



---

**KINGSBURY**

CENTER AT NWEA

# Achievement Gaps and the Proficiency Trap

Michael Dahlin and John Cronin

January 2010





## Table of Contents

INTRODUCTION .....	page 2
THE PROBLEM: FIFTY STATES – FIFTY ACHIEVEMENT GAPS.....	page 2
ONE ALTERNATIVE: VIEWING ACHIEVEMENT GAPS THROUGH PERFORMANCE DISTRIBUTIONS .....	page 4
ANOTHER ALTERNATIVE: VIEWING ACHIEVEMENT GAPS AS GAPS IN STUDENT GROWTH.....	page 6
CONCLUSIONS .....	page 7
REFERENCES .....	page 8

## INTRODUCTION

The achievement gap was a term coined to describe disparities in academic performance between groups of students. Such discussions typically focus on disparities between white students and their African American or Hispanic peers, or between students from middle/high income families and students from families with low income. While the achievement gap is most commonly used to refer to differences in performance on some form of achievement test, it is also used to characterize differences in college completion rates, dropout rates, or course selection (Education Week, September 10, 2004). Closing achievement gaps was a focal point of school reform efforts during the George W. Bush administration and it continues as a priority in the Obama administration, which has committed \$5 billion in stimulus fund to efforts to spur innovation that will close the achievement gap (White House, 2010).

Under the No Child Left Behind Act (NCLB), schools are required to measure the difference in proficiency rates between two or more ethnic or socioeconomic groups of students on their state assessment. For example, if a school reported that 75% of its white students were proficient in mathematics, while only 55% of African-American students met this standard, the school would be said to have a proficiency gap of 20 percentage points. NCLB's requirement that schools achieve one-hundred percent proficiency for all students, including traditionally disadvantaged subgroups, theoretically means that proficiency gaps should be eliminated by 2014. Many parents and educators believe that eliminating proficiency gaps will, by definition, achieve equity within schools.

Defining the achievement gap in terms of proficiency rates, as NCLB requires schools to do, obscures many of the inequities within schools that NCLB was intended to eliminate.

Even if schools are successful in eliminating proficiency gaps by 2014, achievement gaps will very likely persist. After all, just because all students cross a threshold of proficiency does not mean that all students' achievement is equally beyond the threshold. Jennifer Jennings and Sherman Dorn characterized this very phenomenon as the proficiency trap. After all, having all students pass a standard does not necessarily demonstrate that all students perform equally well. The definitions and measures that we use have a profound impact on the results we achieve, and defining achievement gaps as differences in proficiency rates both misrepresents the nature of the problem and points educators toward solutions that won't resolve it.

## THE PROBLEM: FIFTY STATES – FIFTY ACHIEVEMENT GAPS

Imagine two groups of able-bodied high school students, one comprised of track and field athletes, the other made up of students who have never competed in sports. Imagine also that we want to test the rates of athletic proficiency in these two groups, and that our measure for determining athletic proficiency is the ability to jump over a twelve-inch hurdle. With such an easy proficiency standard, one would expect almost no proficiency gap between the athletes and non-athletes, since nearly everyone in both groups would be capable of jumping that high. Nor would there be a proficiency gap if the hurdle were set at twelve feet, because no student in either group could meet that standard, and so both groups would fail equally. Even though there might be no proficiency gap in this example, there is certainly a profound achievement gap between the two groups of students. It is highly unlikely that the non-athletes would be able to jump, on average, as high as the track athletes.

Our point is that proficiency gaps can be deceptive because the size of the gap depends on where the proficiency cut score is placed. More importantly, one can eliminate a proficiency gap without resolving the underlying achievement gap simply by raising or lowering the standard to the point where all groups either pass or fail. And even if we somehow manage to get all students over the hurdle without changing the proficiency standard, this is no assurance that achievement gaps among the groups have been eliminated. Some groups may still do better than others.

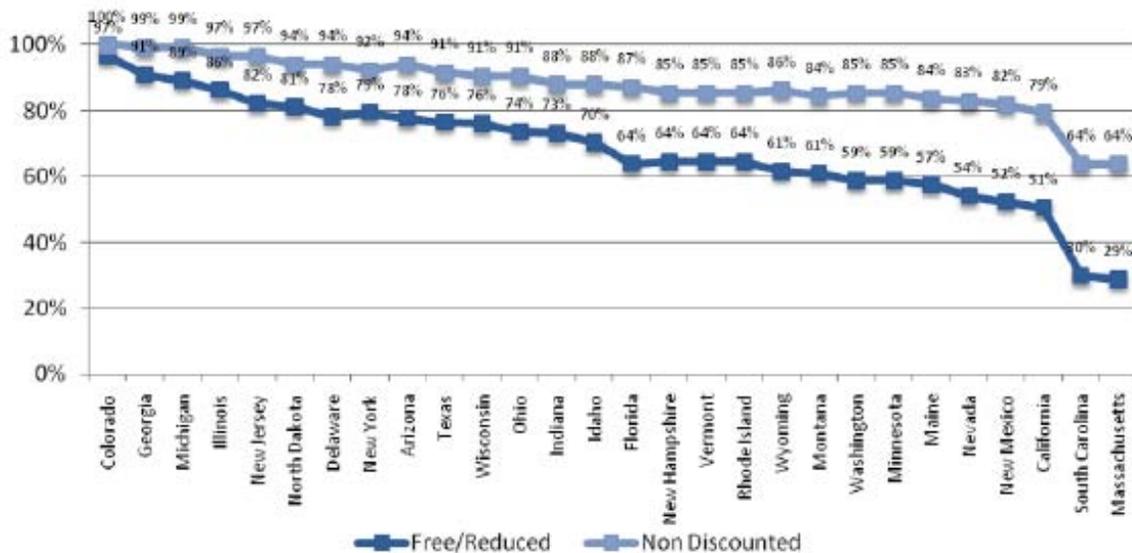
State test proficiency standards are essentially academic hurdles, and thanks to NCLB, those hurdles are set at different heights in nearly every state. The Kingsbury Center at NWEA recently completed a study (Cronin, Dahlin, Xiang, & McCahon, 2008) in which we evaluated the performance of real students in 36 actual schools relative to the proficiency cut scores (that is, the minimum score on the state test corresponding to proficiency) of 28 states (Cronin, Dahlin, Xiang, & McCahon, 2009). Data from one of these schools, Alice Mayberry Elementary (a pseudonym), are shown in Table 1.

This table shows the average math achievement scale scores (and their corresponding norm-based percentile ranks) for two groups: students from families with low-income (i.e., eligible for free or reduced price lunches), and students not eligible for such assistance. These data show a school with many high performing students (as indicated by the highest percentile ranks), but with substantive achievement differences between low-income and other students.

**Table 1 – Mathematics performance of Alice Mayberry Elementary on spring 2006 administration of Measures of Academic Progress®**

Grade	Non-Discounted Students		Students Eligible for Free or Reduced Lunch	
	Average Scale Score	Percentile Rank	Average Scale Score	Percentile Rank
3	209.4	94th	202.1	81st
4	222.8	97th	212.7	78th
5	232.8	93rd	215.5	63rd

**Figure 1 – Mayberry Elementary School’s Mathematics Proficiency Gap in 28 States**



This difference in average scale scores between low income and other students could certainly be called an achievement gap, but how substantial is that gap when the scale scores are reduced to proficiency rates? Figure 1 shows that when Mayberry students are evaluated by lower proficiency cut scores, such as those used in Colorado, Georgia, and Michigan, the proficiency gap between low-income and non-low-income Mayberry students is relatively small, ranging between three and ten percentage points. But if Mayberry were evaluated using the higher proficiency standards of states such as California, South Carolina, and Massachusetts, the gap is considerably larger, between 28 and 35 percentage points. In other words, a single school would have 28 different sized achievement gaps, all dependent upon how stringently the state chooses to define proficiency. If Mayberry happened to be located in Colorado, educators and parents would likely be pleased to learn that there is virtually no achievement gap in their school for students from low income families. Had fate placed Mayberry in Massachusetts instead, those same parents and educators would be shocked and appalled to find an almost insurmountable gap.

Imagine also how educators might react to these data in different states. A Michigan educator might be delighted that proficiency rates for their low-income students were high and that the achievement gap seemed relatively low. For Michiganders, the gap certainly does not appear to be a crisis, and educators might try to address it with relatively modest measures, perhaps focusing their efforts on the lowest performing 10% of the low-income population. The other 90% of low income students, having met proficiency standards, presumably need no improvement plan.

A Massachusetts educator would see an entirely different story in the same data. A 35 point

achievement gap is a crisis, and one unlikely to be resolved with minor school program tinkering. With 71% of the low-income students failing to meet proficiency standards, it is unlikely that all of them could be elevated to proficiency right away. In such a case, students who were not performing near the standard might be triaged in order to focus improvement efforts on the bubble students who could reasonably be expected to help the school make AYP. In short, the same data drawn from the same school and students would produce two vastly different school improvement plans, depending on the standards used.

The phenomenon we described repeated itself across all of the 36 schools we studied. In each school, the size of the proficiency gap varied based on where one set the proficiency bar. Lower proficiency bars had a tendency to diminish the perceived size of the achievement gap, while proficiency bars set in the middle of the distribution made achievement gaps more visible.

## ONE ALTERNATIVE: VIEWING ACHIEVEMENT GAPS THROUGH PERFORMANCE DISTRIBUTIONS

The main weakness of proficiency ratings is that they provide no information about students' actual performance, other than whether they meet or exceed a state's single arbitrary threshold. In that sense, such ratings are similar to the information one might get from a hypothetical bathroom scale designed only to measure whether someone is "Fat/Not Fat". That kind of scale begs the more meaningful question, "How overweight am I?" or put another way "What's the gap between my current weight and my target weight?" The accountability structure in place for NCLB does nothing more than provide a scale which returns "Proficient/Not Proficient". We need something better.



Just as students' heights and weights can vary across a range (or distribution) of scores, so do state achievement test scores. Figure 2 shows the distributions of math achievement scores for low income and non-low income fifth grade students (data come from Cronin, et al., 2009), with the distribution for low income students shown in dark blue, and non-low-income students shown in light blue. Also shown is a hypothetical proficiency threshold at 200 on the scale of the assessment.

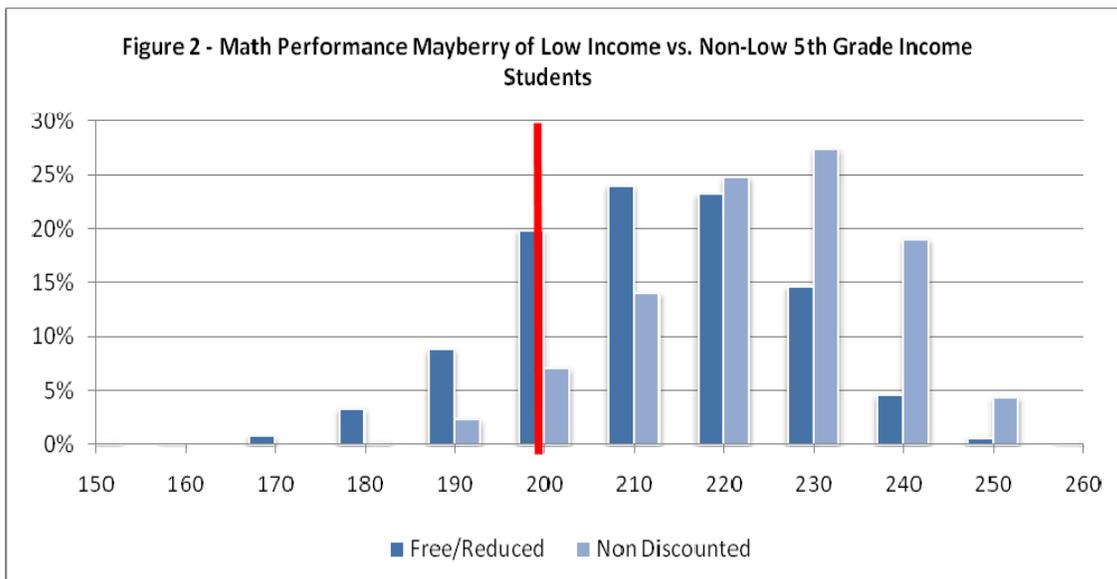
Information about the entire distribution is much more informative than merely knowing what percentage of students fall above or below a threshold value, since the distributions also show the high degree of overlap between the two groups, and more clearly illustrate the relative number of high and low-performers within both groups – a fact easily overlooked when considering only proficiency rates.

Distributions can also be used to measure achievement gaps by asking the question, “Do the distributions of Group A and Group B differ?” When we define achievement gaps using entire distributions rather than proficiency rates, we

make use of the information from all students, not just the ones close to the threshold value.

This is a far more equitable approach, and eliminates the possible temptation to focus only on “bubble kids” very close to the threshold value. It also has the advantage of a century’s worth of scientific precedent, since comparisons of distributions are the primary statistical methods employed by researchers to demonstrate differences in group performance.

Finally, the use of a performance distribution discourages stereotyping groups of students by forcing the people using the data to consider all of the students in the display. In the above display, it is obvious that, while more low income students are low performing than non-discounted students, low income students are not necessarily low performing. Large numbers of them perform in the middle and upper ends of the distribution. Given that fact, one can’t solve the achievement gap by merely focusing on the low end of the distribution, teachers must also focus on making middle performers high performers and they must help high performers reach their full potential.





### ANOTHER ALTERNATIVE: VIEWING ACHIEVEMENT GAPS

Several states are currently experimenting with strategies for measuring change over time, or growth, within the NLCB accountability framework. Like the approach described above, which focuses on entire distributions of performers, growth models may be viewed as more equitable because they give equal emphasis to all students within a school, rather than focusing only on so-called “bubble” students whose performance puts them near the proficiency standard (for example, near the red line in Figure 2). Yet raw growth, without additional context, cannot provide sufficient information to evaluate an individual’s progress. Just as a weight loss of five pounds is more serious for a 10 pound newborn than a 195 pound adult, raw growth cannot be fully correctly interpreted without also knowing about a student’s age and prior ability, or without some standard for what constitutes “typical” growth.

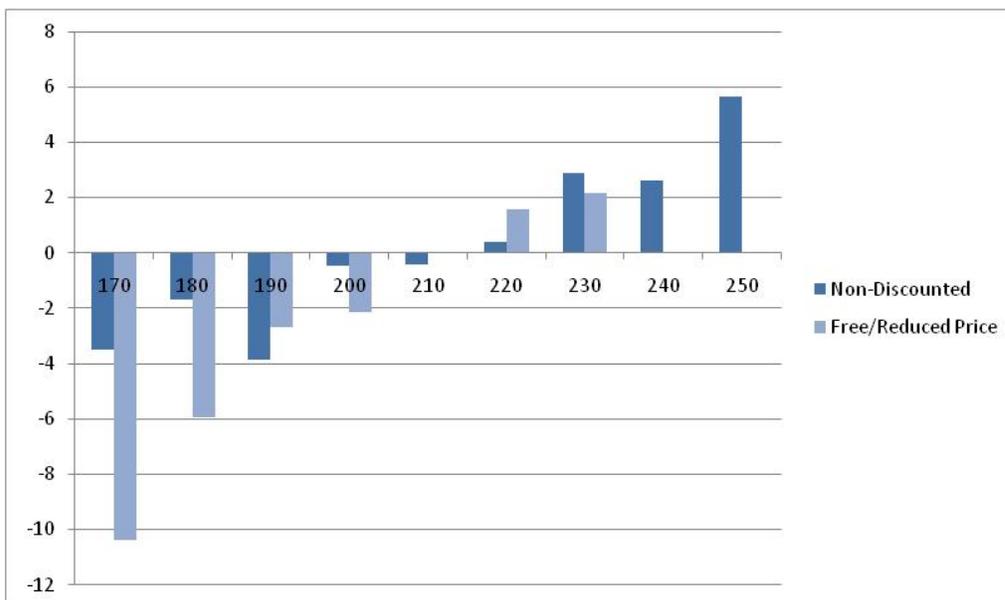
Figure 3 shows just such a solution, depicting average fall-to-spring growth for middle school students at 18 real middle schools across the

### AS GAPS IN STUDENT GROWTH

country, differentiating between students receiving free/reduced price lunches (low income) and higher income groups (data taken from Cronin, et al., 2009). This figure illustrates average “growth index” scores for students at differing initial ability levels, where the growth index is the difference between observed growth and the growth that is typical for students who achieved the same beginning score (Northwest Evaluation Association, 2008). In this context, growth index scores greater than zero imply average growth that is greater than normal, given age and starting achievement, whereas growth index scores less than zero imply less than average growth.

What’s interesting about this approach is that it shows whether schools are creating achievement gaps where none previously existed. For example, consider the students in Figure 3 who started with a score between 180 and 189. Non-discounted students in these schools, on average, lost just under 2 scale score points relative to the NWEA norming group.

**Figure 3 - Fall to Spring Average Growth Index Scores of 18 Real Middle Schools**



Sadly, low income students lost even more, an average of just under 6 points relative to students who started with the same score. So Figure 3 depicts a case in which students who did not have an achievement gap at the beginning of the year, finished with an achievement gap because the students showed different rates of growth.

The advantages to illustrating the growth of low income students in this manner are twofold. One advantage is that the growth index controls for differences in growth attributable to age or starting ability in a way that would not be possible using observed raw growth. This makes it possible to compute averages across age groups in a meaningful way. Another advantage of the growth index statistic is that it permits more meaningful school level comparisons. For example, if a school failed to make AYP over sufficient years, NCLB permits parents to transfer their children to an alternative school. Under such a scenario, the growth index would provide the best comparative indicator for determining whether a student would likely be better off at another school, since it is unlikely that a student would fare better at a school with lower growth indicators. Comparisons of such specificity are simply not possible using the current NCLB school performance metrics.

## CONCLUSIONS

One of the main goals of NCLB was to hold schools accountable for ensuring that all students are meeting high academic standards, and to eliminate disparities in academic performance between traditionally advantaged and disadvantaged groups of students. One could argue that this focus on holding schools accountable for the performance of individual subgroups, rather than considering only aggregated school-wide performance, is a step in the right direction. Still, the metrics specified within NCLB to evaluate school and sub-group

performance (i.e., group proficiency rates), are inadequate for evaluating whether progress has been made towards eliminating racial and/or socio-economic gaps in academic achievement.

As shown in Figure 1, proficiency rates are largely a function of the difficulty of state proficiency standards, which cannot be directly measured or compared when states use different tests and scales. Only when state the proficiency standards are mapped or expressed on a single common scale can proficiency rates provide real information about the relative differences between the groups in question. Furthermore, even when proficiency standards can be expressed on a common scale, the rate itself tells very little about the performance of the groups of interest, other than what percentage meets or exceeds the standard. None of the information conveyed by Figure 2, such as the range of abilities within groups, and the degree to which groups are performing equivalently, can be inferred from group proficiency rates. Finally, nothing about the growth of students, as conveyed in Figure 3 can be inferred from group proficiency rates, and it is growth information that is most relevant when determining how effectively schools are teaching children.

Simply put, even though the goal of NCLB was to eliminate disparities in achievement among groups of students within schools, the performance metrics required under NCLB make it nearly impossible to determine whether schools are actually making progress towards these ends. Only by considering the full distribution of student performance, and by considering growth information along with performance information, can a complete picture be revealed about gaps in academic achievement, and whether schools are making adequate progress towards eliminating such gaps. Using alternative measures such as the ones described here will be critical in that effort.

## REFERENCES

Cronin, J., Dahlin, M.P., Adkins, D., & Kingsbury, G. (2007). *The Proficiency Illusion*. Washington, DC: Thomas B. Fordham Institute. Retrieved from [http://edexcellence.net/detail/news.cfm?news\\_id=376&id=130](http://edexcellence.net/detail/news.cfm?news_id=376&id=130).

Cronin, J., Dahlin, M.P., Xiang, Y., & McCahon, D. (2009). *The Accountability Illusion*. Washington, DC: Thomas B. Fordham Institute. Retrieved from [http://edexcellence.net/index.cfm/news\\_the-accountability-illusion](http://edexcellence.net/index.cfm/news_the-accountability-illusion).

Education Week (2004, September 10). *Achievement Gap*. Retrieved from <http://www.edweek.org/rc/issues/achievement-gap/> on August 6, 2009

Jennings, J. & Dorn, S. *The Proficiency Trap: New York City's Achievement Gap Revisited*. Teachers College Record, September 18, 2008. <http://www.tcrecord.org>. ID Number: 15366, Date Accessed: 8/11/2009 2:12:20 p.m.

Northwest Evaluation Association. (2003). *Technical Manual for the NWEA Measures of Academic Progress and Achievement Level Tests*. Portland, OR: Author.

Northwest Evaluation Association. (2008). *RIT Scale Norms For Use with Measures of Academic Progress*. Portland, OR: Author.

White House (2010). *Education*. Retrieved on January 19, 2010 from <http://www.whitehouse.gov/issues/education>.