Children Left Behind in AYP and Non-AYP Schools: Using Student Progress and the Distribution of Student Gains to Validate AYP

CRESST Report 637

Kilchan Choi, Michael Seltzer, Joan Herman,
and Kyo Yamashiro
CRESST / UCLA

September, 2004

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
GSE& IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Educational Sciences (IES), U.S. Department of Education.

Erratum: Minor changes were made to this report in July 2007 per the request of the first author.

CHILDREN LEFT BEHIND IN AYP AND NON-AYP SCHOOLS: USING STUDENT PROGRESS AND THE DISTRIBUTION OF STUDENT GAINS TO VALIDATE AYP

Kilchan Choi, Michael Seltzer, Joan Herman, and Kyo Yamashiro
CRESST / UCLA

Abstract

The No Child Left Behind Act (NCLB, 2002) establishes ambitious goals for increasing student learning and attaining equity in the distribution of student performance. Schools must assure that all students, including all significant subgroups, show adequate yearly progress toward the goal of 100% proficiency by the year 2014. In this paper, we wish to illustrate an alternative way of measuring AYP that both emphasizes individual student growth over time, and focuses on the distribution of student growth between performance subgroups. We do so through analyses of a longitudinal dataset from an urban school district in Washington. We also examine what these patterns tell us about schools that meet their AYP targets and those that do not. This alternative way of measuring AYP helps bring to light potentially important aspects of school performance that might be masked if we limit our focus to classifying schools based only on current AYP criteria. In particular, we are able to identify some schools meeting the Washington state criteria for AYP, for example, that have above average students making substantial progress but below average students making little to no progress. In contrast, other schools making AYP have below average students making adequate progress but above average students making little gains. These contrasts raise questions about the meaning of "adequate" progress—and to whom the notion of progress is referring. We believe that closely examining the distribution of student progress may provide an important supplementary or alternative measure of AYP.

Introduction

The No Child Left Behind Act (NCLB, 2002) establishes ambitious goals for increasing student learning and attaining equity in the distribution of student performance. Schools must show they are making "adequate yearly progress," or AYP, toward the goal of all students achieving proficiency in mathematics and reading/language arts by the year 2014. In addition, schools must assure that all subgroups of students make such progress, including students who are economically disadvantaged, students with disabilities, and students with limited English proficiency.

With its attention to AYP for traditionally low performing subgroups, NCLB is clearly aimed at closing the gap and bringing those who have been traditionally been left behind up to proficiency. But as currently defined, the AYP metric may not be fully consonant with this mission. Rather than looking at the progress of students from year to year to directly examine whether previously "left behind" children are making progress toward or attaining proficiency, the AYP metric examines the performance of cohorts from year to year in terms of the proportion attaining proficiency on state standards-based tests. If the proportion for the school as a whole and for each numerically significant subgroup is at or greater than that specified in state annual measurement objectives (AMO), then the school is designated a meeting AYP and presumably credited with making progress in closing the achievement gap. But is this actually the case? There is some evidence of unintended consequences that may work against the students in most educational need. That is, by concentrating on the proportion of students achieving a particular cut point, the score needed to be considered proficient, current AYP requirements unintentionally encourage some schools under great pressure to meet AMO goals to triage their efforts to focus on moving students who are closest to the threshold over it, virtually ignoring their lowest performing students. Moreover, schools whose low performing subgroups fail to meet the minimum group size may meet AYP without addressing the needs of these students. The question gets at the heart of the validity of the current AYP indicator.

In contrast, attending to the relationship between where students start (i.e., initial status) and how rapidly they progress (i.e., their rates of change) can provide us with indicators of whether and which children are being left behind and whether the achievement gap within a school is being closed. In other words, these indicators enable us to have direct insight into how equitably student achievement is distributed within a school (Seltzer, Choi, & Thum, 2003). If a school's average growth rate is high, and the relationship between initial status and growth over time is strongly positive, we know that children are being left behind. In this case, lower achieving students are making less progress than their higher achieving peers and as a result, initial gaps in the achievement among students become magnified over time. The high overall growth rate for the school masks important inequity in growth patterns. On the other hand, if a school's average growth rate is only modestly positive, and the relationship between initial status and growth over time is negative or even weakly positive, this would be an indication that the school is doing a reasonable job of leaving no child behind, despite the lower overall gains, though some individual students may still be making poor progress. In this case, progress is being underestimated by the low overall growth rate in the school. A very desirable scenario would be one in which virtually all students in a school are growing at rapid and fairly similar rates, which would be reflected in a high overall rate of growth and an extremely weak relationship between initial status and growth. Another would be one in which virtually all students in a school are making strong progress, but those with low initial status are growing more rapidly, which would be reflected in a high overall rate and a negative relationship between initial status and growth.

The methodology proposed in this paper for measuring growth and how equitably it is distributed provides an indicator for looking at the validity of current AYP designations. The methodology estimates not only school mean growth, but also the expected growth or progress of specific subgroups within a school defined by students' initial performance status—low, medium or high relative to school average performance. While these subgroups are different from those defined by current policy (ethnicity, language ability, disability classification, or income level), it is worth noting that their definition gets to the heart of the goal of closing the achievement gap by providing a direct measure of whether schools are helping their lowest performing students to make substantial progress and/or catch up. The definition also is one that is applicable across all schools, holding them responsible assuring the progress of all their students, and not just current AYP subgroups that meet minimum group size. In the results reported below, we examine expected growth patterns for students who perform at a school's mean initial starting value, as well as for students performing roughly one standard deviation above and below the school's average starting point. By doing this, we can look across schools and see whether the same amount of initial gap within each school widens, narrows or remains the same. Our study draws on longitudinal data from an urban district in the state of Washington. Our criteria for adequate progress are defined by, and driven by, individual students' gains over a 2-year period. This notion of adequate progress, though different from AYP in NCLB, still addresses many of the underlying goals and tenets of NCLB. At the time of the study, the AYP decision for an individual school was based on the percentage of fourth graders scoring proficient on the state's standards-based assessment, in other words a multiple cohort design drawing on different cohorts of students as they passed through the same grade level (fourth grade). Our notion of adequate growth, by contrast, emphasizes monitoring each individual student's progress longitudinally. For example, we measure growth over two subsequent years, using students' test scores at Grade 3 in 2001 and at Grade 5 in 2003. This is an example of a longitudinal panel data design. In this type of design, it is possible to examine how much an individual student progresses each year, or over a span of 2 or 3 years, as well as whether there are differences in progress across different subgroups.

We first classified schools based on the AYP criteria used by the state department of education website in the 2001-2002 school year. "AYP Schools" are those that met the state's AYP targets; "Non-AYP Schools" are those that did not. We re-evaluated each of the AYP and Non-AYP schools based on our alternative framework for measuring progress. As noted above, our analysis leads to an evaluation of the validity of the current AYP measure as well as a possible alternative (or supplement) to it.

Research Questions

We address the following questions in this study:

1. Are there schools that meet AYP yet still have children who are not making substantial progress? That is, are some schools classified as AYP leaving some children behind?
2. Are there schools that do not meet AYP yet still enable students to make substantial progress?
3. Do AYP schools achieve a more equitable distribution of student growth in achievement than Non-AYP schools? In other words, are schools that meet AYP enabling students at all ability levels to make substantial progress?
4. Are there Non-AYP Schools that are reducing the achievement gap and essentially acting to leave no child behind?

Data

Our study draws on elementary school data from a large, racially diverse, urban school district in a state in northwest, where required testing includes the Iowa Test of Basic Skills (ITBS) in Grades 3 and 5 and state's standard-based assessment in Grade 4. While determinations of AYP are based on the state test, the ITBS results provided us with the data needed for analyzing student growth. The difference in the tests, between the one used to measure progress in our analysis and that used by AYP criteria, is admittedly a limitation in our comparisons, particularly given differences in the tests' respective alignment with state standards. However, in order to show growth over time for individuals in elementary schools, we relied on the measure that was used in more than one elementary grade, the ITBS. In addition, equating the state test and the ITBS in order to use adjacent grade levels was not an option with the data provided.

We used the ITBS reading scale scores for the longitudinal analysis. These scale scores are vertically equated developmental scores and thus are appropriate for examining student change over time. The test scores for third graders in 2001, and the test scores for these students when they were in the fifth grade, provided the basis of estimating gains in achievement between Grades 3 and 5 for each student in our sample. The total number of students in our sample was 2,995. Among those students, however, approximately 14% of students have either transit to different schools in the district between Grades 3 and 5, and approximately additional 2% of students do not have test score either at Grades 3 or 5. The analysis includes only students who had test scores in both Grades 3 and 5 and remained in the same school. In all, there are 2,524 students in this sample nested within 72 schools. The number of students in each school ranges from 13 to 87. The average number of students in this sample per school is approximately 35.

Among the sample of 72 schools, 51 are AYP schools (or met the AYP criteria) and the remaining 21 are Non-AYP schools (or did not meet the AYP criteria). As mentioned earlier, decisions regarding AYP in this particular state are based on the state's standards-based test. At the elementary school level, the state test is administered only at Grade 4. Decisions whether each elementary school met AYP in the baseline year (2001-02) and during the second year (2002-03) were made based on fourth grade state test scores. In this study, the AYP and Non-AYP grouping is based on the baseline year (2001-02), when our sampled students were in the fourth grade.

Note that we used standard errors of measurement (SEs) connected with students' ITBS test scores in our analysis. The SE for various ranges of test scores was obtained from the technical manual of the ITBS test booklet. As will be seen in detail in the following section, the SE information is incorporated into our analyses in estimating the true initial status and the true gain score for each student in each school.

Estimating True Gain Based on Two-Time Points with the Standard Errors of Measurement: Latent Variable Regression Hierarchical Model (LVR-HM_Gain)

We use an advanced two-level hierarchical modeling technique in a fully Bayesian framework (Choi & Seltzer, 2003; Seltzer, Choi & Thum, 2003; see also Raudenbush & Bryk, Chpt. 11, 2002 for a maximum likelihood based approach). This modeling methodology incorporates latent variable regression techniques into two-level hierarchical models. In a Level-1 (within-individual) model, time-series observations are modeled as a function of a time metric, and we obtain estimates of initial status and rates of change for individual students. In a Level-2 (between-individual) model, rates of change for each student are modeled as a function of their initial status. The coefficient capturing the relationship between initial status and rates of change represents an expected increase (in the case of a positive coefficient) or decrease (in the case of a negative coefficient) in rates of change when initial status increases one unit.

We apply this modeling technique to settings where we have observations for individuals at two points in time, and the corresponding standard errors of measurement for both observations. In the following Level-1 model, ITBS test scores at Grades 3 and 5 for each of N students (i = 1, …, N) (Y_{ti}) are modeled as a linear function of time ($Time_{ti}$). $Time_{ti}$ takes a value of 0 for Grade 3 and a value of 1 for Grade 5.

$$Y_{ti} = \pi_{0i} + \pi_{1i}\, Time_{ti} + \varepsilon_{ti} \qquad \varepsilon_{ti} \sim N(0, [SE(Y_{ti})]^2) \tag{1a}$$

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \pi_{0i} \\ \pi_{1i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \qquad \varepsilon_{ti} \sim N(0, [SE(Y_{ti})]^2) \tag{1b}$$

By virtue of our time metric, π_{0i} represents initial status for student i, and π_{1i} captures gain or change during the span of time. Note that ε_{ti} are residuals assumed to be normally distributed with a mean of zero and variance represented by $[SE(Y_{ti})]^2$.

To incorporate the standard errors of measurement (SEs) into our model, we re-pose a Level-1 model of the following form:

$$Y^*_{ti} = \pi_{0i} + \pi_{1i}\, Time^*_{ti} + \varepsilon^*_{ti} \qquad \varepsilon^*_{ti} \sim N(0, 1) \tag{2a}$$

$$\begin{pmatrix} Y_{1i} / SE(Y_{1i}) \\ Y_{2i} / SE(Y_{2i}) \end{pmatrix} = \begin{pmatrix} 1/ SE(Y_{1i}) & 0/ SE(Y_{1i}) \\ 1/ SE(Y_{2i}) & 1/ SE(Y_{2i}) \end{pmatrix} \begin{pmatrix} \pi_{0i} \\ \pi_{1i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} / SE(Y_{1i}) \\ \varepsilon_{2i} / SE(Y_{2i}) \end{pmatrix}, \tag{2b}$$

where we scale the left and right hand sides of Equation 1 by the inverse of the SEs for student i such that ε^*_{1i} and ε^*_{2i} are both assumed to be normally distributed and have a mean of zero, but their variance is now 1 ($\varepsilon^*_{ti} \sim N$ (0, 1). By re-scaling outcomes and the time metric based on an estimate of the precision associated with each test score, (1/ SE(Y_{ti})) (see Bryk, Thum, Easton & Luppescu, p. 135, 1998), we are able to obtain estimates of true variance in gain and initial status, and estimate the relationship between true gain and initial status in each school. Note here that both π_{0i} and π_{1i} are latent variables, not observed variables.

We now pose the following Level-2 (between student) model:

$$\pi_{0i} = \beta_{00} + r_{0i} \qquad r_{0i} \sim N(0, \tau_{00}) \qquad (3a)$$

$$\pi_{1i} = \beta_{10} + b(\pi_{0i} - \beta_{00}) + r_{1i} \qquad r_{1i} \sim N(0, \tau_{11}) \quad Cov(r_{0i}, r_{1i}) = 0 \qquad (3b)$$

In Equation 3a, β_{00} represents the population mean initial status and r_{0i} a random effect capturing deviations in the initial status of students from β_{00}. In Equation 3b, true gain for a student is modeled as a function of true initial status. Thus, the latent variable regression coefficient b captures the relationship between initial status and amount of gain. In other words, b captures the expected increase or decrease in the amount of gain (π_{1i}) when initial status (π_{0i}) increases one unit (see for example, Raudenbush & Bryk, 2002; Seltzer et al., 2003). β_{10} now represents the expected gain for student i whose initial status value (π_{0i}) is equal to the mean initial status (β_{00}) for that student's school. As for variance components, τ_{00} captures the extent to which students vary in initial status, and τ_{11} represents a residual variance in gain after taking into account differences in initial status.

Including initial status as a predictor for the gain helps address an important question on how student growth is distributed within a school (Seltzer et al., 2003). The questions are as follows: Does a student who starts out with higher scores tend to gain more than a student who starts out with lower scores? Is the gap among students diminishing or increasing over time? As mentioned above, the latent variable regression coefficient, b, gives us a sense of the patterns underlying the relationship between expected gain and initial status. By attending to b coefficient, we can ask a question on what the expected gain would be for students at different levels of initial status. More specifically, we examine the expected gain for: a) a student whose initial status value is equal to the school mean initial status, b) a student who is initially starting 15 points below the school mean initial status, c) a student who is initially starting 15 points above the school mean initial status. The latter two categories represent students who started approximately at 1 SD below and above each school's mean value. Note that 15 points is the magnitude of the within-school SD in grade 3 ITBS scores one tends to see for the schools in the sample.

Estimating the expected values at different levels of initial status based on the LVR-HM_Gain

We fit each school's data to the model described in Equations 2a, 3a, and 3b. Note that all analyses of the LVR-HM_Gains presented in this paper are conducted using the software program WinBUGS1.4 (i.e., the Windows version of Bayesian Analysis Using the Gibbs Sampler developed by Spiegelhalter et al, 2002). WinBUGS provides a fairly easy means of implementing the Gibbs sampler in a wide range of modeling settings. Upon convergence, the Gibbs sampler essentially provides us with an accurate approximation of the marginal posterior distribution for all of the parameters in our model. Using the resulting marginal posterior distributions, we can obtain the posterior mean, median, standard deviation, and 95% interval based on the .025 and .975 quantiles of the marginal posterior distribution.

We can estimate the expected gains for various levels of initial status using Equation (3b) which defines the relationship between gain and initial status. For example, we are interested in estimating the expected gains for three sets of student initial status values: values equal to the school mean, 15 points below the school mean, and 15 points above the school mean.

First, drawing on Equation 3b, the following equation specifies the expected gain for a student whose initial status value (π_{0i}) is equal to his or her school's mean initial status value (β_{00}):

$$E\ (\pi_{1i} \mid \pi_{0i} = \beta_{00}) = \beta_{10} + b(\pi_{0i} - \beta_{00}) = \beta_{10} + b \times (0) = \beta_{10} \quad (4a)$$

Note that β_{10} can be viewed as the mean gain for a given school.

Second, the expected gain for students starting out 15 points below the school mean initial status value is as follows:

$$E\ (\pi_{1i} \mid \pi_{0i} = \beta_{00} - 15) = \beta_{10} + b(\pi_{0i} - \beta_{00}) = \beta_{10} + b \times (-15) \quad (4b)$$

Similarly, the expected gain for students whose initial status value is 15 points above the school mean is as follows:

$$E\,(\pi_{1i} \mid \pi_{0i} = \beta_{00} + 15) \;=\; \beta_{10} + b(\pi_{0i} - \beta_{00}) = \beta_{10} + b \times (15) \qquad (4c)$$

Note that if the relationship between gains and initial status for a school is, for example, positive (i.e., b is positive), then we see that for a student 15 points above the school's initial status value, the expected gain for such a student is *larger* than the mean gain (β_{10}) for his or her school; for a student 15 points below the mean initial status value, the expected gain is *smaller* then the mean gain (i.e., a certain amount is subtracted from β_{10}). Conversely, when the relationship between gains and initial status is negative, we see that for a student 15 points above the school's initial status value, the expected gain for such a student would be *smaller* than the mean gain; for a student 15 points below the mean initial status value, the expected gain will be *larger* than the mean gain.

Note that via the Gibbs sampler, we can obtain the marginal posterior distributions of the expected gains defined in Equations 4a, b and c, which will provide us with estimates and intervals for the expected gains of interest for the schools in our sample.

Results

Descriptive statistics of data. The district in our sample is a diverse, urban district with several schools that are do not meet AYP criteria. As mentioned earlier, of the 72 schools in our sample, 51 are AYP schools and 21 are Non-AYP schools. As can be seen in Table 1, the percent of students eligible for free/reduced price lunch in AYP schools is 24.7% (c.f. district average = 36.4%). In comparison, 69.6% of students in Non-AYP schools are eligible for free/reduced price lunch, which is almost double the district average. In terms of racial composition, more than half of the students in AYP schools (56.5%) are White. In contrast, African American (38.5%) and Asian students (33.4%) together compose the majority of the population in Non-AYP schools.

Table 1

AYP schools vs. Non-AYP Schools on Measures of the Mean ITBS Reading Scores, the Percentage of Students Eligible for Free/Reduced Price Lunch, and Ethnic Composition

	AYP school (N = 51)			Non AYP school (N = 21)		
	# of Students	Mean	SD	# of Students	Mean	SD
ITBS reading at						
Grade 3	1866	196.2	21.1	658	176.5	16.4
Grade 5	1866	226.7	23.9	658	203.9	20.8
% eligible for free/reduced price lunch	24.7%			69.6 %		
Ethnic composition	White: 56.5%, Afr Am: 11.9% Hispan: 9.0%, Nativ Am.: 3.1% Asian: 19.7%			White: 12.9%, Afr Am: 38.5% Hispan: 12.5%, Nativ Am.: 2.7% Asian: 33.4%		

As to the outcome variable, the district means of ITBS reading scores at Grades 3 and 5 are 191.0 and 221.8, respectively, and the corresponding standard deviations are 21.7 and 25.2. The district average gain between Grades 3 and 5 is approximately 30 points. The ITBS vertical scale score is defined such that median growth between Grades 3 and 4 (spring to spring) is 15 standard score units and median growth between Grades 4 and 5 (also spring to spring) is 14 standard score units. The Median gain from spring Grade 3 to spring Grade 5 is 29 standard score units (Hoover, Dunbar, & Frisbie, 2003).

When looking at the AYP schools and Non-AYP schools separately, AYP schools have higher mean scores than Non-AYP schools at Grade 3 by approximately 20 points, and their mean gains are slightly larger than the mean gain for Non-AYP schools. On this superficial level, judged by the different mean scores, we might, at first glance, tend to agree with the AYP classifications made by the state. However, when looking more closely at student growth trajectories of students staring out at the school mean, 15 points below the school mean, and 15 points above the school mean, we find patterns of growth that might call those classifications into question.

Comparisons between AYP schools and Non-AYP schools. In our final analysis, we compare the state's AYP classifications against the results that we obtained by fitting each school's data to the expected gains described above (e.g., LVR-HM_Gain presented in Equations 4a, 4b, and 4c). Specifically, the comparison is made based on the expected gains for each of the three different levels of initial status. First, for each of the 72 schools, we compare AYP and Non-AYP schools' average expected gains. This comparison provides us with information on a) whether there are AYP schools that, on average, *are not* making substantial yearly progress against our measures, or b) whether there are Non-AYP schools that, on average, *are* making substantial yearly progress against our measures. Second, we compare AYP schools and Non-AYP schools based on the magnitude of the expected gain for students 15 points below the school mean. By conducting this comparison focusing on those students starting below the school average, we see whether those that need the most help tend to be left behind in schools that have been classified as having met AYP criteria. In addition, we can see whether there are exceptional Non-AYP schools that are reducing the achievement gap and moving students forward at a substantial rate. Finally, we also compare AYP and Non-AYP schools based on the expected gains for students starting out 15 points above the school mean. Looking at growth trajectories of students starting above the school mean allows us to see if students that are above average are accelerating at different rates than others in the school; that is, if students that started out above average tend to be languishing in certain schools.

1) The expected mean gain (AYP schools vs. Non-AYP schools)

As can be seen in Table 2, we tabulate expected gains for AYP and Non-AYP schools. Note that we classify schools depending upon whether each school's 95% interval for expected gain is above or below the district average gain. In other words, we considered the point estimate of the expected mean gain and the estimation error around it as well in our classification. As can be seen in Figures 1a through 3b. In these figures, the horizontal line represents the district average gain for our sampled 72 schools, and the bars represent each school's expected mean gain and its 95% interval. The top line, middle circle, and bottom line of each bar correspond, respectively, the .975 quantile, mean, and the .025 quantile of the marginal posterior distribution of the expected mean gain for a given school.

These 95% intervals are Bayesian analogues of confidence intervals. Consider the intervals in Figure 1. In the Bayesian framework, the interpretation of an interval for a given school is that the probability is .95 that the true mean gain for that school lies between the lower and upper boundary of the interval. Furthermore the interval for a school provides a way of assessing the plausibility that the true gain for the school exceeds or lies below the district mean gain of 30 points. For example, if the upper boundary of the interval for a given school (i.e., the .975 quantile of its posterior distribution) lies below a value of 30, we know that the probability is greater than .975 that the true gain for this school lies below the district mean gain (e.g., schools 15, 43, 14, 2, 50, 17, 62, 20, 48, 13). Similarly, if the lower boundary of the interval for a school (i.e., the .025 quantile of its posterior distribution) lies above a value of 30, the probability is greater than .975 that the true mean gain for this school exceeds the district mean gain (e.g., schools 68, 10, 34, 45, 25, 72, 22, 32, 29, 7, 8, 9, 55, 23).

Table 2

Cross tabulation of AYP schools vs. Non-AYP schools with respect to the amount of expected gain* at different levels of initial status

Student initial status values	AYP school (N = 51)		Non-AYP school (N = 21)	
	≥ 30[a]	< 30[b]	≥ 30[a]	< 30[b]
School mean	12 schools (7, 8, 9, 10, 22, 25, 29, 32, 34, 45, 68, 72)	4 schools (2,14,15,43)	2 schools (23, 55)	6 schools (13,17,20,48,50,62)
15 points above school mean	9 schools (7, 8, 9, 25, 28, 32, 34, 65, 68)	1 school (43)	5 schools (6, 23, 38, 55, 64)	3 schools (13, 27, 50)
15 points below school mean	7 schools (8, 19, 22, 25, 45, 63, 72)	4 schools (2, 14, 28,30)	0 schools	5 schools (20, 24, 33, 48, 50)

[a] Lower boundary of 95% interval of gain.
[b] Upper boundary of 95% interval of gain.
Note. The classification of the schools was made based on the boundaries of the 95% intervals and not on the estimates (i.e., posterior means) of the expected gains.

As can be seen in Figure 1a, 15 schools among 51 AYP schools have a point estimate of expected mean gain that is smaller than the district average gain. Furthermore, among the rest of the 36 schools, only 12 have a 95% interval situated above the district mean gain. In other words, the low end of the 95% interval for these

12 schools does not cross the district average gain line. Note that 4 AYP schools (2, 14, 15, 43) had considerably smaller expected mean gains compared to the district average gain. Moreover, their 95% confidence intervals did not capture the district average gain. This result suggests that these AYP schools did not make adequate growth between Grades 3 and 5.

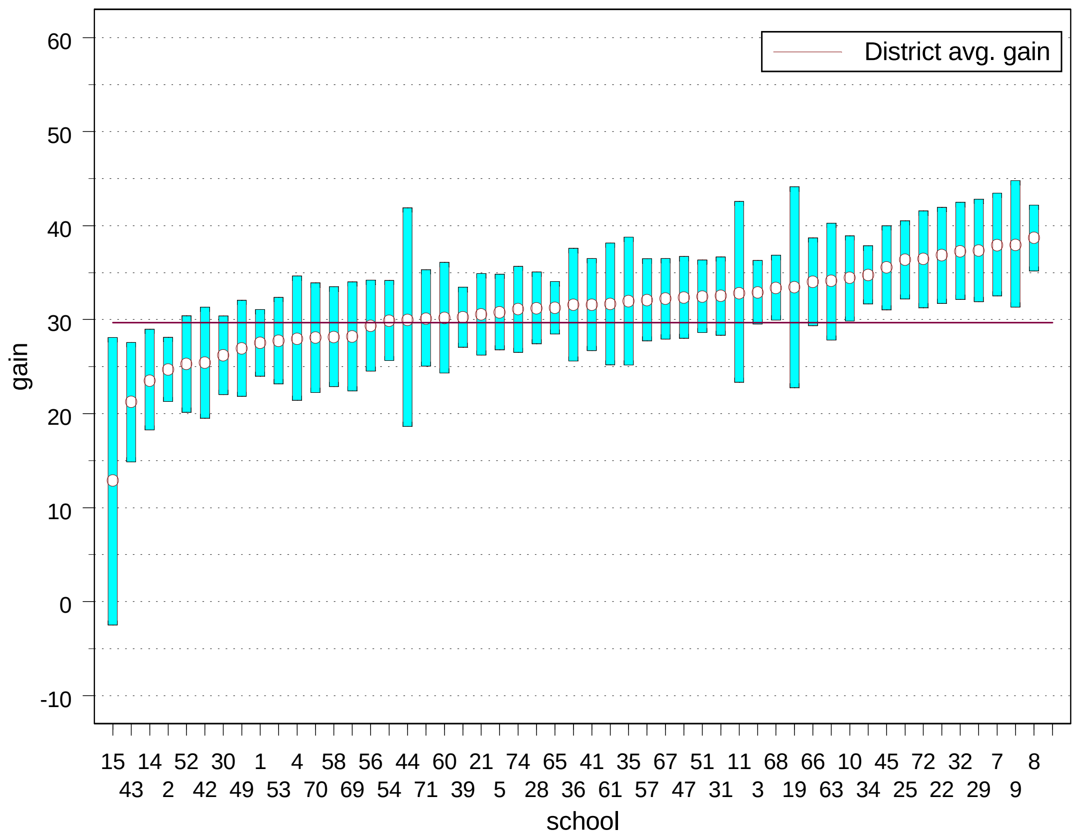

Figure 1a. Expected mean gain in ITBS reading scores for AYP schools. The horizontal line represents the district average gain for our sampled 72 schools. The top line, middle circle, and bottom line of each bar correspond, respectively, to the .975 quantile, mean, and the .025 quantile, of the marginal posterior distribution of the expected mean gain for a given school.

Almost half of the Non-AYP schools (in Figure 1b) have point estimates of expected gains that are higher than the district average. Among those, we have two exceptionally well-performing schools. Schools #23 and #55 have a higher expected mean gain than the district mean gain by approximately 5-10 points, and the lower boundaries of their 95% intervals are above the district average gain. Only one Non-AYP school has an expected mean gain interval that captures a value of 10. Note that 10 points is approximately one standard deviation smaller than the district average gain. As such, we clearly see that there are remarkably large numbers of Non-AYP schools making sizable (close to district average or larger) gains.

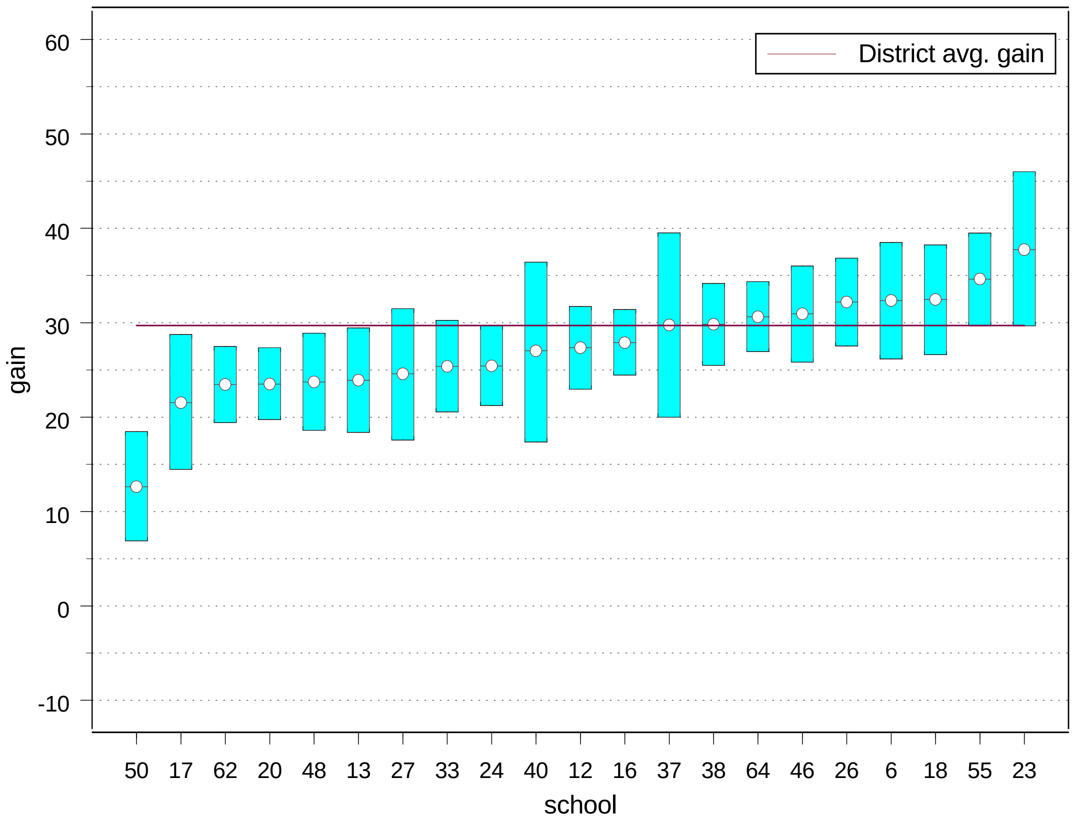

Figure 1b. Expected mean gain in ITBS reading scores for Non-AYP schools.

2) The expected gain for students starting out 15 points below the school mean (AYP schools vs. Non-AYP schools)

For each school, we also examine how much gain is expected for students starting out 15 points below the school mean. The lack of adequate growth for students that start out 15 points below the school mean is a problem for both AYP and Non-AYP schools. As can be seen in Figure 2a, approximately half of the AYP schools (i.e., 25 schools) have point estimates of expected gains that are greater than 30 points. However, there were 4 AYP schools (2, 14, 28, 30) with intervals whose upper boundaries were below the district average gain, and for three of these schools, the point estimates of the expected gains were approximately 20 points. This indicates that there were some students with relatively low initial status values in these schools who were not making adequate progress.

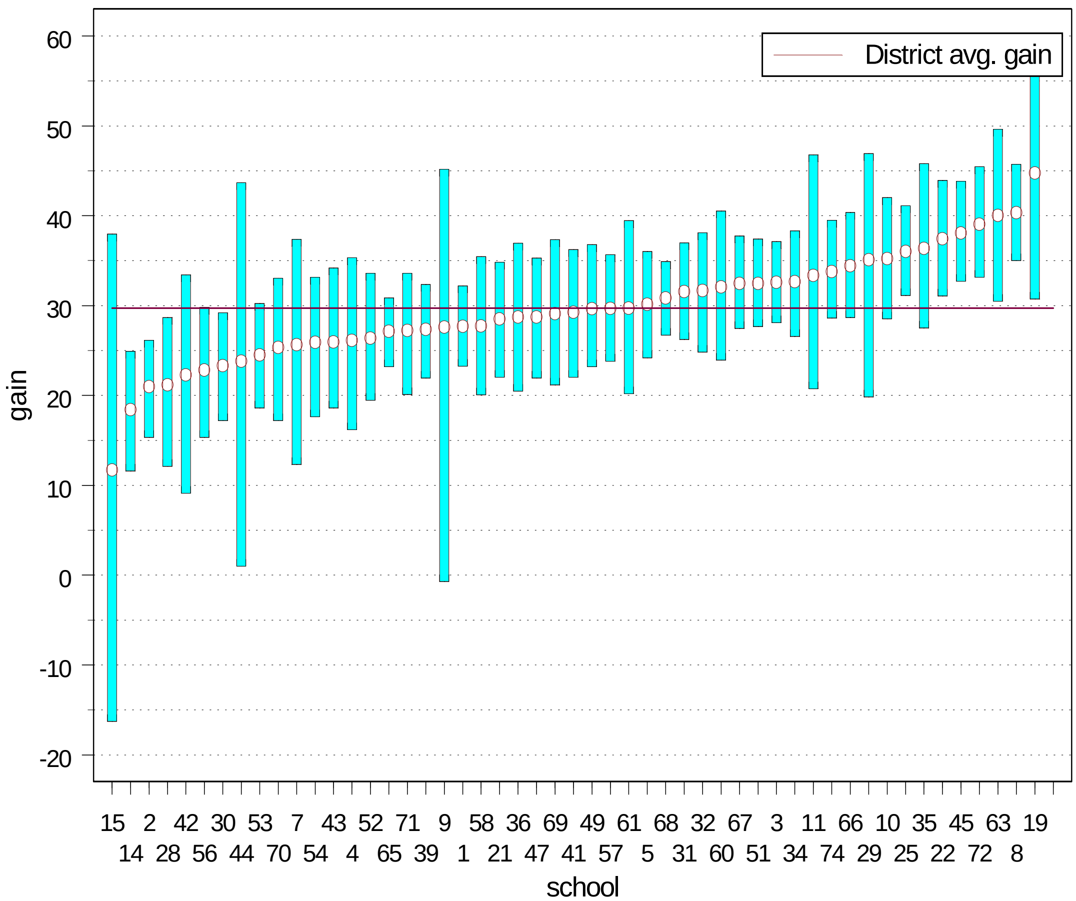

Figure 2a. Expected gain in ITBS reading scores for a student whose initial status value is equal to 15 points below the school mean initial status value (AYP school).

Many of the Non-AYP schools in Figure 2b have point estimates (for students starting out below the mean) that are slightly below the district average; however, 16 of 21 Non-AYP schools have intervals that include the district average of 30. We can see that some of the Non-AYP schools are doing a decent job for students starting out below the school mean. The point estimate of expected gain is larger than the district average gain in 5 schools (#s 23, 46, 27, 26, and 18). However, the expected gains for below average students in five of the non-AYP schools (20, 24, 33, 48, 50) were less than the district average gain by approximately 10 points, and the upper boundaries of the intervals were below the district average line as well.

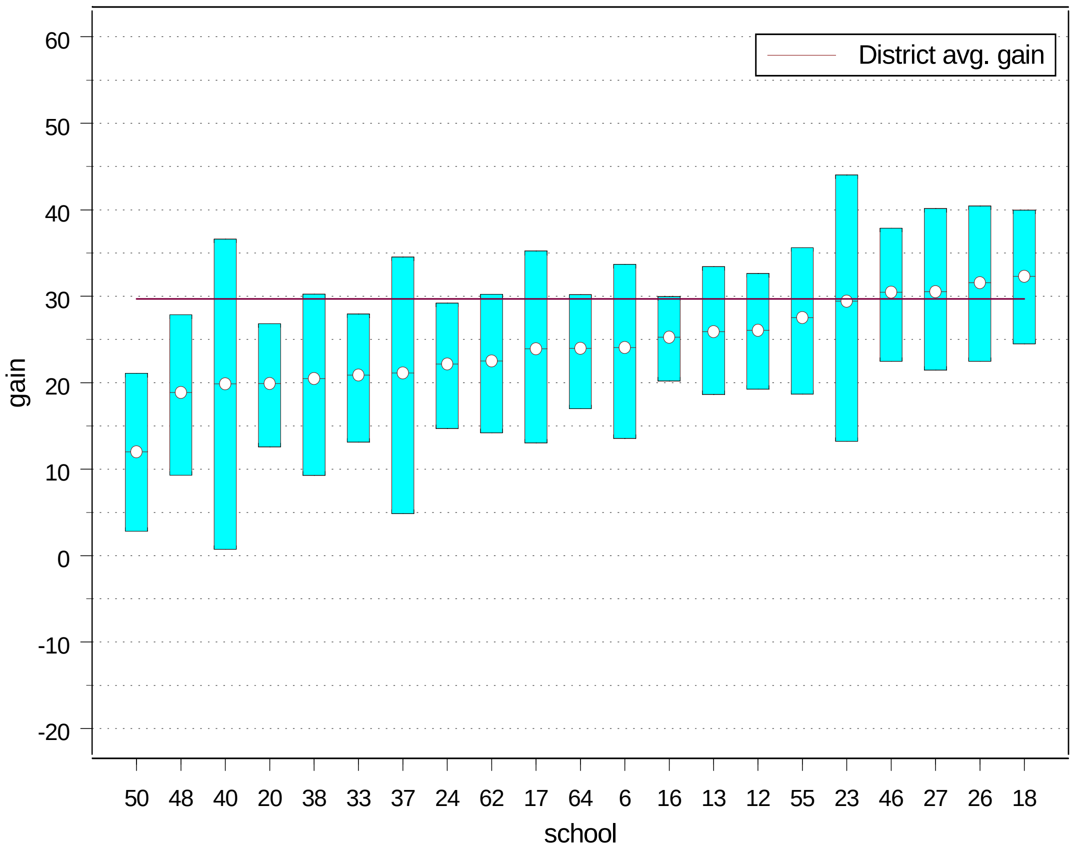

Figure 2b. Expected gain in ITBS reading scores for a student whose initial status value is equal to 15 points below the school mean initial status value (Non-AYP school).

3) The expected gain for students starting out 15 points above the school mean (AYP schools vs. Non-AYP schools)

Figures 3a and 3b display each school's expected mean gains for students starting out 15 points above the school mean.. These figures clearly illustrate that a substantial proportion of schools—both AYP (69%) and Non-AYP (57%)—have point estimates for the expected gain for their above average students that are higher than the district average gain. This implies that students who initially start higher than the school average in a given school are very likely to make adequate gain, regardless of whether they are in AYP or Non-AYP schools.

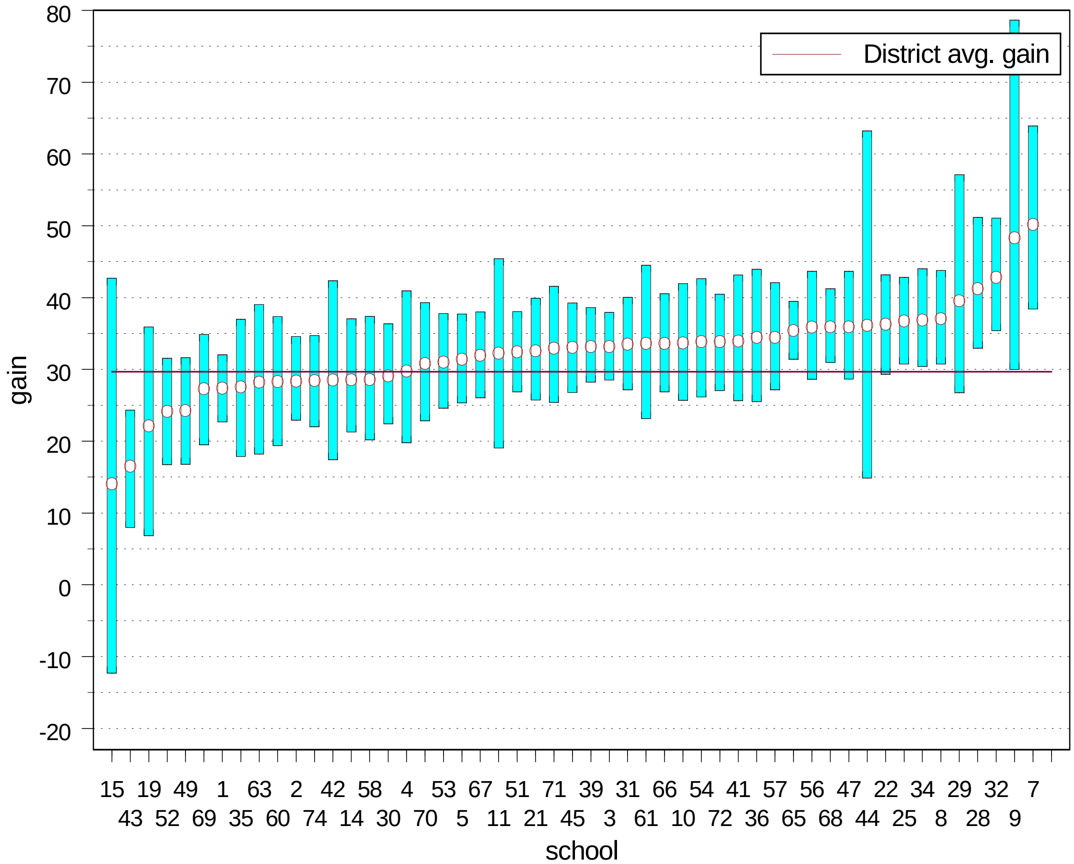

Figure 3a. Expected gain in ITBS reading scores for a student whose initial status value is equal to 15 points above the school mean initial status value (AYP school).

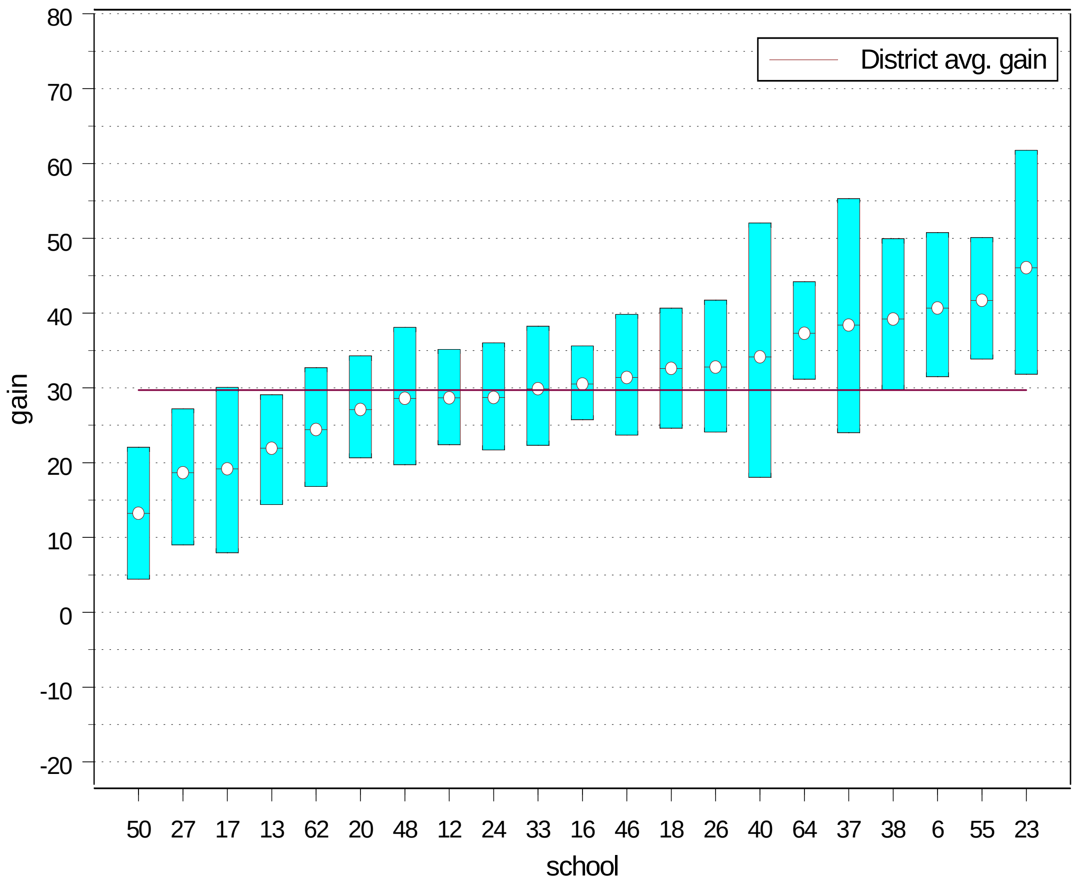

Figure 3b. Expected gain in ITBS reading scores for a student whose initial status value is equal to 15 points above the school mean initial status value (Non-AYP school).

It is noteworthy that there are several AYP and Non-AYP schools in which students with high initial status make large gains. In AYP School #7, for example, the estimate of expected gain is approximately 51 points and the low end of its 95% interval is close to 40 points. Likewise, in three Non-AYP schools (#s 6, 55, and 23), the estimates of expected gains for above average students are greater than 40 points and the corresponding 95% interval falls above the district average gain.

A more interesting finding, however, is that the expected gain for above-average students in one AYP school (43) was approximately equal to 16 points, and the upper boundary of the interval was below the district average gain. In addition, three non-AYP schools (13, 27, 50) also had expected gains below the district average and intervals that excluded the district average. These findings beg the question of whether classifications such as AYP are sufficient if above-average students do not experience adequate gain or progress.

The distribution of student gain in AYP schools and Non-AYP schools. In the above sections, we compared AYP schools with Non-AYP schools in terms of the expected mean gain, and the expected gains for below average and above average students. These analyses show that examining the expected mean gain alone does not provide a complete picture of whether schools may be leaving children behind. Examining the expected gains by different levels of initial status helps us understand how progress varies by performance levels. In this section, we re-examine the distribution of student gains by combining the three performance levels into one picture of school growth. As can be seen in Figure 4, we bring the three expected gain trajectories into one picture. This figure combines Figures 1a, 2a, and 3a in the sense that the circle, triangle, and box lines, respectively, correspond to expected gains for above average, average, and below average students in a given school. Note also that the dotted line (with a diamond icon) represents the district average gain.

Table 3

Distribution of Student Gain for AYP Schools

	15pts > school mean		Equal to the school mean		15pts < school mean	
	Estimate	95% interval	Estimate	95% interval	Estimate	95% interval
Type I						
School #08	37.1	(30.7, 43.8)	38.7	(35.2, 42.2)	40.3	(35.0, 45.7)
School. #22	36.3	(30.0, 43.2)	36.9	(31.7, 42.0)	37.4	(31.1, 43.9)
School. #25	37.7	(30.8, 42.8)	36.4	(32.2, 40.5)	36.0	(31.1, 41.1)
Type II						
School #19	22.1	(6.9, 35.9)	33.5	(22.8, 44.1)	44.8	(30.7, 60.3)
School #63	28.2	(18.2, 39.0)	34.1	(27.8, 40.3)	40.0	(30.5, 49.6)
Type III						
School #28	41.2	(33.0, 51.2)	31.2	(27.4, 35.1)	21.2	(12.1, 28.7)

Note. Expected gains for students whose initial status value is equal to the school mean initial status, at 15 points above the school mean initial status, and at 15 points below the school mean initial status.

These patterns of school growth allow us to note another trend in the growth patterns of these schools, one that involves the closing or widening of achievement gaps in these schools. In what we refer to as Type I schools, the growth trajectories for the three performance subgroups are parallel, while the other types of schools have growth trajectories that either converge or diverge. In Type II schools, the initial gap between above average and below average students in the school closes over time, while the initial gap gets magnified in Type III schools. Much attention is paid to the closing of the achievement gap, and while this phrase is most often used to refer to the racial gap in achievement that exists in many schools, Type II schools in our sample are making progress in closing the gap between an important pair of subgroups – those that start out below average and those that start about above average. When evaluating a school's effectiveness or quality, it is difficult from an ethical point of view to choose the better of Type II or Type III schools. However, from a policy point of view, more resources and sanctions are levied against schools in the Type II category, e.g., through Title I (i.e., No Child Left Behind) funds.

First, consider the distribution of student gains for the illustrative AYP schools presented in Figure 4. We see the distribution of student gain for School #8 as an example of the Type I schools. In this school, it can be seen clearly that all three expected gain trajectories are parallel to each other, and steeper than the district mean gain trajectory. Note that this school has only its 5.7 % of students eligible for free/reduced price lunch. Two other AYP schools (#22 and #25) are classified as Type I schools (see Table 3).

Type II AYP schools (#19 and #63) do not seem to be making exceptional progress with their above average students, but are succeeding with their below average students. In school #19 (see Figure 4), the point estimate of the expected gain for above average students is 22.1 and the 95% interval ranges from 6.9 to 35.9. In contrast, for below average students, the estimate of the expected gain is equal to 44.8 and its 95% interval ranges from 30.7 to 60.3. Given this large difference between above and below average students in their expected gain, the initial gap is substantially diminished over time. Closing the gap among students is very important in addressing equity concerns; however, this exemplary Type II growth pattern presents problems concerning the relatively lower amounts of progress for above average students.

Type III AYP schools (#28) demonstrate a reverse growth pattern of that found in Type II schools. As can be seen in Figure 4, the gap in school #28 is magnified in this school, since below average students are making far slower progress than above average students. In school #28, the estimate of the expected gain for above average students is 41.2 (its 95% interval ranges from 33.0 to 51.2), which is approximately double the expected gain for below average students (expected gain = 21.2; 95% interval = (12.1, 28.7)). While this school (#28) met current AYP criteria, this school is not adequately, or equitably, distributing gains among its student subgroups. Those that need the most help in this school seem to be struggling more. Thus, while this school has been identified as making adequate yearly progress, this analysis raises the question: Adequate progress for whom?

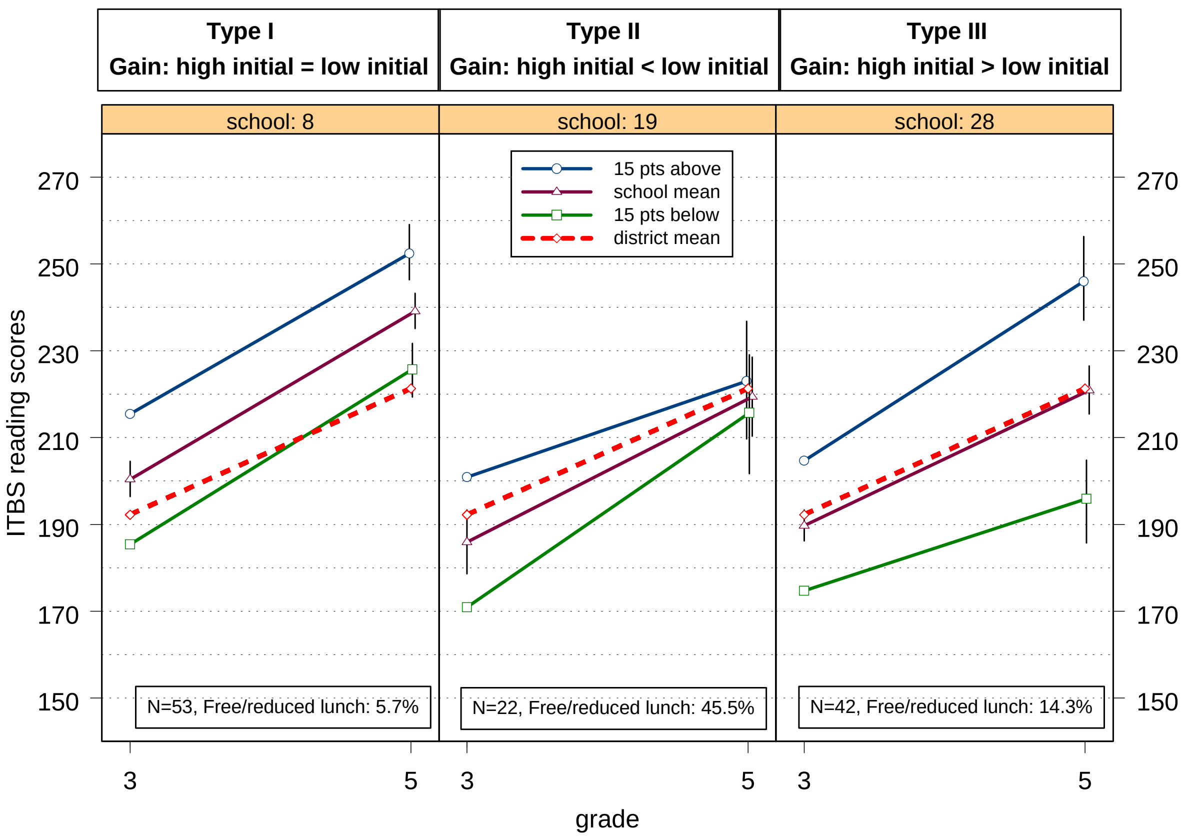

Figure 4. Distribution of student gain for AYP schools. Type I School: Substantial gain of more than 30 points across all performance subgroups. Type II School: Closing the gap. Type III School: Widening the gap.

Similarly, we present Non-AYP schools with the same three types of distribution of student gain in Table 4 and Figure 5. Though the amount of gain is slightly smaller than for Type I AYP schools, one Non-AYP school (#26) displays the most equitable Type I growth pattern (all subgroups growing in parallel). As we can see in Figure 5, school mean initial status for school 26 is approximately 20 points below the district average mean initial status, and even the above average subgroup starts out slightly below the district average. However, all three expected trajectories are, by and large, parallel to—or slightly steeper than – the district average trajectory. Specifically, the estimates of expected gain are 32.8, 32.2, and 31.6, respectively, for above average, average, and below average students. Given the higher proportion of low SES students in this school (66.0% of students are eligible for free/reduced lunch), the amount of expected gain and the distribution of student gains that this school achieves would seem to merit special attention, despite the fact that this school did not meet AYP criteria.

Table 4

Distribution of Student Gain for Non-AYP Schools

	15pts > school mean		Equal to the school mean		15pts < school mean	
	Estimate	95% interval	Estimate	95% interval	Estimate	95% interval
Type I						
School #26	32.8	(24.1, 41.7)	32.2	(27.6, 36.9)	31.6	(22.5, 40.4)
Type II						
School #27	18.7	(9.0, 27.2)	24.6	(17.6, 31.5)	30.5	(21.5, 40.1)
Type III						
School #06	40.7	(31.5, 50.8)	32.4	(26.2, 38.5)	24.1	(13.6, 33.7)
School #38	39.2	(30.0, 49.9)	29.9	(25.5, 34.2)	20.5	(9.3, 30.2)
School #64	37.3	(31.2, 44.2)	30.6	(26.9, 34.4)	24.0	(17.0, 30.2)

Note. Expected gains for students whose initial status value is equal to the school mean initial status, at 15 points above the school mean initial status, and at 15 points below the school mean initial status.

In addition, school #27 represents a Type II Non-AYP school that appears to be struggling. Almost all students in this school are eligible for free/reduced price lunch. The mean initial status is 25 points below the district average. However, this school is doing a reasonable job of making progress with many students, particularly the below average students: the estimate of the expected mean gain is close to 25 points and the expected gain for below average students is greater than 30 points. As a result, the gap between above and below average students is closing. In fact, the intervals for the expected fifth grade status values of the three performance subgroups are overlapping substantially, indicating that that the gap remaining between the subgroups is fairly negligible.

Finally, school #6 represents our Non-AYP Type III School, where above average students are progressing at a faster rate than are others in the school. Again, however, this school is another struggling school, as demonstrated by the higher performing students in this school starting out below the district average. Those students near the district average, however, made the steepest gains (i.e., estimate = 40.7, 95% interval = (31.5, 50.8)), with those students starting below the school mean gaining far less (i.e., estimate = 24.1, 95% interval = (13.6, 33.7)).

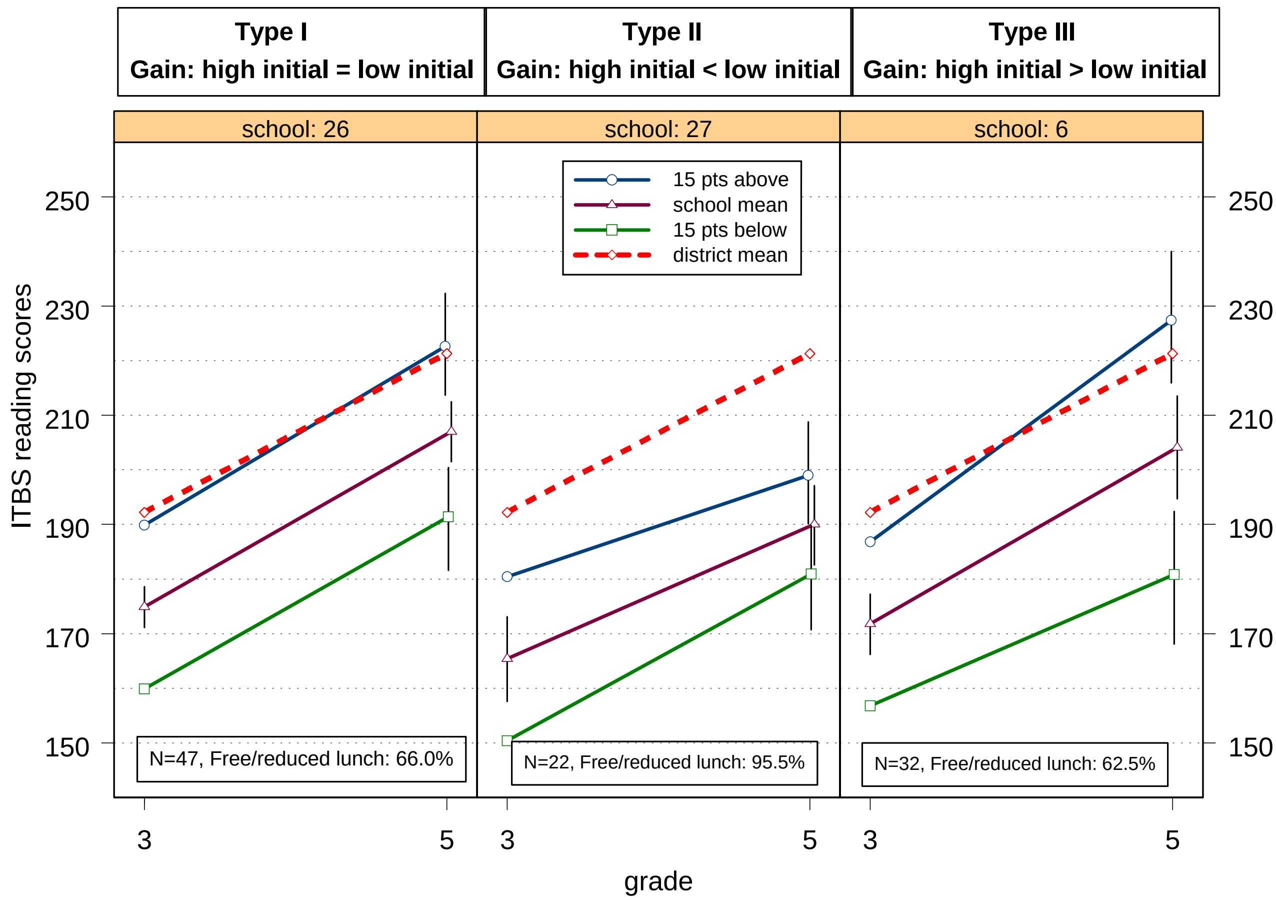

Figure 5. Distribution of student gain for AYP schools. Type I School: Substantial gain of more then 30 points across all performance subgroups. Type II School: Closing the gap. Type III School: Widening the gap..

Discussion

Our analyses of estimating expected gains across different performance subgroups provides an alternative framework for characterizing whether and how schools are achieving NCLB's underlying precepts and also provide food for thought in examining the validity of the current AYP school designations. Based on the analysis of how student gain is distributed within a school, the methodology illustrated through this study helps bring to light potentially important aspects of school performance that can be masked if we limit our focus, as AYP currently does, to classifying schools based on a single school-wide measure, or to measures that focus on changes at a cutpoint. While in looking at general school-wide averages, our analyses do confirm that AYP schools tended to show significantly more growth in student performance than did non-AYP schools, the details of who within schools showed progress was illuminating. For example, students in non-AYP schools with relatively low initial status values were least likely to show exceptional gains. Because non-AYP schools generally serve a higher proportion of students in poverty, more ethnically and language diverse students and perform more poorly academically, this finding suggests that students most at risk are still being left behind.

At the same time, we were able to identify some schools that failed to meet AYP who were making exceptional progress, relative to the district, for their average and above average initial status students. These successes were masked by their AYP designation. We also found some AYP schools with intervals for low initial status students that fell below the District average of 30 points. In addition, we found some AYP schools in which the expected gains for high initial status students were below the District average, while the expected gains for low initial status students clearly exceeded the District average. These patterns show major differences in whether and how the achievement gap may be closing and who may be "left behind."

These contrasts raise questions about the meaning of "adequate" progress and of closing the achievement gap, for whom we wish to target progress, and the basis on which we label schools as making AYP or not. These are primarily policy rather than methodological issues. Nonetheless, we believe that closely examining the distribution of student achievement provides vital perspective and may be a key in supplementary or alternative measures of AYP. Our hope is that our approach provides a more informative picture of student progress beyond a single AYP indicator and, perhaps most importantly, may stimulate discussion among teachers and administrators to identify students in need earlier (see Seltzer et al., 2003). We also believe that encouraging educators to think about achievement levels rather than (or in addition to) current subgroup categories may be both more productive and actionable.

Our approach to AYP measures can be extended in many ways. First, our approach could easily include differences in student and school characteristics. As we saw in an illustrative Non-AYP school, and as the literature suggests, low SES schools have a more difficult time in making AYP than do high SES schools, and more diverse schools have a harder time making AYP than do less diverse schools (Novak & Fuller, 2003). As we know, it is possible to take these factors into account in evaluating school performance. If we re-estimate the distribution of student progress using a model where we include and adjust for differences in student background and in school characteristics (Choi & Seltzer, 2003), we may see downward shifts in the posterior means and intervals for expected gains in many schools that are currently designated as making AYP.

It is also possible to extend our approach in settings where we wish to compare the performance of different cohorts of students. Our analyses focused on one cohort of students who were in Grade 3 in 2001. However, we might see some differences between this cohort and the 2000 or the 2002 cohort of students in this district. Using multiple measures in multiple cohorts of students (i.e., multi-panel multiple-cohort data), we would be able to examine differences in initial status and expected growth among and between the cohorts of students (see Bryk et al., 1998). This longitudinal perspective is important because particular school reform efforts or changes in school characteristics that may impact achievement schoolwide or at a particular grade level might take place in some years and not others (e.g., when affected by increment of school budget, numbers of qualified teachers, student demographic composition, and so on). This multi-panel, multi-cohort design would increase the robustness of our design by adding a longitudinal component, and would also help us to track the cohort differences due to these various factors.

References

Bryk, A., Thum, Y. M., Easton, J., & Luppescu, S. (1998). Assessing school academic productivity: The case of Chicago school reform. *Social Psychology of Education*, 2, 103-142.

Choi, K., & Seltzer, M. (2003). Addressing questions concerning equality in longitudinal studies of school effectiveness and accountability: Modeling heterogeneity in relationships between initial status and rates of change. CSE technical report. Los Angeles, CA: Center for the Research on Evaluation, Standards, and Student Testing, University of California Los Angeles.

Hoover, H.D., Dunbar, S.B., & Frisbie, D.A. (2003). *Iowa Tests of Basic Skills: Guide to Research and Development, Forms A and B*. Chicago: The Riverside Publishing Company.

Novak, J., & Fuller, B. (2003). Penalizing diverse schools? Similar test scores, but different students, bring federal sanction. PACE Policy brief. Berkeley, CA: Policy Analysis for California Education.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, 2nd edition*. Newbury Park, CA: Sage Press.

Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining Relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Education Evaluation and Policy Analysis*, 25, 263-286.

Spiegelhalter, D., Thomas, A., & Best, N. (2002). *WinBUGS: windows version of Bayesian inference using Gibbs sampling, version 1.4, User Manual*. MRC Boistatistics Unit, Cambridge University.

Thum, Y. M (2003). *No Child Left Behind: Methodological Challenges & Recommendations for Measuring Adequate Yearly Progress* CSE Technical Report 590. Los Angeles, CA: Center for the Research on Evaluation, Standards, and Student Testing, University of California Los Angeles.