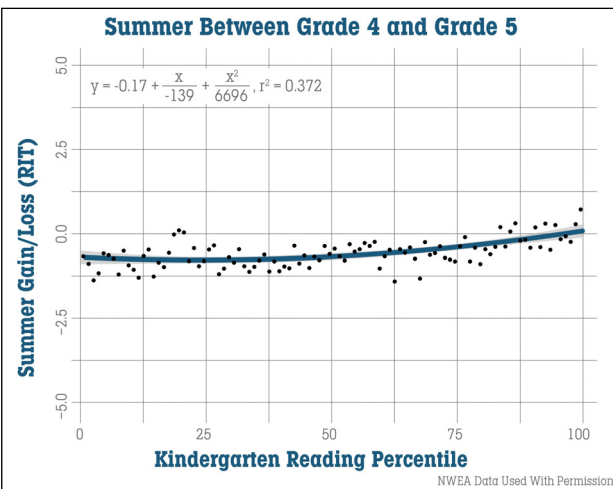
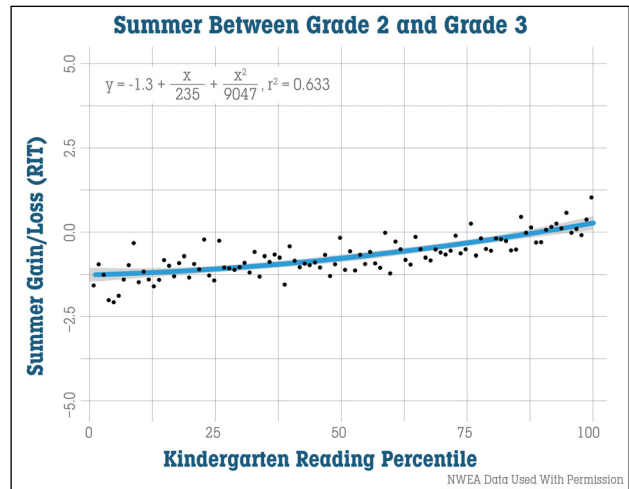
During the vacation between kindergarten and Grade 1, in contrast to all other grade levels, students of all reading abilities show some modest improvement on average. Like other grades, the higher-skill students gain more than the lower-skill group (by about 2.1 RIT points), with over half the differentiation occurring among the top third of students due to a curve in the relationship.



In contrast to the kindergarten summer vacation, the largest average losses occur between Grades 1 and 2. This vacation also sees the most differentiation between the lowest and highest students (3.7 RIT points), and the greatest curvature: the third (quadratic) coefficient is about four times larger in this model than any other. The model predicts that percentile 33 suffers the biggest losses (-4.0 RIT points) with the very lowest percentiles performing marginally better (P1 = -3.1 RIT).

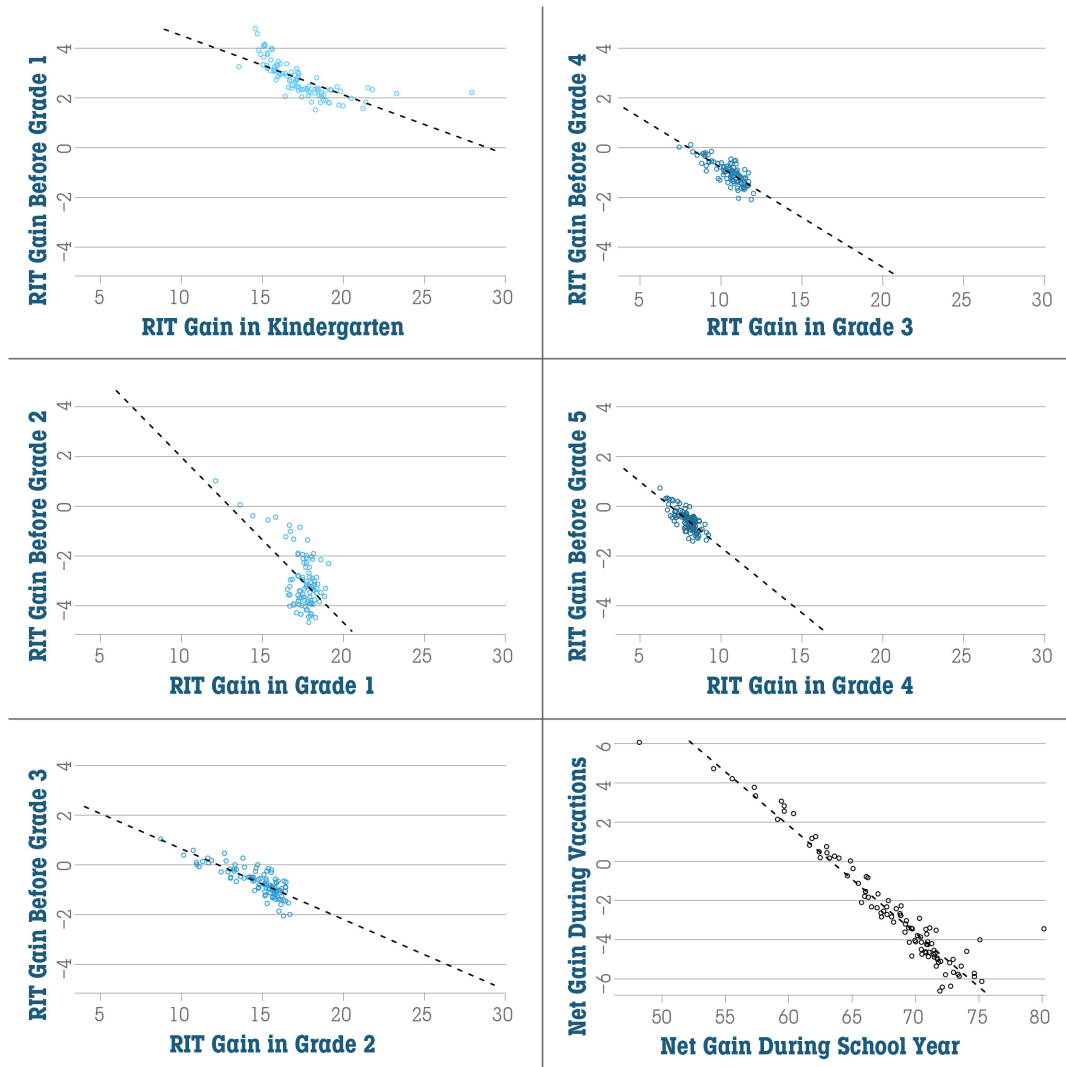


The next three vacations are qualitatively similar: there is a fairly straight-line relationship between kindergarten percentile and summer gain, though the difference between the top and bottom is less than in earlier vacations (1.5, 1.1, and 0.85 RIT points respectively), and consequently the noise in the data begins to overwhelm the trend. Note that only the top few percentiles consistently gain RIT points; all others lose some.



The cumulative effect of these changes is quite striking for the average behavior of percentile groups. However, individual variation in summer gain/loss is much larger than for almost any other metric in the dataset, so this trend--while a powerful indication of group behavior--is a poor predictor of individual behavior. (See figures in Part II.6.)

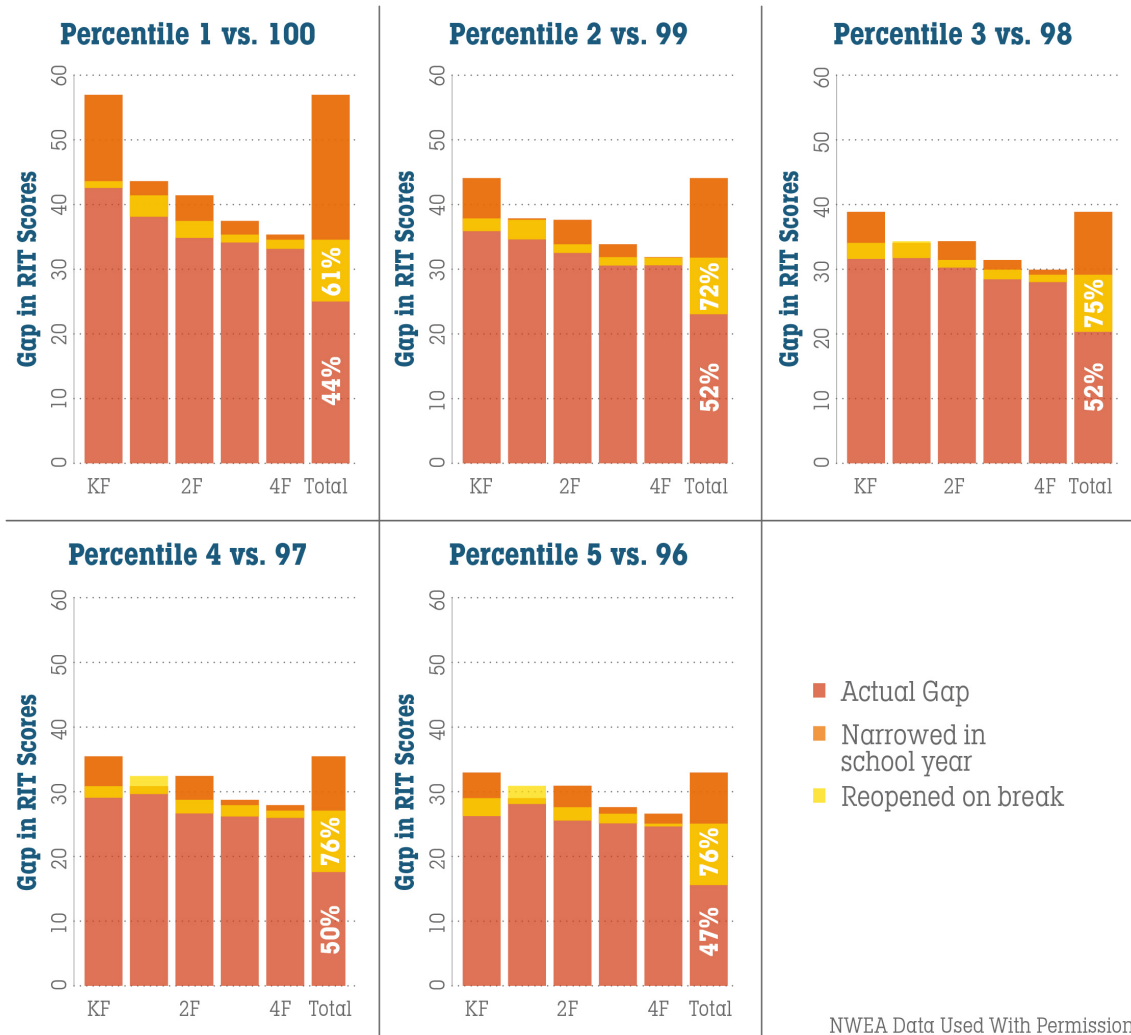
Students who gain many RIT points during the school year also tend to lose more over the summer vacations. Some of this differential loss is due to the tendency of low-scoring students both to gain more points during the year and to lose more over the summer than their high-scoring peers. However, there are other possible explanations. For example, this difference could be demonstrating a “cramming effect,” where students are able to over perform on spring tests after intense study, but they then forget the most and have the biggest relative underperformance the next fall. Students less skilled at cramming will appear to have learned relatively little on the spring test but also will appear to have forgotten relatively little on the next fall’s test.



NWEA Data Used With Permission

Alternatively, students who have an exceptionally bad fall test for whatever reason (inattention, illness, more difficult material, etc.) will appear to have a large vacation loss, followed by a large school-year gain. It is unclear the extent to which the phenomenon graphed below is a result of such auto-correlated time-series.

## Reading Achievement Gap



The gap between the top students and bottom students tends to shrink by about 25% between kindergarten and Grade 5. Some of this is due to regression to the mean: some of the students who scored exceptionally high or low on the first test of kindergarten were over/underperforming and so the group average of the extreme percentiles tends to pull towards the middle. However, the relationship holds up for a trim of 1-10%.

Interestingly, almost all of this narrowing of the achievement gap occurs during the school year; the gap tends to widen again during summer vacations. Without these vacations, the achievement gap would have narrowed by about 50%, not just 25%. In the graph below, the total bar height represents the achievement gap at the start of each school year, the orange-and-yellow bar demonstrates how far the gap decreased during the school year, and the yellow bar shows the amount the gap widened over summer break. The final bar represents the total percentage of the initial achievement gap in kindergarten remaining by Grade 5.

**4. Impact of initial starting points in reading using a simple correlation.**

The effect of entry-level reading test scores (RKF = reading, kindergarten, fall) on subsequent performance was evaluated using univariate least-squares regression. In this and all subsequent equation tables, linear models and correlations were evaluated using Cran-R ver 3.4.3's summary.lm() command. These regressions were performed on both individuals and on percentile groups.

| Formula                             | # of groups | Group Size | R      | R <sup>2</sup> | Error (RIT) | Error (SD) |
|-------------------------------------|-------------|------------|--------|----------------|-------------|------------|
| <b>INDIVIDUALS</b>                  |             |            |        |                |             |            |
| <b>(predicting from K)</b>          |             |            |        |                |             |            |
| <b>R1F = 39.3 + 0.866*RKF</b>       | 61,691      | 1          | 0.6913 | 0.4779         | 6.919       | 0.5615     |
| <b>R2F = 45.7 + 0.923*RKF</b>       | 61,691      | 1          | 0.5807 | 0.3372         | 10.27       | 0.6571     |
| <b>R3F = 74.1 + 0.823*RKF</b>       | 61,691      | 1          | 0.5494 | 0.3019         | 9.595       | 0.6517     |
| <b>R4F = 90.5 + 0.774*RKF</b>       | 61,691      | 1          | 0.5388 | 0.2903         | 9.086       | 0.643      |
| <b>R5F = 101 + 0.75*RKF</b>         | 61,691      | 1          | 0.5307 | 0.2817         | 8.904       | 0.6405     |
| <b>(predicting adjacent grades)</b> |             |            |        |                |             |            |
| <b>R2F = 28.2 + 0.916*R1F</b>       | 61,691      | 1          | 0.7226 | 0.5221         | 8.583       | 0.5493     |
| <b>R3F = 60.8 + 0.737*R2F</b>       | 61,691      | 1          | 0.7826 | 0.6125         | 7.073       | 0.4805     |
| <b>R4F = 51.6 + 0.781*R3F</b>       | 61,691      | 1          | 0.8139 | 0.6624         | 6.299       | 0.4458     |
| <b>R5F = 45.6 + 0.811*R4F</b>       | 61,691      | 1          | 0.824  | 0.6789         | 6.015       | 0.4327     |
| <b>PERCENTILE GROUPS</b>            |             |            |        |                |             |            |
| <b>(predicting from K)</b>          |             |            |        |                |             |            |
| <b>R1F = 39 + 0.868*RKF</b>         | 100         | 617        | 0.9925 | 0.9852         | 0.6634      | 0.0488     |
| <b>R2F = 45.4 + 0.925*RKF</b>       | 100         | 617        | 0.9894 | 0.979          | 0.886       | 0.0572     |
| <b>R3F = 73.8 + 0.824*RKF</b>       | 100         | 617        | 0.993  | 0.986          | 0.6078      | 0.0404     |
| <b>R4F = 90.3 + 0.776*RKF</b>       | 100         | 617        | 0.993  | 0.986          | 0.5945      | 0.0349     |
| <b>R5F = 101.1 + 0.753*RKF</b>      | 100         | 617        | 0.9933 | 0.9867         | 0.5584      | 0.0339     |

5. **Impact of initial starting points in reading using math as an additional variable in a two-variable reading prediction model.** This table is a summary of correlation coefficients to emphasize the benefits of including fall kindergarten math scores.

| Grade Predicted                     | G1     | G2     | G3     | G4     | G5     |
|-------------------------------------|--------|--------|--------|--------|--------|
| 500 Groups of 123<br>K Reading only | 0.9495 | 0.9517 | 0.9416 | 0.9309 | 0.9239 |
| K Reading and Math                  | 0.9884 | 0.9842 | 0.9864 | 0.9862 | 0.9856 |
| Individuals<br>K Reading only       | 0.6913 | 0.5807 | 0.5494 | 0.5388 | 0.5307 |
| K Reading and Math                  | 0.7225 | 0.6025 | 0.5781 | 0.5738 | 0.5694 |

| Formula                                | # of groups | Group Size | R      | R <sup>2</sup> | Error (RIT) | Error (SD) |
|--|-------------|------------|--------|----------------|-------------|------------|
| <b>USING ONLY READING</b>              |             |            |        |                |             |            |
| R1F = 39.1 + 0.867 * RKF               | 500         | 123        | 0.9495 | 0.9015         | 2.8160      | 0.3139     |
| R2F = 45.5 + 0.924 * RKF               | 500         | 123        | 0.9517 | 0.9058         | 2.9276      | 0.3069     |
| R3F = 74 + 0.824 * RKF                 | 500         | 123        | 0.9416 | 0.8866         | 2.8922      | 0.3367     |
| R4F = 90.4 + 0.775 * RKF               | 500         | 123        | 0.9309 | 0.8666         | 2.9868      | 0.3652     |
| R5F = 101 + 0.752 * RKF                | 500         | 123        | 0.9239 | 0.8535         | 3.0574      | 0.3827     |
| <b>USING READING AND MATH</b>          |             |            |        |                |             |            |
| R1F = 49.1 + 0.804*RKF + -0.444*RKF    | 500         | 123        | 0.9884 | 0.9770         | 0.6015      | 0.0488     |
| R2F = 43.6 + 0.936*RKF + 0.079*RKF     | 500         | 123        | 0.9842 | 0.9687         | 0.8931      | 0.0572     |
| R3F = 59.2 + 0.919*RKF + 0.647*RKF     | 500         | 123        | 0.9864 | 0.9730         | 0.5946      | 0.0404     |
| R4F = 70.3 + 0.904*RKF + 0.883*RKF     | 500         | 123        | 0.9862 | 0.9726         | 0.4927      | 0.0349     |
| R5F = 82.5 + 0.871*RKF + 0.817*RKF     | 500         | 123        | 0.9856 | 0.9714         | 0.4714      | 0.0339     |
| <b>PREDICTING INDIVIDUALS</b>          |             |            |        |                |             |            |
| R1F = 31.1 + 0.561 * RKF + 0.561 * MKF | 61,691      | 1          | 0.7225 | 0.5220         | 8.5194      | 0.6914     |
| R2F = 37.8 + 0.626 * RKF + 0.626 * MKF | 61,691      | 1          | 0.6025 | 0.3630         | 12.4718     | 0.7981     |
| R3F = 65.7 + 0.51 * RKF + 0.51 * MKF   | 61,691      | 1          | 0.5781 | 0.3342         | 12.0123     | 0.8159     |
| R4F = 81.8 + 0.445 * RKF + 0.445 * MKF | 61,691      | 1          | 0.5738 | 0.3292         | 11.5721     | 0.8190     |
| R5F = 92.3 + 0.411 * RKF + 0.411 * MKF | 61,691      | 1          | 0.5694 | 0.3243         | 11.4267     | 0.8220     |

6. **Comparison to poverty correlations.**

Note that the correlations listed in the tables above are about twice as large as those commonly reported between poverty and test scores, both on the individual (0.2 to 0.3) and the group (0.68) level. Because the NWEA dataset does not include poverty data, we cannot explicitly demonstrate that poverty impacts later test performance; rather, we hypothesize that its influence modifies entry-level performance.

7. Impact on initial starting points in *math* using a simple correlation.
8. Impact on later math performance of starting points using reading as a second variable.

| Grade predicted          | G1     | G2     | G3     | G4     | G5     |
|--------------------------|--------|--------|--------|--------|--------|
| <b>500 groups of 103</b> |        |        |        |        |        |
| K-Math only              | 0.9674 | 0.9467 | 0.9612 | 0.9656 | 0.9694 |
| K-Math+K-Read            | 0.9937 | 0.9860 | 0.9893 | 0.9889 | 0.9887 |
| <b>Individuals</b>       |        |        |        |        |        |
| K-Math only              | 0.7073 | 0.6122 | 0.6060 | 0.6000 | 0.5942 |
| K-Math + K-Read          | 0.7295 | 0.6406 | 0.6258 | 0.6164 | 0.6076 |

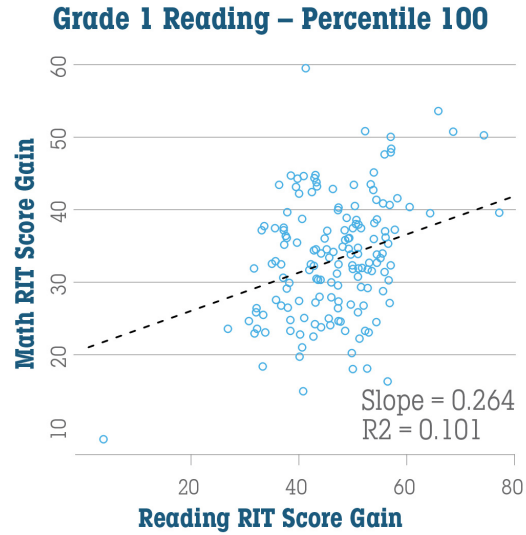
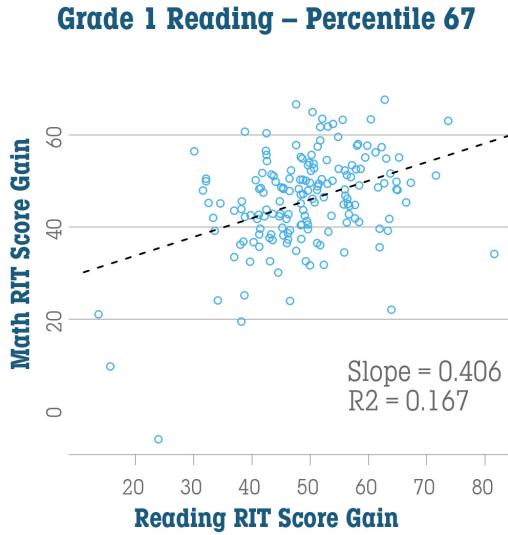
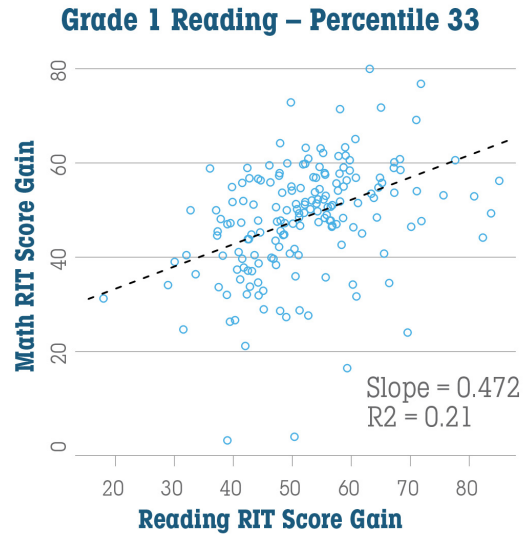
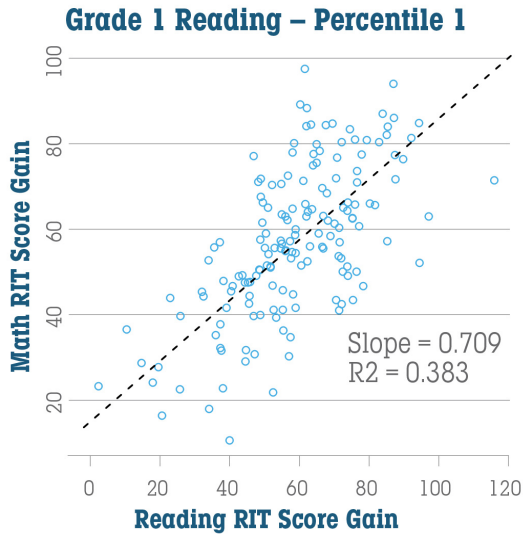
Correlations extracted from the table below emphasize improvement to the model that can be attributed to the addition of scores from kindergarten reading tests.

| Formula  | # of groups | Group Size | R      | R <sup>2</sup> | Error (RIT) | Error (SD) |
|--|-------------|------------|--------|----------------|-------------|------------|
| <b>PREDICTING PERCENTILE GROUPS (using just kindergarten math)</b> |             |            |        |                |             |            |
| M1F = 44.4 + 0.836*MKF   | 100         | 518        | 0.9957 | 0.9915         | 0.7132      | 0.0549     |
| M2F = 82.2 + 0.678*MKF   | 100         | 518        | 0.9886 | 0.9774         | 0.9294      | 0.0764     |
| M3F = 101 + 0.645*MKF  | 100         | 518        | 0.9944 | 0.9889         | 0.6219      | 0.0532     |
| M4F = 109 + 0.667*MKF  | 100         | 518        | 0.9945 | 0.9891         | 0.6330      | 0.0518     |
| M5F = 109 + 0.732*MKF  | 100         | 518        | 0.9952 | 0.9904         | 0.6636      | 0.0490     |
| <b>(predicting adjacent grades)</b>                                |             |            |        |                |             |            |
| M2F = 45.6 + 0.814 * M1F   | 100         | 518        | 0.9968 | 0.9936         | 0.6067      | 0.0802     |
| M3F = 23.8 + 0.943 * M2F   | 100         | 518        | 0.9975 | 0.9950         | 0.5049      | 0.0706     |
| M4F = 4.65 + 1.03 * M3F  | 100         | 518        | 0.9992 | 0.9985         | 0.2896      | 0.0391     |
| M5F = -9.77 + 1.1 * M4F  | 100         | 518        | 0.9990 | 0.9979         | 0.3684      | 0.0454     |
| <b>PREDICTING SMALL GROUPS (using just math)</b>                   |             |            |        |                |             |            |
| M1F = 44.4 + 0.837 * MKF   | 500         | 104        | 0.9674 | 0.9358         | 2.407       | 0.2533     |
| M2F = 82.1 + 0.678 * MKF   | 500         | 104        | 0.9467 | 0.8962         | 2.5359      | 0.3222     |
| M3F = 100 + 0.645 * MKF  | 500         | 104        | 0.9612 | 0.9239         | 2.0337      | 0.2758     |
| M4F = 109 + 0.668 * MKF  | 500         | 104        | 0.9656 | 0.9324         | 1.9765      | 0.2600     |
| M5F = 109 + 0.733 * MKF  | 500         | 104        | 0.9694 | 0.9397         | 2.0392      | 0.2455     |
| <b>(using math and reading)</b>                                    |             |            |        |                |             |            |
| M1F = 28 + 0.594 * MKF + 0.594 * RKF                               | 500         | 104        | 0.9937 | 0.9875         | 1.0611      | 0.1117     |
| M2F = 65.7 + 0.434 * MKF + 0.434 * RKF                             | 500         | 104        | 0.9860 | 0.9722         | 1.3131      | 0.1668     |
| M3F = 87.4 + 0.451 * MKF + 0.451 * RKF                             | 500         | 104        | 0.9893 | 0.9787         | 1.0755      | 0.1458     |

|  |     |     |        |        |        |        |
|--|-----|-----|--------|--------|--------|--------|
| $M4F = 96.2 + 0.485 * MKF + 0.485 * RKF$ | 500 | 104 | 0.9889 | 0.9780 | 1.128  | 0.1484 |
| $M5F = 96.7 + 0.551 * MKF + 0.551 * RKF$ | 500 | 104 | 0.9887 | 0.9775 | 1.2452 | 0.1499 |

|   |       |   |        |        |        |        |
|---|-------|---|--------|--------|--------|--------|
| PREDICTING INDIVIDUALS<br>(using just math) |       |   |        |        |        |        |
| $M1F = 44.6 + 0.835 * MKF$                  | 51781 | 1 | 0.7073 | 0.5002 | 9.1792 | 0.7069 |
| $M2F = 82.3 + 0.677 * MKF$                  | 51781 | 1 | 0.6122 | 0.3748 | 9.6128 | 0.7907 |
| $M3F = 101 + 0.644 * MKF$                   | 51781 | 1 | 0.6060 | 0.3673 | 9.2988 | 0.7955 |
| $M4F = 109 + 0.667 * MKF$                   | 51781 | 1 | 0.6000 | 0.3600 | 9.7754 | 0.8000 |
| $M5F = 109 + 0.732 * MKF$                   | 51781 | 1 | 0.5942 | 0.3530 | 10.889 | 0.8043 |
| <b>(using math and reading)</b>             |       |   |        |        |        |        |
| $M1F = 28.1 + 0.592 * MKF + 0.592 * RKF$    | 51781 | 1 | 0.7295 | 0.5322 | 8.8811 | 0.6840 |
| $M2F = 65.9 + 0.437 * MKF + 0.437 * RKF$    | 51781 | 1 | 0.6406 | 0.4104 | 9.3354 | 0.7679 |
| $M3F = 87.6 + 0.453 * MKF + 0.453 * RKF$    | 51781 | 1 | 0.6258 | 0.3916 | 9.1181 | 0.7800 |
| $M4F = 96.4 + 0.486 * MKF + 0.486 * RKF$    | 51781 | 1 | 0.6164 | 0.3799 | 9.6218 | 0.7874 |
| $M5F = 96.8 + 0.551 * MKF + 0.551 * RKF$    | 51781 | 1 | 0.6076 | 0.3692 | 10.752 | 0.7942 |

9. **Math increases with reading increases.** There is a strong correlation between the reading gain and the math gain between Grades 1 and 5. Because verifying this statement did not require kindergarten scores, we were able to derive a linear model using a much larger dataset ( $n=310,000$ ). The correlation between gains was  $r=0.469$ , with a slope of 0.45, which indicates that each point gained in reading was associated with almost half a point gained in math (and vice versa). The relationship appeared linear, though it was both stronger and steeper at lower percentiles, as shown in the graph below. Data points shown on the graph are 300 students, drawn randomly from the pool of 3,100 students in each percentile, in order to not obscure the trendline.



NWEA Data Used With Permission

10. **Science correlations.** The correlation between science test scores (in the spring of Grades 5 and 8) and various other models is shown below. Note the gradual increase in predictability as the variables approach Grade 5, the general parity between math and reading as predictors, and the failure of gender or ethnicity to improve model fit by more than a few correlation points.

| Equation / Variables                     | n      | r      | r <sup>2</sup> | %1SD  | %2SD  | er(RIT) | err(SD) |
|--|--------|--------|----------------|-------|-------|---------|---------|
| Sci8S = 45.3 + 0.836 * Sci5S             | 6,949  | 0.7492 | 0.5612         | 0.131 | 0.149 | 8.21    | 0.67    |
| Sci5S = 127 + 0.247 * RKF + 0.313 * MKF  | 10,101 | 0.5357 | 0.287          | 0.173 | 0.214 | 8.79    | 0.81    |
| Sci5S = 123 + 0.234 * R1F + 0.286 * M1F  | 40,415 | 0.5915 | 0.3498         | 0.696 | 0.857 | 8.70    | 0.81    |
| Sci5S = 113 + 0.184 * R2F + 0.343 * M2F  | 39,008 | 0.6401 | 0.4097         | 0.69  | 0.83  | 8.25    | 0.76    |
| Sci5S = 96.1 + 0.249 * R3F + 0.331 * M3F | 38,954 | 0.6943 | 0.4821         | 0.711 | 0.832 | 7.73    | 0.72    |
| Sci5S = 86.3 + 0.292 * R4F + 0.307 * M4F | 39,356 | 0.7318 | 0.5355         | 0.737 | 0.843 | 7.32    | 0.68    |

|   |        |        |        |       |       |      |      |
|---|--------|--------|--------|-------|-------|------|------|
| <b>Sci5S = 81.8 + 0.32 * R5F + 0.277 * M5F</b>  | 38,493 | 0.7605 | 0.5784 | 0.733 | 0.826 | 7.03 | 0.65 |
| <b>Sci8S = 128 + 0.304 * RKF + 0.315 * MKF</b>  | 1,335  | 0.5009 | 0.2509 | 0.023 | 0.028 | 9.51 | 0.78 |
| <b>Sci8S = 123 + 0.239 * R1F + 0.344 * M1F</b>  | 10,855 | 0.6036 | 0.3644 | 0.188 | 0.231 | 9.70 | 0.80 |
| <b>Sci8S = 109 + 0.18 * R2F + 0.427 * M2F</b>   | 10,479 | 0.6668 | 0.4447 | 0.189 | 0.223 | 9.05 | 0.74 |
| <b>Sci8S = 88.4 + 0.236 * R3F + 0.438 * M3F</b> | 10,547 | 0.7079 | 0.5011 | 0.195 | 0.226 | 8.58 | 0.70 |
| <b>Sci8S = 78.3 + 0.283 * R4F + 0.404 * M4F</b> | 10,566 | 0.7387 | 0.5456 | 0.199 | 0.226 | 8.22 | 0.68 |
| <b>Sci8S = 73.3 + 0.311 * R5F + 0.374 * M5F</b> | 11,235 | 0.759  | 0.5761 | 0.214 | 0.241 | 7.92 | 0.65 |
| <b>Sci5S = RKF, MKF, Gender, Ethnicity</b>      | 10,093 | 0.5717 | 0.3269 | 0.176 | 0.214 | 8.54 | 0.79 |
| <b>Sci5S = R1F, M1F, Gender, Ethnicity</b>      | 40,407 | 0.6267 | 0.3928 | 0.709 | 0.859 | 8.41 | 0.78 |
| <b>Sci5S = R2F, M2F, Gender, Ethnicity</b>      | 39,000 | 0.665  | 0.4422 | 0.701 | 0.831 | 8.02 | 0.74 |
| <b>Sci5S = R3F, M3F, Gender, Ethnicity</b>      | 38,946 | 0.7117 | 0.5066 | 0.72  | 0.833 | 7.55 | 0.70 |
| <b>Sci5S = R4F, M4F, Gender, Ethnicity</b>      | 39,348 | 0.7456 | 0.556  | 0.744 | 0.843 | 7.15 | 0.66 |
| <b>Sci5S = R5F, M5F, Gender, Ethnicity</b>      | 38,485 | 0.7728 | 0.5973 | 0.74  | 0.826 | 6.87 | 0.64 |

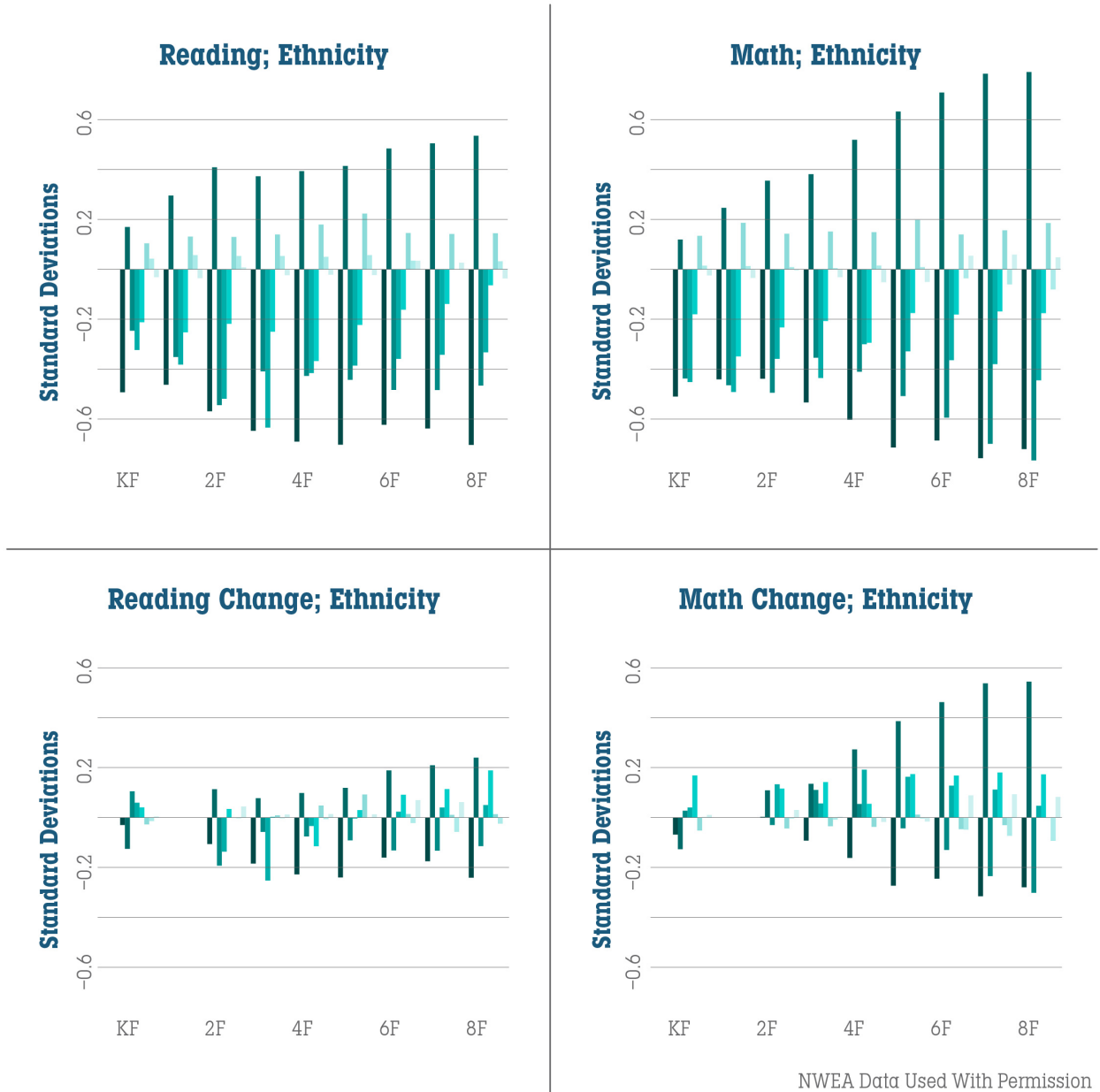
**11. Relative Impact.**

**12. Other factors.**

- a. Gender: The illusory nature of the “gender gap” was evaluated by predicting the average Grade 5 test scores of the 500 kindergarten groups using just the gender composition (PctMale, or the percentage of each of the 500 groups that was male), then using gender and entry-level test scores together. Because of the extremely large sample size, gender remained a “significant” predictor, but the effect size was negligible, approximately 1/20th of a RIT point.

|                                | <b>Effect on G5 for each % Male</b> | <b>% of variation explained by model</b> |
|--------------------------------|-------------------------------------|--|
| <b>R5F ~ %Male</b>             | -0.6099                             | 0.2739                                   |
| <b>R5F ~ %Male + RKF + MKF</b> | -0.0527                             | 0.9729                                   |
| <b>M5F ~ %Male</b>             | -0.3990                             | 0.1348                                   |
| <b>M5F ~ %Male + RKF + MKF</b> | 0.0412                              | 0.9786                                   |

- b. Ethnicity: While many ethnicities at most grade levels showed substantial departures from the mean, these differences were mostly preexisting, and not due to differential performance in elementary school. All achievement gaps were decreased substantially by adjusting achievement by entry-point into Grade 1. With the possible exception of Ethnicities 1 and 2 in math, no trends were apparent, and achievement gaps were smaller than 3.5 RIT points (0.25 standard deviations).



NWEA Data Used With Permission

# APPENDIX A

## The NWEA Testing and Reporting Platforms

**NWEA test information:** The Northwest Evaluation Association, founded in 1977 is a nonprofit testing association based in Portland Oregon. Last year, it tested over ten million students in 9,500 of the nation's 14,500 school districts as well as students in 145 different countries. The assessments utilize a longitudinal RIT scale stretching from kindergarten through Grade 12, so raw scores can be compared within grade levels, across grade levels, across states, and across decades. The RIT scale is in the form of a logarithm where the average RIT points of gain decrease in reading and math each year after kindergarten.

**Adaptive testing.** NWEA pioneered the adaptive testing model, first with a paper-and-pencil model and subsequently with a computer adaptive model. The obvious advantage is, if a given test has 50 questions, and 35 of those questions center at grade level, there are only seven questions to test the actual ability level of those one and two years above grade level and eight questions to test the ability of those one to three years below grade level. This distribution of questions results in the least accurate information being provided for the students who need it the most--those who lag furthest behind. This is one of the reasons that competitors have widely copied the NWEA test design in the last two decades.

**RIT Scale.** The scale is calibrated to provide a 70-point variation in fall reading kindergarten tests scores (110 – 180) for individual children. This test range remains fairly constant as students progress through school.

**Test error.** There is a 70-point variation in kindergarten fall reading test scores (110-180) for individual children as they enter elementary school. This range remains fairly constant as they progress through the grades. On average, most of this difference is real. However, for individual children at any point in time, there can be significant error associated with a specific test result. The tests themselves do not measure performance exactly the same each time the test is taken. The standard deviation (SD) is plus/minus 3 points (i.e., 95% of the time the score will be within plus/minus 6 points of the true score. The overall error, however, reduces with the square of the number of tests taken.

In addition, the tests are given by school districts at different points in time which are generally classified as fall, winter, and spring. We used fall scores to reduce the "error problem". Fall scores have less student fluctuation than spring ones. Individual spring scores show greater variance by the following fall than individual fall-to-fall scores. They are not affected as much by

when the tests are given, which may be as much as six weeks,<sup>23</sup> or by the differential impact of summer loss. These two error possibilities seem to be less influential in fall scores than in spring scores.

**Reading range.** The range in the language and literacy skill levels of incoming kindergarten children remains constant at about five years or normal growth when measured by their fall Northwest Evaluation Association RIT scores between the 10<sup>th</sup> and the 90<sup>th</sup> percentiles each year and between the 50<sup>th</sup> percentiles (or midpoints of quintiles at Grades 1 through 5). This means that at any given year from kindergarten through Grade 8, students in the 10<sup>th</sup> percentile will not score in reading what the students in the 90<sup>th</sup> percentile score until five years later.

**Math range.** The range in math skill levels of incoming kindergarten children appears to start at a little more than three years and achieves about seven years of normal growth by Grade 8 when measured by their Fall NWEA RIT scores between the 10<sup>th</sup> and the 90<sup>th</sup> percentiles within each year and between the 50<sup>th</sup> percentiles (or midpoints of quintiles at Grades 1-5).<sup>24</sup> This means that in kindergarten and Grade 1, students in the 10<sup>th</sup> percentile will not score in math what the students in the 90<sup>th</sup> percentile score until three years later; each year, the gap widens so that by the Grades 6 and 7, the gap is seven years.<sup>25</sup>

The individual based reading prediction model can be further improved by using spring kindergarten reading scores (rather than fall). While this information does not improve the predictability of the 500 groups (correlation remains 0.986), individual scores now have a correlation of 0.633.

|          | No Math | With Math |
|----------|---------|-----------|
| Fall R   | 0.5308  | 0.5706    |
| Spring R | 0.6042  | 0.6329    |

<sup>23</sup> In the Elgin Project in Kentucky, fall tests at individual schools occurred within +/- one week, while the spring tests were +/- 3 weeks). Tests taken later in spring will obviously measure more growth than those taken earlier.

<sup>24</sup> To a greater degree than language, math builds on what comes before. For example, the inability to “carry numbers” in addition and subtraction will likely limit normal advancement into algebra, geometry, or calculus.

<sup>25</sup> The initial range in math at kindergarten seems to be slightly less than four years, growing to five years by Grade 5 and seven years by late middle school. See Fielding, Kerr, and Rosier, *Annual Growth for All Children, Catch-up Growth for Those Who Are Behind* (2007), 178-181.

# APPENDIX B

## The Initial Ask, Metadata, and Numerical Counts

**DATASET INFORMATION.** The initial data come from five cohorts of students who started kindergarten in fall 2006 through fall 2010, provided by the Northwest Evaluation Association (NWEA) located in Portland, Oregon. (See Appendix A.) That dataset consists of 9,368,549 reading and math scores from tests administered in the fall, winter, and spring for 370,308 students in Grades K-8. It includes some science scores at Grade 5.

This paper focuses on two subsets of that dataset. The reading data set consists of 61,691 students with complete reading scores from the fall of their kindergarten year to the spring of their Grade 5. 10.5% of these scores (n=101,580) were interpolations of a missed test, based on the average of the student's previous and subsequent tests.

The equivalent math dataset includes 51,781 students who had test scores for every math test from kindergarten through Grade 5, and at least a kindergarten reading score. There are a total of 795,978 test scores in this dataset, 11.7% (n=92,859) of which were interpolations.

Our initial request to the Northwest Evaluation Association was for about 330,000 students who had reading test scores for both the fall of Grade 1 and the fall of Grade 5, less students for whom NWEA was missing two or more consecutive fall scores.

To get to about 330,000 students, the authors initially requested a pull of the fall and spring reading and math scores for 330,000 students who started Grade 1 in the falls of 2006, 2007, 2008, 2009, and 2010 for Grades 1-5.

For this core student group, the authors then asked for a subsequent pull for

- a. the kindergarten fall and spring reading and math scores that were available
- b. the scores for those students' tests in Grades 6-8, if available, and
- c. available science scores for the springs of Grade 5 and 8.

While in theory, the request should have resulted in an exact matched dataset of Grade 1 and Grade 5 fall reading scores, in fact, it resulted in some 342,000 reading scores in Grade 1 and 328,000 in Grade 5, and some 320,000 math scores in Grade 1 and 301,000 in Grade 5. That number was reduced to 127,000 in Grade 7, and 28,000 in Grade 8. The authors also obtained science scores for both Grade 5 and Grade 8.

See following pages for a count of available scores by cohort by subject by testing period.

**Metadata in data set.**

|                        |   |
|------------------------|---|
| STUDENT_ID             | Student identifier, randomized so students are not identifiable   |
| ASSESS_STUDENT_GRADE   | The grade of student for the score                                |
| STUDENT_GENDER         | male/female   |
| STUDENT_ETHNIC_GRD_KEY | Ethnicity of student  |
| SCHOOL_ID              | School building (one of 50,000 in US) - also randomized           |
| DISTRICT_ID            | School district (one of 15,000 in US) - also randomized           |
| STATE_NAME             | not randomized  |
| TERM_NAME              | will either be fall test and year or spring test and year         |
| ASSESS_SUBJECT         | 1 = math, 2= reading, 4= science                                  |
| TEST_TYPE_NAME         | NWEA specific test type, generally not relevant                   |
| TEST_NAME              | Name of test -  |
| ASSESS_TEST_START_DT   | date of test and time it started                                  |
| ASSESS_TEST_DURATION   | length of test in seconds (tests under 15 minutes are less valid) |
| ASSESS_RIT             | NWEA uses a longitudinal scale from K to grade 12                 |
| ASSESS_STD_ERR         | standard error  |
| ASSESS_TEST_%ILE       | student's national percentile rank                                |
| PCT_CORRECT            | % correct on the test   |
| PROJECTED_PROFICIENCY  | 1="yes", 0="no"   |

**See the table below for the actual test score counts by year by subject by season.**

| READING |        |        |        |        |         |         |
|---------|--------|--------|--------|--------|---------|---------|
| Grade   | Term   | 2006   | 2007   | 2008   | 2009    | 2010    |
| K       | Fall   | 0      | 5,493  | 11,803 | 25,326  | 36,250  |
| K       | Spring | 0      | 12,935 | 27,196 | 46,447  | 61,950  |
| 1       | Fall   | 19,018 | 46,794 | 73,587 | 103,255 | 127,637 |
| 1       | Spring | 18,701 | 45,990 | 72,514 | 101,971 | 125,731 |
| 2       | Fall   | 18,307 | 44,983 | 70,608 | 99,480  | 122,701 |
| 2       | Spring | 18,367 | 44,971 | 70,417 | 98,772  | 122,176 |
| 3       | Fall   | 18,832 | 45,250 | 70,737 | 98,618  | 121,238 |
| 3       | Spring | 18,770 | 45,261 | 70,487 | 98,685  | 121,186 |
| 4       | Fall   | 18,865 | 45,592 | 70,933 | 99,386  | 121,973 |
| 4       | Spring | 18,903 | 45,851 | 71,378 | 98,308  | 122,729 |
| 5       | Fall   | 19,018 | 46,788 | 73,566 | 68,148  | 127,576 |
| 5       | Spring | 17,778 | 45,486 | 70,836 | 64,454  | 122,351 |
| 6       | Fall   | 0      | 41,568 | 64,559 | 60,252  | 104,489 |
| 6       | Spring | 0      | 40,124 | 62,542 | 60,640  | 93,552  |
| 7       | Fall   | 0      | 36,932 | 54,394 | 71,032  | 0       |
| 7       | Spring | 0      | 35,523 | 51,214 | 64,449  | 0       |
| 8       | Fall   | 0      | 29,899 | 46,603 | 0       | 0       |
| 8       | Spring | 0      | 26,139 | 40,644 | 0       | 0       |

| MATH  |        |        |        |        |        |         |
|-------|--------|--------|--------|--------|--------|---------|
| Grade | Term   | 2006   | 2007   | 2008   | 2009   | 2010    |
| K     | Fall   | 0      | 5,189  | 12,332 | 25,356 | 36,071  |
| K     | Spring | 0      | 12,290 | 25,600 | 43,613 | 57,912  |
| 1     | Fall   | 17,053 | 43,660 | 70,767 | 99,639 | 122,282 |
| 1     | Spring | 16,812 | 42,068 | 68,682 | 96,667 | 118,919 |
| 2     | Fall   | 17,721 | 43,849 | 69,330 | 96,809 | 119,537 |
| 2     | Spring | 17,864 | 43,773 | 68,115 | 96,597 | 119,564 |
| 3     | Fall   | 18,634 | 44,389 | 69,656 | 93,663 | 117,690 |
| 3     | Spring | 18,047 | 43,061 | 67,315 | 94,321 | 117,099 |
| 4     | Fall   | 18,604 | 44,914 | 68,736 | 95,852 | 118,545 |
| 4     | Spring | 17,840 | 43,770 | 67,984 | 94,886 | 118,665 |
| 5     | Fall   | 18,873 | 46,582 | 72,349 | 66,924 | 126,086 |
| 5     | Spring | 17,718 | 42,919 | 68,763 | 60,442 | 116,227 |
| 6     | Fall   | 0      | 37,223 | 61,531 | 46,976 | 100,072 |
| 6     | Spring | 0      | 38,184 | 60,152 | 57,573 | 94,300  |
| 7     | Fall   | 0      | 32,305 | 51,228 | 66,961 | 0       |
| 7     | Spring | 0      | 33,979 | 47,677 | 64,526 | 0       |
| 8     | Fall   | 0      | 27,149 | 43,450 | 0      | 0       |
| 8     | Spring | 0      | 26,167 | 41,061 | 0      | 0       |

| SCIENCE |        |       |       |       |        |        |
|---------|--------|-------|-------|-------|--------|--------|
| Grade   | Term   | 2006  | 2007  | 2008  | 2009   | 2010   |
| 5       | Spring | 1,596 | 7,507 | 8,285 | 11,785 | 12,851 |
| 8       | Spring | 1,057 | 4,474 | 5,764 | 0      | 0      |

# APPENDIX C

## Summary of Linear Models Used in this Paper

| Formula                                | # of groups | Size of groups | r      | r <sup>2</sup> | Mean Avg Err (RIT) | Mean Avg Err (SD) | Preds w/in 1 SD | Preds w/in 2 SD |
|--|-------------|----------------|--------|----------------|--------------------|-------------------|-----------------|-----------------|
| R1F = 39 + 0.868 * RKF                 | 100         | 617            | 0.9926 | 0.9852         | 1.050              | 0.122             | 1               | 1               |
| R2F = 45.4 + 0.925 * RKF               | 100         | 617            | 0.9894 | 0.9790         | 1.337              | 0.145             | 1               | 1               |
| R3F = 73.8 + 0.824 * RKF               | 100         | 617            | 0.9930 | 0.9860         | 0.970              | 0.118             | 1               | 1               |
| R4F = 90.3 + 0.776 * RKF               | 100         | 617            | 0.9930 | 0.9860         | 0.911              | 0.118             | 1               | 1               |
| R5F = 101 + 0.752 * RKF                | 100         | 617            | 0.9934 | 0.9868         | 0.859              | 0.115             | 1               | 1               |
| R2F = 3.86 + 1.07 * R1F                | 100         | 617            | 0.9964 | 0.9928         | 0.781              | 0.085             | 1               | 1               |
| R3F = 34.4 + 0.886 * R2F               | 100         | 617            | 0.9972 | 0.9943         | 0.617              | 0.075             | 1               | 1               |
| R4F = 21 + 0.94 * R3F                  | 100         | 617            | 0.9989 | 0.9977         | 0.367              | 0.048             | 1               | 1               |
| R5F = 13.8 + 0.968 * R4F               | 100         | 617            | 0.9991 | 0.9983         | 0.309              | 0.041             | 1               | 1               |
| R1F = 39.1 + 0.867 * RKF               | 500         | 123            | 0.9495 | 0.9015         | 2.816              | 0.314             | 0.996           | 1               |
| R2F = 45.5 + 0.924 * RKF               | 500         | 123            | 0.9517 | 0.9058         | 2.928              | 0.307             | 0.994           | 1               |
| R3F = 74 + 0.824 * RKF                 | 500         | 123            | 0.9416 | 0.8866         | 2.892              | 0.337             | 0.994           | 1               |
| R4F = 90.4 + 0.775 * RKF               | 500         | 123            | 0.9309 | 0.8666         | 2.987              | 0.365             | 0.994           | 1               |
| R5F = 101 + 0.752 * RKF                | 500         | 123            | 0.9239 | 0.8535         | 3.057              | 0.383             | 0.992           | 1               |
| R1F = 30.8 + 0.557 * RKF + 0.557 * MKF | 500         | 123            | 0.9884 | 0.9770         | 1.362              | 0.152             | 0.998           | 1               |
| R2F = 37.4 + 0.623 * RKF + 0.623 * MKF | 500         | 123            | 0.9842 | 0.9687         | 1.688              | 0.177             | 0.998           | 1               |
| R3F = 65.4 + 0.506 * RKF + 0.506 * MKF | 500         | 123            | 0.9864 | 0.9730         | 1.412              | 0.164             | 0.998           | 1               |
| R4F = 81.4 + 0.44 * RKF + 0.44 * MKF   | 500         | 123            | 0.9862 | 0.9726         | 1.354              | 0.166             | 1               | 1               |
| R5F = 91.9 + 0.407 * RKF + 0.407 * MKF | 500         | 123            | 0.9856 | 0.9714         | 1.351              | 0.169             | 1               | 1               |
| R1F = 31.1 + 0.561 * RKF + 0.561 * MKF | 61,691      | 1              | 0.7225 | 0.5220         | 8.519              | 0.691             | 0.862           | 0.991           |
| R2F = 37.8 + 0.626 * RKF + 0.626 * MKF | 61,691      | 1              | 0.6025 | 0.3630         | 12.472             | 0.798             | 0.784           | 0.99            |
| R3F = 65.7 + 0.51 * RKF + 0.51 * MKF   | 61,691      | 1              | 0.5781 | 0.3342         | 12.012             | 0.816             | 0.796           | 0.979           |
| R4F = 81.8 + 0.445 * RKF + 0.445 * MKF | 61,691      | 1              | 0.5738 | 0.3292         | 11.572             | 0.819             | 0.809           | 0.974           |
| R5F = 92.3 + 0.411 * RKF + 0.411 * MKF | 61,691      | 1              | 0.5694 | 0.3243         | 11.427             | 0.822             | 0.813           | 0.974           |
| R1F = 39.3 + 0.866 * RKF               | 61,691      | 1              | 0.6913 | 0.4779         | 8.903              | 0.723             | 0.845           | 0.989           |
| R2F = 45.7 + 0.923 * RKF               | 61,691      | 1              | 0.5807 | 0.3372         | 12.722             | 0.814             | 0.775           | 0.989           |
| R3F = 74.1 + 0.823 * RKF               | 61,691      | 1              | 0.5494 | 0.3019         | 12.301             | 0.836             | 0.786           | 0.975           |
| R4F = 90.5 + 0.774 * RKF               | 61,691      | 1              | 0.5388 | 0.2903         | 11.903             | 0.842             | 0.799           | 0.971           |
| R5F = 101 + 0.75 * RKF                 | 61,691      | 1              | 0.5307 | 0.2817         | 11.781             | 0.848             | 0.802           | 0.972           |
| M1F = 44.4 + 0.836 * MKF               | 100         | 518            | 0.9957 | 0.9915         | 0.854              | 0.092             | 1               | 1               |
| M2F = 82.2 + 0.678 * MKF               | 100         | 518            | 0.9886 | 0.9774         | 1.136              | 0.150             | 1               | 1               |
| M3F = 101 + 0.645 * MKF                | 100         | 518            | 0.9944 | 0.9889         | 0.753              | 0.105             | 1               | 1               |
| M4F = 109 + 0.667 * MKF                | 100         | 518            | 0.9945 | 0.9891         | 0.773              | 0.104             | 1               | 1               |
| M5F = 109 + 0.732 * MKF                | 100         | 518            | 0.9952 | 0.9904         | 0.795              | 0.098             | 1               | 1               |
| M2F = 45.6 + 0.814 * M1F               | 100         | 518            | 0.9968 | 0.9936         | 0.607              | 0.080             | 1               | 1               |
| M3F = 23.8 + 0.943 * M2F               | 100         | 518            | 0.9975 | 0.9950         | 0.505              | 0.071             | 1               | 1               |
| M4F = 4.65 + 1.03 * M3F                | 100         | 518            | 0.9992 | 0.9985         | 0.290              | 0.039             | 1               | 1               |
| M5F = -9.77 + 1.1 * M4F                | 100         | 518            | 0.9990 | 0.9979         | 0.368              | 0.045             | 1               | 1               |
| M1F = 44.4 + 0.837 * MKF               | 500         | 104            | 0.9674 | 0.9358         | 2.407              | 0.253             | 1               | 1               |

|   |        |     |        |        |        |       |       |       |
|---|--------|-----|--------|--------|--------|-------|-------|-------|
| $M2F = 82.1 + 0.678 * MKF$  | 500    | 104 | 0.9467 | 0.8962 | 2.536  | 0.322 | 1     | 1     |
| $M3F = 100 + 0.645 * MKF$   | 500    | 104 | 0.9612 | 0.9239 | 2.034  | 0.276 | 1     | 1     |
| $M4F = 109 + 0.668 * MKF$   | 500    | 104 | 0.9656 | 0.9324 | 1.977  | 0.260 | 1     | 1     |
| $M5F = 109 + 0.733 * MKF$   | 500    | 104 | 0.9694 | 0.9397 | 2.039  | 0.246 | 1     | 1     |
| $M1F = 28 + 0.594 * MKF + 0.594 * RKF$  | 500    | 104 | 0.9937 | 0.9875 | 1.061  | 0.112 | 1     | 1     |
| $M2F = 65.7 + 0.434 * MKF + 0.434 * RKF$  | 500    | 104 | 0.9860 | 0.9722 | 1.313  | 0.167 | 0.998 | 1     |
| $M3F = 87.4 + 0.451 * MKF + 0.451 * RKF$  | 500    | 104 | 0.9893 | 0.9787 | 1.075  | 0.146 | 1     | 1     |
| $M4F = 96.2 + 0.485 * MKF + 0.485 * RKF$  | 500    | 104 | 0.9889 | 0.9780 | 1.128  | 0.148 | 1     | 1     |
| $M5F = 96.7 + 0.551 * MKF + 0.551 * RKF$  | 500    | 104 | 0.9887 | 0.9775 | 1.245  | 0.150 | 1     | 1     |
| $M1F = 44.6 + 0.835 * MKF$  | 51,781 | 1   | 0.7073 | 0.5002 | 9.179  | 0.707 | 0.855 | 0.99  |
| $M2F = 82.3 + 0.677 * MKF$  | 51,781 | 1   | 0.6122 | 0.3748 | 9.613  | 0.791 | 0.806 | 0.983 |
| $M3F = 101 + 0.644 * MKF$   | 51,781 | 1   | 0.6060 | 0.3673 | 9.299  | 0.795 | 0.814 | 0.979 |
| $M4F = 109 + 0.667 * MKF$   | 51,781 | 1   | 0.6000 | 0.3600 | 9.775  | 0.800 | 0.814 | 0.98  |
| $M5F = 109 + 0.732 * MKF$   | 51,781 | 1   | 0.5942 | 0.3530 | 10.889 | 0.804 | 0.807 | 0.98  |
| $M1F = 28.1 + 0.592 * MKF + 0.592 * RKF$  | 51,781 | 1   | 0.7295 | 0.5322 | 8.881  | 0.684 | 0.868 | 0.992 |
| $M2F = 65.9 + 0.437 * MKF + 0.437 * RKF$  | 51,781 | 1   | 0.6406 | 0.4104 | 9.335  | 0.768 | 0.819 | 0.986 |
| $M3F = 87.6 + 0.453 * MKF + 0.453 * RKF$  | 51,781 | 1   | 0.6258 | 0.3916 | 9.118  | 0.780 | 0.822 | 0.982 |
| $M4F = 96.4 + 0.486 * MKF + 0.486 * RKF$  | 51,781 | 1   | 0.6164 | 0.3799 | 9.622  | 0.787 | 0.821 | 0.981 |
| $M5F = 96.8 + 0.551 * MKF + 0.551 * RKF$  | 51,781 | 1   | 0.6076 | 0.3692 | 10.752 | 0.794 | 0.812 | 0.982 |
| $Sci8S = 45.3 + 0.836 * Sci5S$  | 6,949  | 1   | 0.7492 | 0.5612 | 8.214  | 0.675 | 0.131 | 0.149 |
| $Sci5S = 127 + 0.247 * RKF + 0.313 * MKF$   | 10,101 | 1   | 0.5357 | 0.287  | 8.788  | 0.814 | 0.173 | 0.214 |
| $Sci5S = 123 + 0.234 * R1F + 0.286 * M1F$   | 40,415 | 1   | 0.5915 | 0.3498 | 8.702  | 0.806 | 0.696 | 0.857 |
| $Sci5S = 113 + 0.184 * R2F + 0.343 * M2F$   | 39,008 | 1   | 0.6401 | 0.4097 | 8.254  | 0.764 | 0.69  | 0.83  |
| $Sci5S = 96.1 + 0.249 * R3F + 0.331 * M3F$  | 38,954 | 1   | 0.6943 | 0.4821 | 7.734  | 0.716 | 0.711 | 0.832 |
| $Sci5S = 86.3 + 0.292 * R4F + 0.307 * M4F$  | 39,356 | 1   | 0.7318 | 0.5355 | 7.317  | 0.677 | 0.737 | 0.843 |
| $Sci5S = 81.8 + 0.32 * R5F + 0.277 * M5F$   | 38,493 | 1   | 0.7605 | 0.5784 | 7.031  | 0.651 | 0.733 | 0.826 |
| $Sci8S = 128 + 0.304 * RKF + 0.315 * MKF$   | 1,335  | 1   | 0.5009 | 0.2509 | 9.508  | 0.781 | 0.023 | 0.028 |
| $Sci8S = 123 + 0.239 * R1F + 0.344 * M1F$   | 10,855 | 1   | 0.6036 | 0.3644 | 9.697  | 0.797 | 0.188 | 0.231 |
| $Sci8S = 109 + 0.18 * R2F + 0.427 * M2F$  | 10,479 | 1   | 0.6668 | 0.4447 | 9.047  | 0.743 | 0.189 | 0.223 |
| $Sci8S = 88.4 + 0.236 * R3F + 0.438 * M3F$  | 10,547 | 1   | 0.7079 | 0.5011 | 8.575  | 0.704 | 0.195 | 0.226 |
| $Sci8S = 78.3 + 0.283 * R4F + 0.404 * M4F$  | 10,566 | 1   | 0.7387 | 0.5456 | 8.216  | 0.675 | 0.199 | 0.226 |
| $Sci8S = 73.3 + 0.311 * R5F + 0.374 * M5F$  | 11,235 | 1   | 0.759  | 0.5761 | 7.917  | 0.650 | 0.214 | 0.241 |
| $Sci5S = 126 + 0.243 * RKF + 0.279 * MKF + 1.8 * Gender + Eth2 \{10.7, 2.12, 5.43, 8.07, 5.9, 5.44, 5.59\}$     | 10,093 | 1   | 0.5717 | 0.3269 | 8.539  | 0.790 | 0.176 | 0.214 |
| $Sci5S = 121 + 0.24 * R1F + 0.248 * M1F + 1.94 * Gender + Eth2 \{10.5, 2.21, 5.46, 5.78, 6.65, 6.58, 5.67\}$    | 40,407 | 1   | 0.6267 | 0.3928 | 8.410  | 0.779 | 0.709 | 0.859 |
| $Sci5S = 113 + 0.194 * R2F + 0.298 * M2F + 1.76 * Gender + Eth2 \{8.28, 2.56, 4.97, 6.32, 6.44, 6.16, 5.16\}$   | 39,000 | 1   | 0.665  | 0.4422 | 8.024  | 0.743 | 0.701 | 0.831 |
| $Sci5S = 97.4 + 0.255 * R3F + 0.293 * M3F + 1.65 * Gender + Eth2 \{6.62, 1.48, 3.76, 3.89, 5.03, 5.01, 3.75\}$  | 38,946 | 1   | 0.7117 | 0.5066 | 7.549  | 0.699 | 0.72  | 0.833 |
| $Sci5S = 87.8 + 0.295 * R4F + 0.276 * M4F + 1.71 * Gender + Eth2 \{4.81, 0.685, 2.63, 3.35, 4.05, 3.78, 3.17\}$ | 39,348 | 1   | 0.7456 | 0.556  | 7.154  | 0.662 | 0.744 | 0.843 |
| $Sci5S = 82.7 + 0.324 * R5F + 0.252 * M5F + 1.81 * Gender + Eth2 \{3.9, 0.77, 2.68, 3.28, 3.83, 3.5, 2.69\}$    | 38,485 | 1   | 0.7728 | 0.5973 | 6.872  | 0.636 | 0.74  | 0.826 |

# APPENDIX D

## Data for Figures in Part II

**Data for Figure 3**

| Formula                     | # of groups | Group Size | R      | R <sup>2</sup> | Error (RIT) | Error (SD) |
|-----------------------------|-------------|------------|--------|----------------|-------------|------------|
| $R5F = 101.1 + 0.753 * RKF$ | 100         | 617        | 0.9933 | 0.9867         | 0.5584      | 0.0339     |

**Data for Figure 4**

| Formula                    | # of groups | Group Size | R      | R <sup>2</sup> | Error (RIT) | Error (SD) |
|----------------------------|-------------|------------|--------|----------------|-------------|------------|
| $R1F = 39 + 0.868 * RKF$   | 100         | 617        | 0.9926 | 0.9852         | 1.0504      | 0.1218     |
| $R2F = 3.86 + 1.07 * R1F$  | 100         | 617        | 0.9964 | 0.9928         | 0.7807      | 0.0847     |
| $R3F = 34.4 + 0.886 * R2F$ | 100         | 617        | 0.9972 | 0.9943         | 0.6169      | 0.0754     |
| $R4F = 21 + 0.94 * R3F$    | 100         | 617        | 0.9989 | 0.9977         | 0.3673      | 0.0477     |
| $R5F = 13.8 + 0.968 * R4F$ | 100         | 617        | 0.9991 | 0.9983         | 0.3087      | 0.0413     |

**Data for Figure 5**

| Formula                                  | # of groups | Group Size | R      | R <sup>2</sup> | Error (RIT) | Error (SD) |
|--|-------------|------------|--------|----------------|-------------|------------|
| $R5F = 101 + 0.752 * RKF$                | 500         | 123        | 0.9239 | 0.8535         | 3.0574      | 0.3827     |
| $R5F = 82.5 + 0.871 * RKF + 0.817 * MKF$ | 500         | 123        | 0.9856 | 0.9714         | 0.4714      | 0.0339     |
| $R5F = 101 + 0.75 * RKF$                 | 61,691      | 1          | 0.5307 | 0.2817         | 8.9040      | 0.6405     |
| $R5F = 92.3 + 0.411 * RKF + 0.411 * MKF$ | 61,691      | 1          | 0.5694 | 0.3243         | 11.4267     | 0.8220     |

The table below compares the accuracy of reading predictions with math predictions. Even though math groups are smaller (which should make average group behavior more difficult to predict), the correlation between kindergarten and Grade 5 tests is even higher (0.995 vs. 0.993).

| Formula                     | # of groups | Group Size | R      | R <sup>2</sup> | Error (RIT) | Error (SD) |
|-----------------------------|-------------|------------|--------|----------------|-------------|------------|
| $R5F = 101.1 + 0.753 * RKF$ | 100         | 617        | 0.9933 | 0.9867         | 0.5584      | 0.0339     |
| $M5F = 109 + 0.732 * MKF$   | 100         | 518        | 0.9952 | 0.9904         | 0.7949      | 0.0979     |

#### Data for Section 5

| Formula                                  | # of groups | Group Size | R      | R <sup>2</sup> | Error (RIT) | Error (SD) |
|--|-------------|------------|--------|----------------|-------------|------------|
| $R5F = 101 + 0.752 * RKF$                | 500         | 123        | 0.9239 | 0.8535         | 3.0574      | 0.3827     |
| $M5F = 109 + 0.733 * MKF$                | 500         | 104        | 0.9694 | 0.9397         | 2.0392      | 0.2455     |
| $R5F = 82.5 + 0.871 * RKF + 0.817 * MKF$ | 500         | 123        | 0.9856 | 0.9714         | 0.4714      | 0.0339     |
| $M5F = 96.7 + 0.551 * MKF + 0.551 * RKF$ | 500         | 104        | 0.9887 | 0.9775         | 1.2452      | 0.1499     |
| $R5F = 101 + 0.75 * RKF$                 | 61,691      | 1          | 0.5307 | 0.2817         | 8.9040      | 0.6405     |
| $M5F = 109 + 0.732 * MKF$                | 51,781      | 1          | 0.5942 | 0.3530         | 10.8891     | 0.8043     |
| $R5F = 92.3 + 0.411 * RKF + 0.411 * MKF$ | 61,691      | 1          | 0.5694 | 0.3243         | 11.4267     | 0.8220     |
| $M5F = 96.8 + 0.551 * MKF + 0.551 * RKF$ | 51,781      | 1          | 0.6076 | 0.3692         | 10.7524     | 0.7942     |