APPENDIX I METHODOLOGY

Two cohorts were created for the main study. (The sample for the secondary analyses, highlighted in the sidebar titled *A Closer Look at High Flyers in High-Poverty Schools,* is explained in Appendix II.) They consisted of public-school students in grades three through eight (Cohort 1) and in grades six through ten (Cohort 2). Among these cohorts, high achievers were those students who performed at or above the 90th percentile, based on NWEA's 2008 norms, on their third-grade (Cohort 1) or sixth-grade (Cohort 2) Measures of Academic Progress (MAP) tests.

Students in **Cohort 1** were followed from third grade in the 2004-05 school year until the end of eighth grade in the 2009-10 school year (Table A-1). The elementary/middle school cohort consisted of 81,767 students in math and 93,182 students in reading, and was drawn from more than 1,500 schools in thirty states.

Students in **Cohort 2** were followed from sixth grade in the 2005-06 school year through tenth grade in the 2009-10 school year.[12] This cohort consisted of 43,423 students in math and 48,220 students in reading, and was drawn from more than 800 schools in twenty-eight states.

Each cohort included only students who had data in both the initial and final grades studied. Students were not required to test in each intervening grade as a condition for inclusion in the cohort, because this requirement would have decreased the size of the sample dramatically.

TABLE A-1

Cohorts 1 and 2 by Grade and Year

Cohort	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10
Cohort 1 (Elementary/Middle School)	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Cohort 2 (Middle/High School)	N/A	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10

The proportions of high achievers in the study varied by cohort and by subject, though every group saw its proportion of high achievers grow over the course of the study (Table A-2). The percentage of high flyers in math at the elementary/middle level, for instance, grew from 12.4 percent of all students in third grade to 14.1 percent in eighth grade.

Demographic characteristics of the full sample, by cohort and subject, are shown in Table A-3.[13] Parallel demographic breakdowns of the high-achieving groups within each cohort and subject are shown in Table A-4 (initial year of study) and Table A-5 (final year of study). A quick comparison of the two tables reveals that females were underrepresented among high achievers in math, and both minority students and students in high-poverty schools were underrepresented in all categories of high achievers.

12 We did not track students through higher grades due to the fact that smaller numbers of students participate in testing at those grades.

13 Minority students were defined as children from traditionally disadvantaged ethnic groups and included African American, Hispanic, and Native American students. Non-minority students included Anglo and Asian students. Low poverty was defined as schools in which less than 50 percent of students received free or reduced-price lunch, while high poverty refers to a school in which more than 50 percent did so.

TABLE A-2

High Achievers in Initial and Final Years

	Total Number of Students in Cohort	Number of High Flyers in Initial Year	Percentage of High Flyers in Initial Year	Number of High Flyers in Final Year	Percentage of High Flyers in Final Year	Change in High-Flyer Percentage
COHORT 1 (ELEMENTARY/MIDDLE SCHOOL)						
Math	81,767	10,116	12.4%	11,544	14.1%	+1.7%
Reading	93,182	10,925	11.7%	12,429	13.3%	+1.6%
COHORT 2 (MIDDLE/HIGH SCHOOL)						
Math	43,423	2,912	6.7%	4,779	11.0%	+4.3%
Reading	48,220	4,394	9.1%	4,677	9.7%	+0.6%

TABLE A-3

Demographic Characteristics of Total Sample of Students

	Gender		Ethnicity		School Poverty Status	
	Female	Male	Minority	Non-minority	High Poverty	Low Poverty
Cohort 1 Math	49.7%	50.3%	23.3%	76.7%	30.6%	69.4%
Cohort 1 Reading	49.3%	50.7%	24.4%	75.6%	29.4%	70.6%
Cohort 2 Math	49.2%	50.8%	25.1%	74.9%	31.0%	69.0%
Cohort 2 Reading	48.9%	51.1%	23.5%	76.5%	30.1%	69.9%

TABLE A-4

Demographic Characteristics of High Achievers (Initial Year of Study)

	Gender		Ethnicity		School Poverty Status	
	Female	Male	Minority	Non-minority	High Poverty	Low Poverty
Cohort 1 Math	41.9%	58.1%	8.2%	91.8%	19.4%	80.6%
Cohort 1 Reading	51.7%	48.3%	9.0%	91.0%	13.5%	86.5%
Cohort 2 Math	39.0%	61.0%	7.3%	92.7%	18.1%	81.9%
Cohort 2 Reading	49.8%	50.2%	6.7%	93.3%	16.6%	83.4%

TABLE A-5

Demographic Characteristics of High Achievers (Final Year of Study)

	Gender		Ethnicity		School Poverty Status	
	Female	Male	Minority	Non-minority	High Poverty	Low Poverty
Cohort 1 Math	44.0%	56.0%	8.2%	91.8%	16.1%	83.9%
Cohort 1 Reading	53.0%	47.0%	9.4%	90.6%	13.4%	86.6%
Cohort 2 Math	41.7%	58.3%	7.8%	92.2%	15.3%	84.7%
Cohort 2 Reading	52.6%	47.4%	7.3%	92.7%	14.7%	85.3%

Beyond the composition of the two cohorts, there are five key technical issues and/or limitations that require clarification:

1. **Possible misclassification attributable to error on the assessment.** The issue of measurement error is relevant to this study because it is possible that students whose measured MAP scores placed them just under or over the 90th percentile threshold for high achievement might "truly" belong on the other side. In other words, it is possible that the true scores of some high achievers may fall below the 90th percentile and the true scores of some non-high achievers may fall above the cut line. (For individual students, the misclassification probability is 50 percent for students whose scores are exactly at the threshold, but that probability decreases quickly as students' measured scores rise higher above or fall further below the high-achievement cut off).

   To account for measurement error, tests often provide a standard error of measurement (SEM) along with an observed score. Standard errors indicate how much potential measurement error may exist. For example, if a student's measured score was 110 and the standard error of measurement was two, then that student's "true" score most likely falls within the range of 108 to 112, and almost certainly within the range of 106 to 114. With computerized adaptive tests such as MAP, standard errors of measurement are roughly constant for all observable MAP scores; typical SEMs on the test are about three points.

   Because this study examines the attrition rate among students who perform at or above the 90th percentile on NWEA's norms, it is important to address questions regarding the likely proportion of students scoring at that level who may have been misidentified due to SEM. Hence, we calculated the "maintenance rate" that would be expected for the high-achieving group if measurement error were taken into account. The calculation depends on the score distribution among the sample of high achievers (how close initial high achievers scored to the cut off) and the measurement error associated with each score. Based on this information, we estimated the expected maintenance rates were 83.4 percent for math and 83.3 percent for reading for the elementary/middle school cohort. Put another way, we would expect about 83 percent of the high-achieving math group to remain intact if the entire group of high achievers were immediately retested, and the attrition in the group would be explained by the measurement error associated with the test. The report indicates that the actual maintenance rate for this group was considerably lower; only 57 percent of the elementary/middle math cohort remained high-achieving. That means that the expected attrition, at about 17 percent, was much higher in reality, at 43 percent. The difference between the anticipated attrition and the observed attrition is perhaps the best representation of the actual attrition within the high-achieving group.[14] In other words, a total of 26 percent of students became Descenders for reasons other than measurement error.

2. **Risk of regression toward the mean.** Fixed-form tests have a relatively high risk of score regression toward the mean, partly due to ceiling effects and partly because they exact a relatively high penalty on inadvertent errors. Because fixed-form tests have to assess performance across the entire spectrum of achievement, they provide a limited number of items that can be targeted to any one group. Thus, the number of items used to measure the performance of high achievers tends to be small, generally only five to ten in a fifty item test. This contributes to a ceiling effect. In addition, because so few items discriminate among high achievers, a high achiever who inadvertently misses an item (forgets to "carry a one" on an addition problem, for example) often finds his score takes a large (and unrecoverable) penalty. Adaptive tests such as MAP, however, have lower risk of regression to the mean because they offer more appropriately targeted items to high-performing students and exact relatively small penalties for inadvertent errors. Further, in an adaptive test, the primary "penalty" for missing an item is that the student receives an easier item; in fact, students are expected to miss approximately 50 percent of the items on the test. Thus, they have many opportunities to "recover" from an inadvertently missed item.

14 While this addresses the likelihood of overclassification—i.e., classifying students as high-performing who are not—there are also students not included in the initial high-achieving cohort because they were underclassified due to measurement error. In a normal distribution, the number of underclassified students is likely to be slightly largely than the number overclassified because of their position in the distribution (there are fewer students just above the 90th percentile than there are just below the 90th percentile). Were measurement error to be completely resolved—an impossibility, unfortunately—the actual estimate of high achievers might be slightly higher than what we report.

3. **Use of a normative cut score to define eligibility for the sample.** This study used a normative standard, the 90th percentile on the 2008 NWEA norms, to define the cut point for high achievement. While the cut point was normative, the use of that standard did not limit the number of students in the study sample who could be identified as high-achieving, nor did it limit the number of students who could enter or leave high-achieving status. In other words, we did not design the study so that any student moving above the 90th percentile norm had to replace another student. In fact, the overall count of high-achieving students increased over time within the study group. Thus while a normative cut score was applied, its application did not create a zero-sum result for the study sample.

4. **Limiting the sample to students with scores at the beginning and ending grades.** We limited the sample to students who had test records at both the beginning and ending grades. We considered restricting the sample further to require students to have scores at every grade, but rejected that option because it would have excluded the vast majority of students from the sample. Limiting the sample in this way could impact measurement error as discussed above. However, it should be noted that in the secondary analyses (Appendix II) we used data points in the intervening years to measure growth in the hierarchical linear model (HLM).

5. **Measurement of student improvement.** To estimate improvement rates, we compared high achievers (students performing at or above the 90th percentile on NWEA norms) with low achievers (students performing below the 10th percentile) and middle achievers (students performing between the 45th and 54th percentiles, inclusive). We used all available data points to depict their trend lines, but the calculation was not based on individual growth trajectories. For the measure of growth (see *A Closer Look at High Flyers in High-Poverty Schools* on page 15), we estimated growth trajectories for the school-defined groups via HLM. For more information on this analysis, see Appendix II.