Appendix D: Methodology

I. Short Cycle Assessment data

For all analyses, we rely on 2007-08 and 2008-09 district-wide short cycle assessments (SCA). In both math and reading, these are KDPR (spring 2007-08) and A2L (spring 2007-08 through spring 2008-09). In reading, we also use DRA2/EDL2 (spring 2007-08 through spring 2008-09). Note the necessary lag; the guide will be used to compare 2009-10 student results to those from the 2008-09 cohort. This guide will be updated yearly in order to compare student results with the previous year's cohort.

Kindergarten Developmental Progress Report (KDPR). This district-created kindergarten assessment determines students' proficiency in mathematics and language arts through a combination of teacher observations, group exercises, and individual exercises. Each item is given a rating of 0 - 3, where 0 = "area of need," 1 = "emergent," 2 = "nearing proficiency," and 3 = "proficient." For the purposes of this process, we summed results from questions 17 to 30 to obtain math scores and summed results from questions 33 to 44 to obtain reading scores (for spring 2007-08).

Developmental Reading Assessment/Evaluación del Desarrollo de la Lectura (DRA2/EDL2). This SCA, published by Pearson, determines students' independent reading level and is used district-wide for grades K, 1, and 2. Only kindergarteners who are developmentally ready take this assessment. In the spring of 2008, approximately 91% of kindergarteners were assessed. For third grade, schools have a choice of using either DRA2/EDL2 or A2L/DBA as the reading SCA. (Some schools use both.) In 2008-09, approximately 29% of 3rd graders were assessed with the DRA2/EDL2. In this assessment, teachers determine a student's independent reading level based on oral reading fluency, comprehension, and reading strategies. For the purposes of these guidelines, we have recoded and treated independent reading levels as an interval-level variable ranging from 1 to 21. The English and Spanish versions of the assessment are assumed to be equivalent; and students who took the assessment in both languages are assigned the higher earned level. Levels of DRA2/EDL2 are listed in Appendix B.

Assess To Learn (A2L)/ now District Benchmark Assessment (DBA). This assessment, published by Riverside and designed to reflect New Mexico standards, is the district's SCA in reading and math for grades 3 through 8. A2L/DBA scores are the percentage of multiple choice items answered correctly; therefore scores range from 0 to 100. Some elementary schools use DRA2/EDL2 for reading for third grade instead. In 2008-09, approximately 66% to 76% of 3rd grade students were assessed with A2L for reading (depending on the testing window). Some elementary schools use this SCA for 1st and 2nd grade math assessment as well. In 2008-09, approximately 34% of 1st graders and 43% of 2nd graders were assessed in math with A2L.

In 2009-10, elementary-level DBA math changed from being a survey test, which assesses the entire year's material at each window, to a benchmark test, which assesses only information that a student should have been taught and should know up to that point in time. Additionally, district staff made significant changes in items in order to align the assessment to standards to a greater extent. The likely consequence of these changes in terms of this guide is the over-identification of weaknesses. (There were few changes to DBA in reading, which was essentially already a benchmark assessment.)

Appendix D: Methodology

II. Identifying weaknesses in achievement

The definition of "weakness in achievement" is based on NM Public Education Department's current eligibility guidelines for a specific learning disability – an SCA score at least 1.5 standard deviations below the district average. All students in a grade level at the time who were administered an assessment (including students identified as having a specific learning disability) are included in the computation of each SCA's average and standard deviation. Assuming scores are normally distributed, approximately 7% of students will be identified as having a weakness in achievement.

It is worth noting that standard deviations tend to be high, perhaps somewhat more in reading than in math. For instance, the standard deviation for the DRA2/EDL2 is usually three or four levels, suggesting that 'normal' scores cover a wide range of seven to nine levels. For A2L, standard deviations tend to be around 21 points in reading and 17 points in math. The consequence is that students must earn a score or level quite far below average to be considered to have a weakness in achievement.

III. Identifying weaknesses in growth

Again, the definition of "weakness in growth" is based on NM Public Education Department's current eligibility guidelines for a specific learning disability – an SCA score at least 1.5 standard deviations below the district average growth rate. We have interpreted standard deviation to mean standard error of slope (change in achievement over time). Assuming growth is normally distributed, approximately 7% of students will be identified as having a weakness in growth.

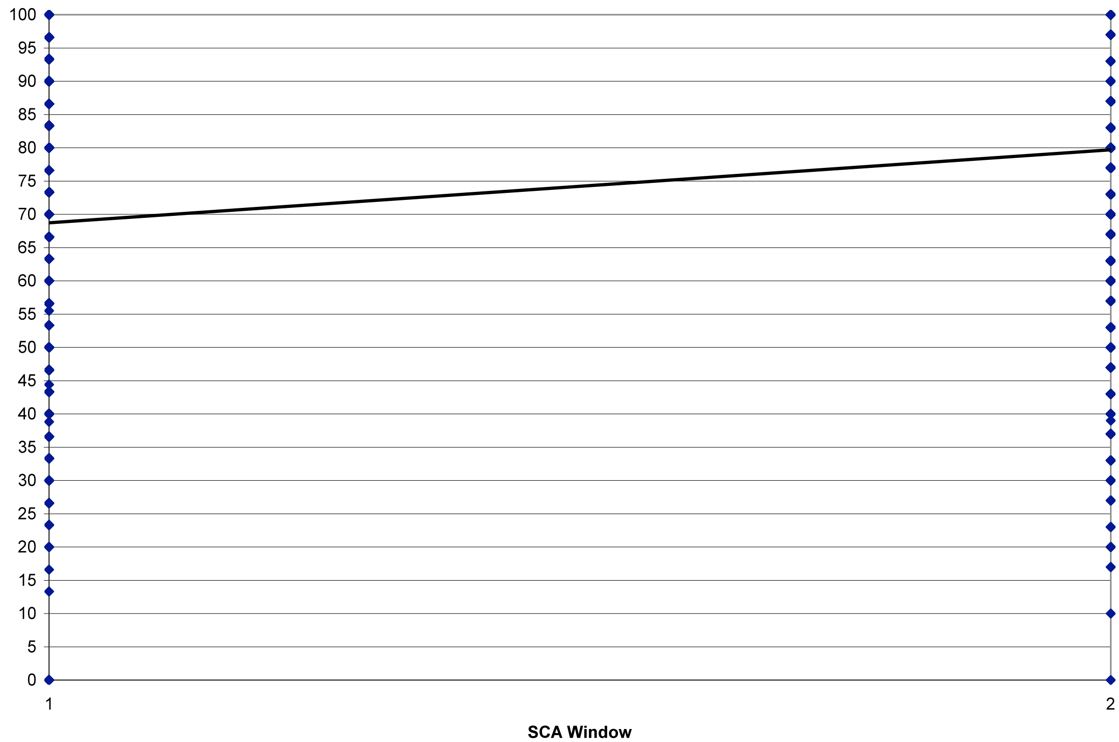
Interestingly, standard errors tend to be low (particularly when Method B, described below, is used to determine growth). For instance, on average 2nd grade students increase their reading abilities by 2.1 DRA2/EDL2 levels between winter and spring; and the standard error is only .1 levels. The consequence is difficulty in distinguishing normal and weak growth.

Method A. Two methods are used to determine growth (slope), depending on available data. The first ("Method A") is used when the same assessment is used for both time periods. This is the more common method. It has the advantage of including all available assessment information, whether or not a student took the assessment for both time periods. Additionally, this method results in a single value (slope or change in achievement) to identify a weakness that applies to every student in the grade level. In other words, the maximum change in achievement does not depend on how a student scored during the first time period.

We used OLS regression to predict all scores with the single independent (explanatory) variable, time (i.e. SCA window), with the first time period coded "1" and the second coded "2". The standard error of the coefficient for the independent variable, time, was used to determine 1.5 standard errors below average.

Figure 1 is an example of Method A. It shows 1st graders' math scores in the y-axis, and benchmark window in the x-axis. The first time period is winter, when the average math A2L score for 1st graders is 70.8. The second time period is spring, when the average math A2L score is 79.7. The slope of the line between the two is 8.9, meaning that on average students' scores increase by about 8.9 points. With a standard error of .4, 1.5 standard errors below the average slope is 8.2. A weakness in growth is based on an individual student's change in scores from winter to spring of 8.2 points or less, regardless of their winter score. Every student's growth could be plotted on this graph. Any student whose growth rate is less than 8.2 (either flatter or negative) would be considered weak in growth.

Figure 1: 1st Grade Math A2L Scores in Winter and Spring, 2008-09



Method B. In some cases, it is not possible to look at change over time directly because assessments with different scales are used, making it impossible to graph scores on the y-axis and time on the x-axis as in Method A. This commonly happens when schools use different assessments in different grades. In these cases, we have elected to determine average growth by predicting scores on the second assessment using scores from the first assessment. This means we must omit information from students who were not assessed in both time periods. The likely consequence of this action is to calculate average growth that is artificially high (though probably not by much).

We used OLS regression to determine the relationship between current and previous SCA scores, which results in an equation of a line, commonly expressed as $y = mx + b$, that defines average growth. In this case, y = the score on the current assessment, x = the score on the previous assessment, and b is the score on the current assessment when the score on the previous assessment is 0 (the intercept). We altered this equation to define weak growth by lowering the coefficient for x (i.e. m) by 1.5 standard errors. So, all points falling on or below this new line represent weak growth.

Appendix D: Methodology

Figure 2 is an example of Method B. It shows 1st graders' fall A2L math scores on the y-axis; the average score is 56.2. The x-axis is comprised of the same cohort's KDPR math scores from the previous spring; the average score is 38.6. The equation of the linear trend line is: y = 18.75 +.96x, suggesting that A2L scores increases by .96 for every 1 point increase in KDPR. With a standard error of .04, the equation for the line defining a weakness is growth becomes y<=18.75 +.90x, or, solving for x, x>=(y-18.75)/.90. All points falling on or below this line represent weak growth. For instance, a student earning a 35 on the fall A2L would be considered to have a weakness in growth if he or she scored an 18 or *higher* on the spring KDPR. If the student had scored below an 18, than his or her "point" would land above the line, implying normal or high growth.

Figure 2: 1st Grade Fall 2008 Math A2L by Kindergarten Spring 2008 Math KDPR