10-13-2011

NEPC Review: Do High Flyers Maintain Their Altitude?

Jaekyung Lee
SUNY University at Buffalo, JL224@buffalo.edu

NEPC
NATIONAL EDUCATION POLICY CENTER

Review of *Do High Flyers Maintain Their Altitude?*

Reviewed By

Jaekyung Lee

University at Buffalo, SUNY

October 2011

Summary of Review

The research report reviewed here concludes that many initially high-achieving students are falling further and further behind over the course of their years in school. The report intends to raise the alarm and to advocate for improved programs for these students. It is, however, a false alarm due to biased methodology and misleading arguments. The report's norm-referenced framework guarantees "losers" as well as "winners," regardless of any true improvement made by the students. Also, the "regression to the mean" effect produces a false illusion of a tradeoff of over-progress by low achievers at the cost of under-progress for high achievers. Finally, its prescription for stronger school accountability for high-achieving students under NCLB does not follow research-based guidance on how to improve student learning. Other research, including that conducted by this reviewer, finds that students who are high achievers and low achievers make approximately equal academic progress in reading and math, while the achievement gaps remain large. Moreover, a substantial part of the variation in student progress is attributable to teacher and school effects beyond students' initial status and background characteristics.

Kevin Welner
Editor

William Mathis
Managing Director

Erik Gunn
Managing Editor

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: 303-735-5290
Fax: 303-492-7090

Email: NEPC@colorado.edu
http://nepc.colorado.edu

Publishing Director: Alex Molnar

GREAT LAKES CENTER
FOR EDUCATION RESEARCH & PRACTICE

Review of *Do High Flyers Maintain Their Altitude?*

Jaekyung Lee, University at Buffalo, SUNY

I. Introduction

This Fordham/North West Evaluation Association research report raises concerns about the performance trends of high-achieving students. The central question (and title of the report) is: *Do High Flyers Maintain Their Altitude?* or, in plain language, "Do high-achieving students maintain their high academic ranking?" The motivation for this study is stated in the foreword to the report:

> If America is to remain internationally competitive with other advanced nations, we need to maximize the potential of our top students. Yet many analysts worry that various policies and programs, including the federal No Child Left Behind Act (NCLB), tend to "level" student achievement by focusing on the lowest-achieving students and ignoring—or, worse, driving resources away from—our strongest students (p. 1).[1]

Specifically, the study examines reading and math achievement trends for students who scored extremely well on the Measures of Academic Progress (MAP). Although this study is unique in terms of its exclusive focus on high achievers at the student level (pitted against low achievers), the underlying framework differs little from many previous policy studies.[2] As illustrated by this reviewer in Figure 1 below, the report's classification and labeling of

The authors appear to switch from one metric to another throughout the report.

students is based on two variables: performance level (initial status) and improvement (growth). The initial status measure represents the baseline status of achievement—how well students perform in the first year their scores enter the database. The students are then tracked and classified based on "growth," specifically how much students improve their achievement over time. By cross-classifying students according to these two dimensions it is possible to examine how many high achievers versus low achievers are making more or less progress. Middle achievers can be added to the framework, but they

are not considered in this review, for the sake of sharpening the contrast between high and low achievers. Among initially high-performing students (dubbed "High Flyers" in the report), those students who fall within cell B in the table below are designated "Descenders." The study identifies those in cell A as "Steady High Flyers" whose performance remained consistently above the top 10-percent bar. This raises the question of how many of the students maintain or lose their academic edge over time (and why). The study also contrasts those in cell D, designated as "Never High Flyers" versus those in cell C, designated as "Late Bloomers."

Table 1. Classification of students by the level of performance (initial status) and improvement (growth)

		Performance (Initial Status)	
		High	Low
Improvement (Growth)	High	High on both Performance and Improvement (A)	High on Improvement and Low on Performance (C)
	Low	Low on Improvement and High on Performance (B)	Low on Both Performance and Improvement (D)

II. Findings and Conclusions of the Report

The report's key findings and implications are summarized as follows.

A majority of high flyers (nearly three in five such students) maintained their status over time, but substantial numbers were found to have "lost altitude." The report notes that there are real consequences for graduates who descend from the 90th to 70th percentile in terms of merit-based aid and choice of college. The findings show that most Descenders do not fall that far, however. Although the Descenders no longer performed at or above the 90th percentile, as they did in third grade, the vast majority remained above the 70th percentile in eighth grade. Additionally, the study ended up with more high achievers overall than it started with. The proportion of "Late Bloomers" surpassed that of the "Descenders." Most of the Late Bloomers were already above average at the starting point, with nearly all performing between the 50th and 89th percentile.

The report shows that high flyers grew academically at similar rates to low and middle achievers in math but grew at slightly slower rates than low and middle achievers in reading. In reading, high achievers grew about half as fast from third grade to eighth grade as low-achieving elementary/middle school students, reducing the gap between the two groups by more than a third. The report claims one of the factors possibly contributing to these results could be NCLB's focus on low-performing schools or Reading First's focus on struggling readers.

The report further claims that high-achieving students attending high-poverty schools made about the same amount of academic growth over time as their high-achieving peers in low-poverty schools. The authors interpret this finding as challenging the notion that wealthy suburban schools produce greater academic gains for students than their poorer counterparts. They speculate that growth over time for the highest-achieving students has little to do with the schools they attend and much to do with what's happening for them personally and at home.

III. Review of the Report's Methods, Plus an Alternative Analysis

The study used samples of public school students in grades 3 through 8 (Cohort 1) and in grades 6 through 10 (Cohort 2). Among these cohorts, high achievers were those students who performed at or above the 90th percentile, based on 2008 norms, on their third-grade (Cohort 1) or sixth-grade (Cohort 2) MAP tests. The methodology section in the appendix gives sample size but not information on the population or the sampling method.[3] The sample size is very large, but it is unclear whether it is a nationally representative random

There is little information in the report that can help differentiate between high-improving and low-improving students.

sample. It gives demographic breakdown by gender, race and poverty. Approximately 75% of the total sample was non-minority, so it seems that the sample included more Whites than what the nationally representative sample is expected to have (56% White as of fall 2007, according to the Digest of Education Statistics).[4]

The most critical aspects of this study concern how it operationally defined high achievers and how it tracked academic progress over time. These aspects are unclear and confusing because the authors appear to switch from one metric to another throughout the report. For example, the use of percentile ranks in each grade is based on normative comparisons and thus produces winners and losers. In contrast, it appears that the study also used developmental scale scores, which allow for continuing growth regardless of change in relative status; thus all students can be winners or losers. Why the authors chose particular metrics and also shifted between metrics is not explained.

There are several threats to the validity of how growth was measured in the study. To illustrate, to investigate these threats, and to check the robustness and validity of their findings, I decided to conduct a comparable study. Because the MAP dataset that the study used is not publicly available, this review examined similar data from the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K). My analysis, presented briefly below in the section titled "Review of the Validity of the Findings and Conclusions," was carried out using the Item Response Theory (IRT) scale scores from reading and math achievement tests, with a focus on fall kindergarten test scores (initial status) and K-to-8th-grade gain scores (growth).

One threat to validity involves so-called ceiling effects. For both the MAP study and the ECLS-K study, there may be ceiling effects, or limits to achievement gains, that could explain the observed decline in high-achieving students' trajectory. Although the report noted that measurement errors are relatively small for high achievers, it did not check and report how observed test scores were distributed (skewness) at the end of the achievement spectrum. This becomes important when studying changes among the highest achievers in a distribution.

A second threat concerns so-called regression to the mean.[5] Such regression artifacts complicate the evaluation of student progress. Regression to the mean occurs when we examine the difference between two imperfectly correlated measures. As one would expect, the size of correlations between repeated measures drops substantially as the time intervals grow farther apart in a longitudinal study. Lower-performing students tend to improve their performance status more than higher-performing students.

The correlations in ECLS-K turned out to vary substantially according to which of the two metrics used in the report was applied. With percentile ranks, the correlations were moderately negative (r = -.51 for reading and r = -.47 for math), meaning that the higher the initial status, the smaller the subsequent gains. In contrast, with IRT scale scores, the correlation between fall kindergarten scores and K-8 gain scores in ECLS-K were close to zero (r = -.004 for reading and r = .09 for math), meaning that students gain at essentially equal rates no matter their starting scores. This suggests that the regression-to-the-mean phenomenon could be a more serious problem when using percentile rank as the achievement outcome metric.

Given the threat posed by regression to the mean, the central question is the extent to which students' academic progress reflects real change as opposed to a statistical artifact. One way to check this is to conduct time-reversed analyses to determine whether the gain really is a regression artifact (see Campbell & Kenny, 1999).[6] Unfortunately, the information that the report provides under the title of "regression to the mean" in the Appendix was actually about the ceiling-effect problem.[7]

Beyond the regression phenomenon, there is little information in the report that can help differentiate between high-improving and low-improving students. Unless we understand the mechanism (such as school and teacher variables) that facilitates or constrains the different patterns of growth, a simple presentation of these differences is not informative. As mentioned above, the report includes a separate analysis that employed hierarchical linear modeling (HLM) to evaluate the results for high achievers in high- and low-poverty schools and found no systematic relationship between school poverty and improvement of high achievers. This is consistent with prior research.[8] While school-related effects may include a broad array of factors, it is also worth differentiating between classroom-level (teacher) and school-level effects to capture value-added contributions of teachers. The study's model, however, does not capture these effects at all, so all between-classroom variance is hidden in the within-school variance. Moreover, the relative comparison of growth among schools does not capture the absolute amount of school or teacher contributions to students' academic growth; consider the well-known phenomenon of

summer learning loss for disadvantaged students. Simple comparison of relative achievement gains can lead to the false impression that academic growth is primarily an individual student matter and that schools hardly make a difference.

IV. Review of the Literature

The report provides a sidebar (p. 6) in which the authors cite a couple of selected studies and point out that the body of research regarding the academic performance of high achievers and the impact of accountability policy on that group is relatively limited. However, the report could have acknowledged some comprehensive reviews of related studies that give broader views and raise concerns about equity as well as excellence. For example, the National Research Council has issued reports including NRC (2002) on the issue of minority representation in gifted education and NRC (2011) on the issue of high-stakes accountability policy impact and risks; another important study is Ceci & Papierno (2005), on the issue of who benefits more from universal versus targeted interventions. [9] The question of excellence in education is not restricted to high-performing students, and a helpful literature review would need to address both excellence and equity in a balanced manner.

V. Review of the Validity of the Findings and Conclusions

Given the lack of sufficient technical information in the report, it is difficult to assess all of the aforementioned threats to validity. To cast light on the importance of these issues, I therefore present here the findings of my analysis using ECLS-K data.[10] These analyses and findings illustrate potential threats to validity and demonstrate the variations in results that can occur when using different measures and methods.

Scale Scores versus Percentile Ranks

Figures 2 and 3 show K-8 reading and math achievement trajectories of students in ECLS-K. The results show a sharp difference between using scale scores (top panel) and percentile ranks (bottom panel).[11] Unlike the percentile ranks (bottom panel) that show decline among high achievers, scale scores (top panel) show continuing growth among high achievers. Moreover, high achievers end up meeting the national standard (NAEP proficient level of achievement) in both subjects before they reach grade 8, whereas low achievers remain off track.[12]

This contrast between scale scores and percentile ranks shows how researchers can come to totally different conclusions simply based on the choice of metrics. If we use the percentile rank, low achievers bloom but high achievers wither. But this impression is invalid due to the regression-to-the-mean problem, which does not pose a threat when the metric used is scale scores because the developmental scale helps measure the time-varying amount of growth in individual achievement across grades. ECLS-K scale scores are also relatively more immune from ceiling effect problems by design.[13]

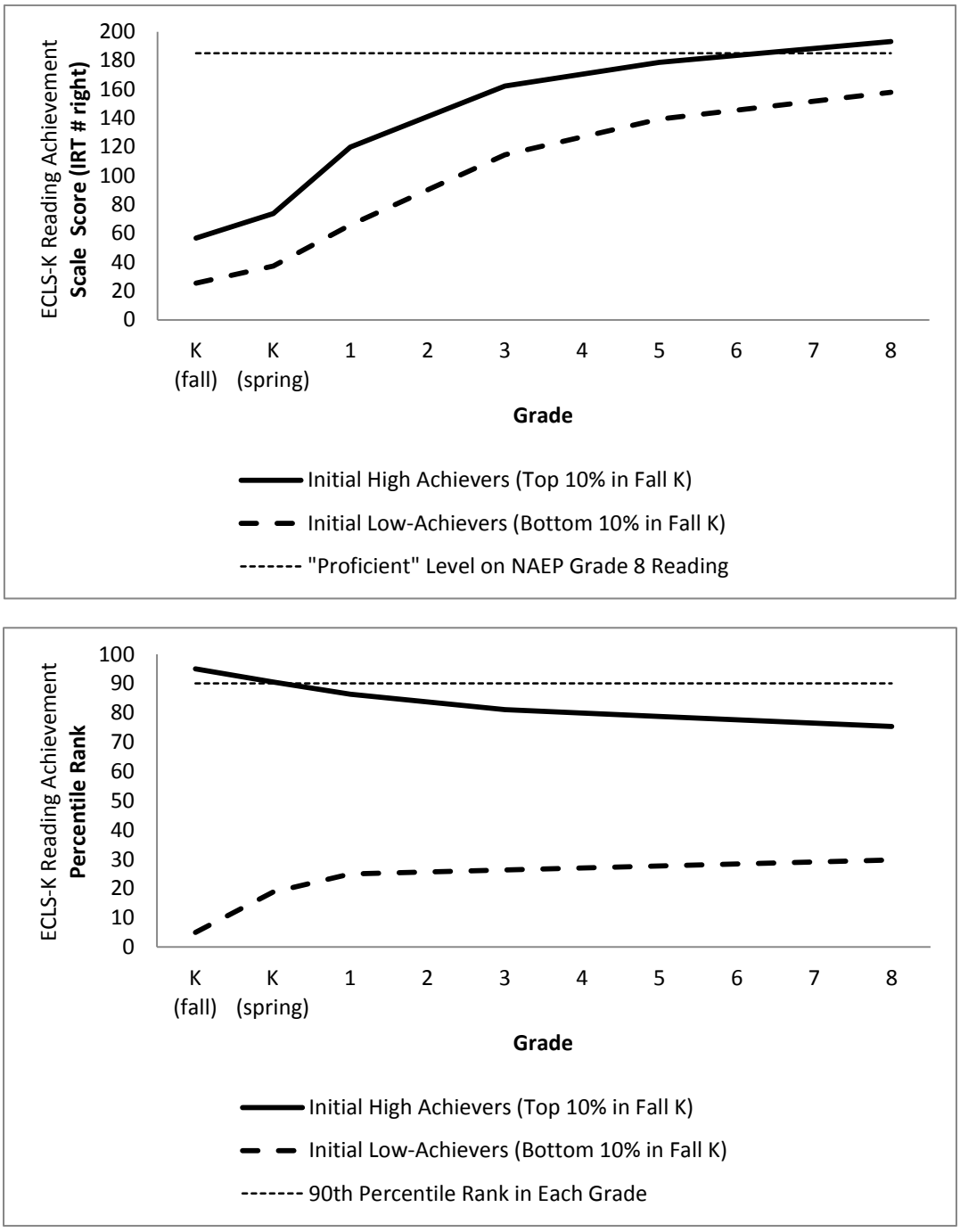


Figure 2. High vs. low-achieving Kindergarteners' reading achievement growth trajectories during K-8 period, using the metric of IRT scale score (top panel) and percentile rank (bottom panel)

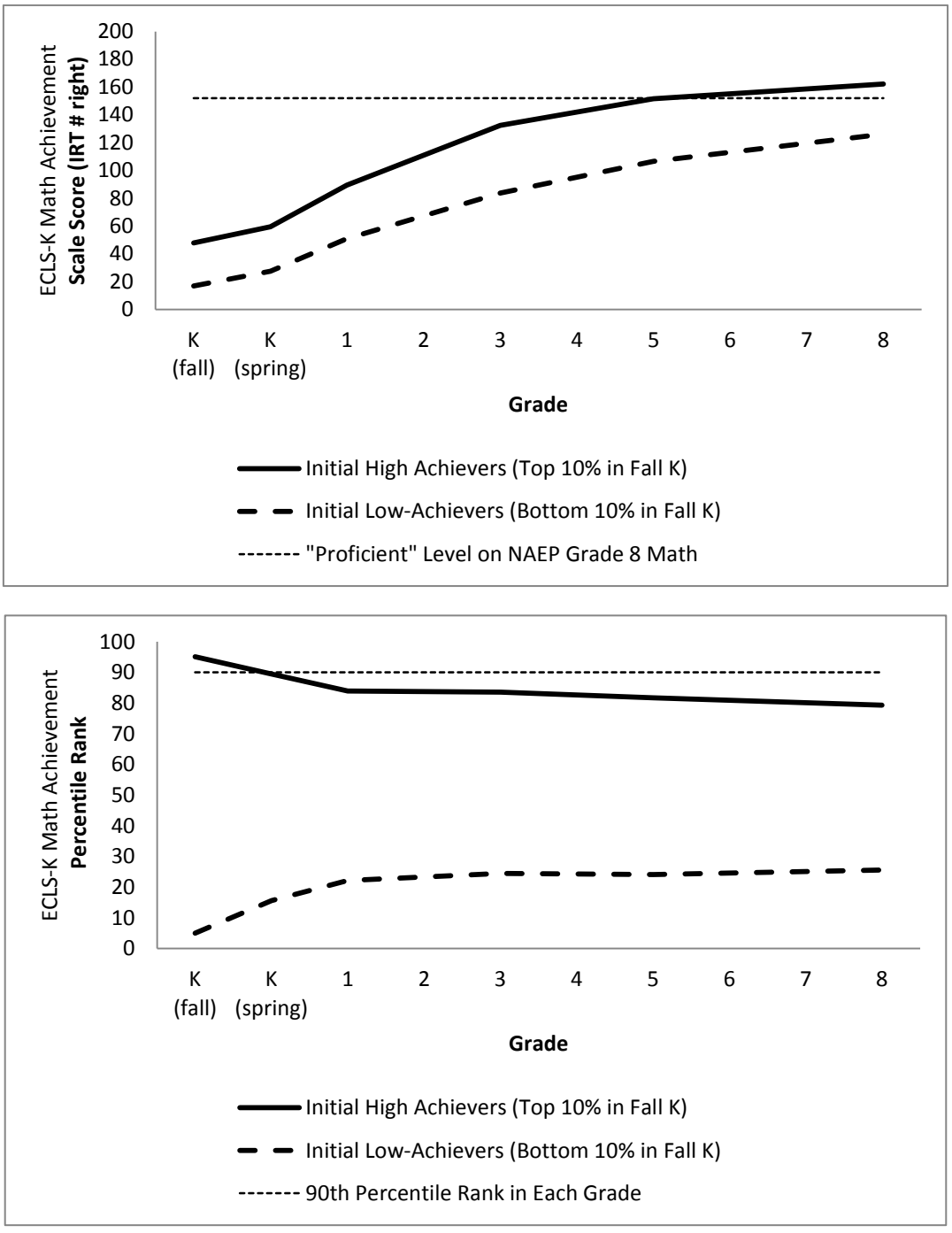


Figure 3. High vs. low-achieving Kindergarteners' math achievement growth trajectories during K-8 period, using the metric of IRT scale score (top panel) and percentile rank (bottom panel)

How does a researcher determine the size of the regression-to-the-mean effect and adjust for it? Time-reversed analysis can help detect the bias. In this study, for example, one can flip the question: how well did high achievers in the eighth grade perform back then when they were in kindergarten (tracking performance backward from a higher grade to a lower

grade)? The results of time-reversed analysis show that the high achievers based on grade 8 percentile rank performed less well in kindergarten; this pattern is the mirror image of what we saw in Figure 3, using a forward tracking method. In order to remove this regression artifact from gain scores, one can use the initial test score as a control variable in a regression model, which allows the researcher to produce residualized gains.

For the new results presented here, I therefore used the analysis of covariance (ANCOVA) method to control for initial status and compared adjusted gain scores with raw gain scores; the question is how much the change would have occurred if everyone had started at the same point. Before presenting the results, I should note that, as Cronbach and Furby (1970) point out, while residualized gain is a way of removing the effect of pretest status from a posttest score, it is not a corrected measure of true change because the portion discarded may include some genuine and important changes in the subjects.[14]

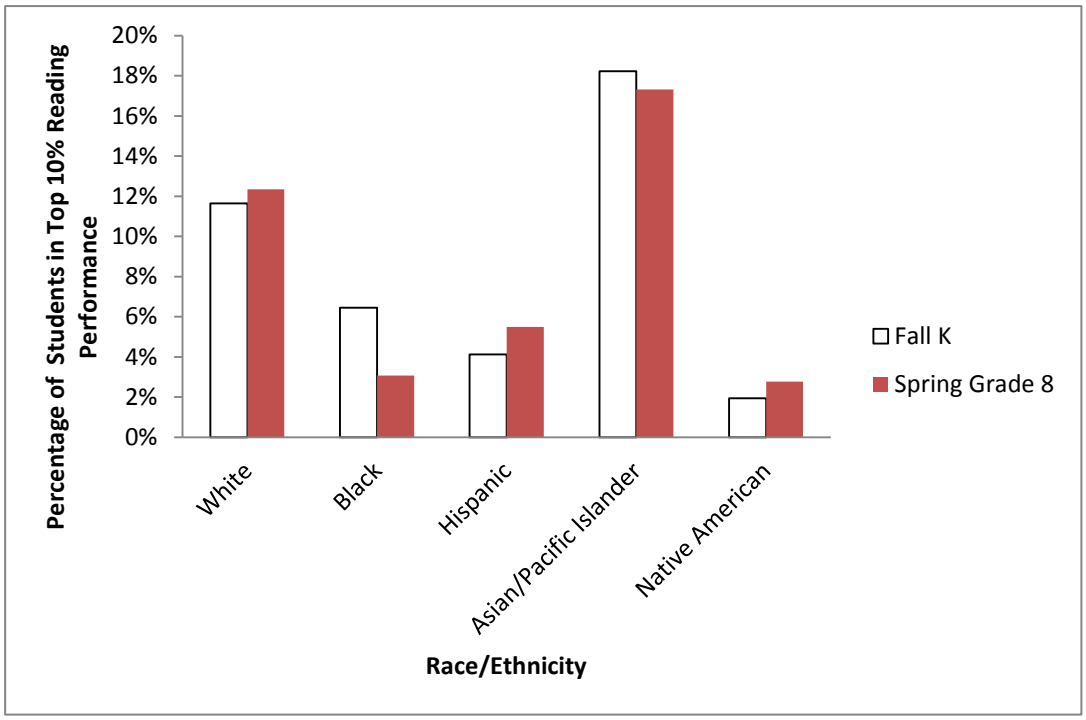
Before adjustment, K-8 reading and math percentile gains for initially high achievers were large and negative (-20 points in reading and -16 points in math), as shown in Figures 2 and 3. After adjustment for the effect of initial status, the gains for initially high achievers

The regression-to-the-mean tendency may have created an illusion of over-progress in percentile ranks for low achievers and over-decline for high achievers.

became positive or close to zero (+ 9 points in reading and -1 point in math). Exactly opposite patterns occur among initially low achievers: Their unadjusted gains were highly positive (+25 points in reading and +21 points in math), whereas adjusted gains turned negative or much smaller positive (-4 points in reading and +6 point in math). The point here is not that this adjusted gain is more valid than the original measure of gain, but that the measure of change using percentile rank could be biased in one way or the other.

This suggests that the regression-to-the-mean tendency may have created an illusion of over-progress in percentile ranks for low achievers and over-decline for high achievers. Alternatively, when we use the IRT scale score, this problem is less likely to be an issue, since the developmental scale helps measure the varying amount of growth in individual achievement across grades. Specifically, the top panels in Figures 2 and 3 show that (a) initially high and low achievers make equal academic progress in reading and math through the K-8 schooling period; (b) the achievement gaps between high and low achievers persist through the period, and low achievers perform well below the proficient level.

What is particularly serious is the presence of gaps among racial and poverty groups. Figure 4 shows underrepresentation of certain minority and high-poverty groups in the high-achiever category consistently over time. This persistence of the gap problem is not adequately reported and acknowledged in the report.



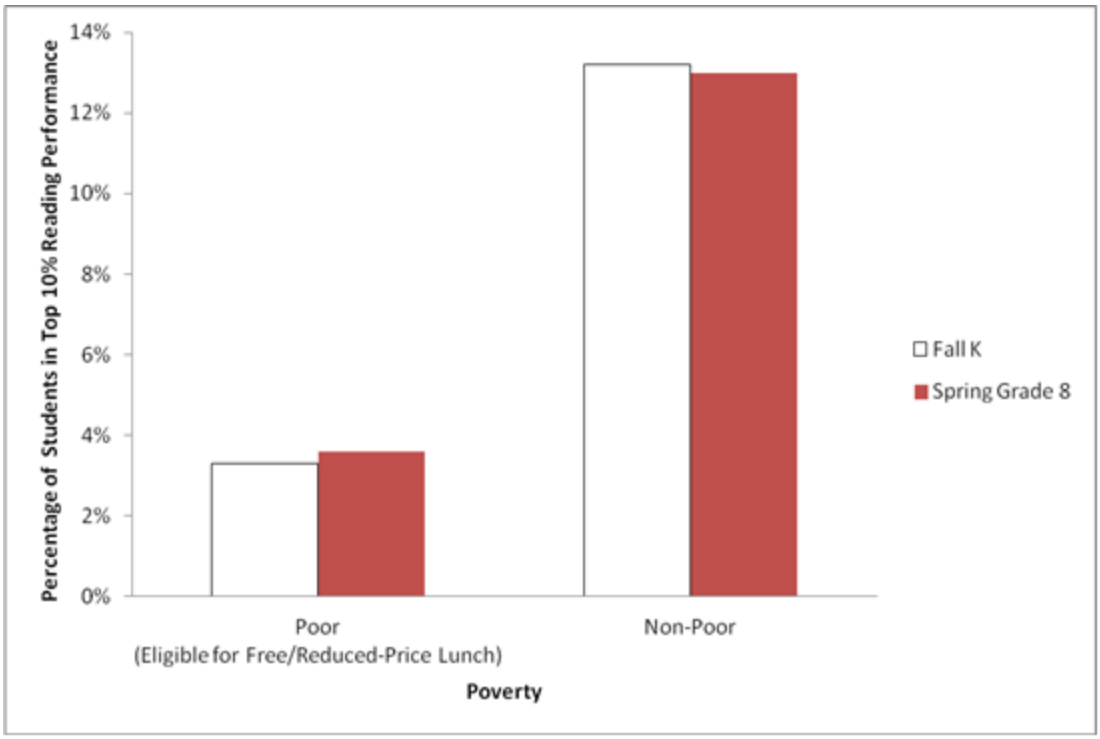


Figure 4. Percentages of students in high reading achievement (top 10%) in Kindergarten and grade 8, by race/ethnicty (top panel) and poverty (bottom panel)[15]

Lastly, one of the key questions is why this uneven progress, beyond regression to the mean, may happen among top-achieving students. How much of the change is attributable

to teacher or school effects? Answers to these questions may have important policy implications. The percent of the difference occurring at the teacher or school levels may give an upper-bound estimate of possible teacher effects or school effects. My analysis of ECLS-K data shows that some variations in student progress are indeed attributable to teacher and school effects beyond students' initial test scores or demographic background characteristics: teacher-level effects and school-level effects account for about 10 percent of the total variance each (after controlling for students' initial test scores and demographics) and the effect is not negligible.[16]

In conclusion, the report makes the following policy recommendation, which implies that the current test-driven accountability system is working for low achievers but at the expense of high achievers:

> If we are truly serious about providing excellence in education for all students, then we should consider changing accountability systems to place emphasis on the growth of low-, middle-, and high-achieving students alike. Our results suggest that this type of accountability would subject some wealthy, underperforming suburban schools to fair and welcome scrutiny.

This logic is flawed for two reasons. First, there is no scientific evidence that high-stakes accountability policies such as NCLB are working consistently well.[17] The cited finding, showing low achievers improving relatively more than high achievers on NAEP over the last decade, does not support causal attribution of the effect to NCLB. Further, a great need continues to exist for a continuation of the federal role over the past half-century, focusing attention on struggling students in high-poverty schools. Although the current federal policy has problems with its approach, this does not deny the continued importance of the federal role in targeting disadvantaged low-achieving students and their schools. Emphasis on growth is needed for all students, but more so for low-achieving students, due to the large opportunity gaps and to these students' serious underperformance against national standards.

VI. Usefulness of the Report for Guidance of Policy and Practice

This study, which tracked longitudinal academic growth of elementary and middle school cohorts, may help bring more attention to the issues of high-achieving students in accountability designs. However, the report's flawed analysis and interpretation leads to biased results and to an unsupported conclusion that many high-performing students do not maintain their academic edge while more low-performing students catch up. The reported decline in high-performing students' achievement is likely an artifact of the measure being used and of the regression to the mean phenomenon. As I demonstrated through analysis of similar national data, the results are very sensitive to the choice of measures and to analytical methods. Thus, the report's arguments about the loss of potential human capital and about a purported trade-off between excellence and equity can be more harmful than helpful.

"Do High Flyers Maintain Their Altitude?" The answer to this question depends on how researchers define and measure altitude and progress. The report's norm-referenced classification method guarantees winners and losers independent of their true performance and improvement levels. There will always be winners and losers when the calculation is based on comparisons between students' relative performance rather than against absolute benchmarks. Further, this norm-referenced approach may pose potential conflicts with the criterion-referenced approach required by law. My alternative analysis of performance trends added such a criterion-referenced perspective, and it shows that initially high-flying students continue to meet the NAEP standard of proficiency, while initially low flyers almost never reach such a high goal. So the good news or bad news, depending on one's predilection, is that everybody improves to more or less the same extent over time, implying equal benefits of schooling. However, if we are concerned about the issue of equity, the picture looks gloomy. The clear bad news is that the achievement gap between high and low achievers is large and does not narrow over time in general. And more specifically, racial and poverty gaps also do not narrow (and sometimes widen) over time.

The utility of this report is further limited by its black-box approach that assumes a link between its findings and NCLB-related policies. Even assuming the validity of the report's findings, such causal assumptions are problematic given the study's failure to examine specific teacher or school characteristics associated with differences between low-improving versus high-improving students that had high initial performance status. To investigate questions of this nature, a study will have to begin with valid and reliable measures of differences in academic growth and will also have to include measures of school and teacher factors that may have caused these differences. Using the framework in Figure 1,what the study does not address is how we can help students move from cell B to cell A and from cell D to cell C. The nation's education can become more excellent and equal, not simply by sorting, labeling, and tracking students by initial test scores, but by investing more in high-quality educational practices for all students.

Notes and References

1 Xiang, Y., Dahlin, M., Cronin, J., Theaker, R. & Durant, S. (September, 2011). *Do High Flyers Maintain Their Altitude?* The Thomas B. Fordham Institute/Northwest Evaluation Association. Retrieved October 11, 2011, from http://www.edexcellence.net/publications-issues/publications/high-flyers.html.

2 For example, see Ushomirsky, N. & Hall, D. (2010). *Stuck Schools*. Washington, DC: The Education Trust. The Education Trust report focuses on school-level performance trends and focuses on low-performing and low-improving schools. The unique aspects of the Fordham report reviewed here are that the unit of analysis is students rather than schools, and the focus is on high achievers rather than low achievers.

3 In an email to Professor Kevin Welner, Director of the NEPC, Fordham's Michael Petrilli stated that the report's authors were working on a technical report. Welner contacted John Cronin to ask when the technical report would be available, but he did not receive a response.

4 See National Center for Education Statistics (2009). *The digest of education statistics 2009*. Retrieved on September 31, 2011, from http://nces.ed.gov.

5 In blog entries subsequent to the report's publication, one or more of the report's authors attempt to explain and defend their approach to the regression to the mean concerns. This review of the report does not also take on a review of those blog entries, but nothing in those entries alleviates the concerns described here about these threats to the validity of the report's findings.

6 Time-reversed analysis refers to the analysis of data by switching the flow of time and determining if the results change.

See Campbell, D. T. & Kenny, D. A. (1999). *A Primer on Regression Artifacts*. New York: Guilford Press.

7 This text in the report's appendix reads as follows:

> *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.

8 See Raudenbush, S. W. (2004). *Schooling, Statistics, and Poverty: Can We Measure School Improvement?* Angoff Lecture No. 9. Princeton, NJ: ETS.

9 National Research Council (2002). *Minority Students in Special and Gifted Education*. Committee on Minority Representation in Special Education, M. S. Donovan and C. T. Cross (Editors). Washington, DC: National Academy Press.

National Research Council (2011). *Incentives and Test-Based Accountability in Public Education*. Committee on Incentives and Test-Based Accountability in Public Education, M. Hout and S. W. Elliott (Editors). Washington, DC: National Academies Press.

Ceci, S. J., Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "have nots" gain but the "haves" gain even more. *American Psychologist, 60*, 149-160.

10 For detailed information on ECLS-K data sampling and measurement, see the following NCES reports (available at www.nces.ed.gov/ecls):

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., and Najarian, M. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99* (ECLS-K), *Combined User's Manual for the ECLS-K Eighth-Grade and K–8 Full Sample Data Files and Electronic Codebooks* (NCES 2009–004). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Najarian, M. Pollack, J.M., & Sorongon, A.G. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for the Eighth Grade* (NCES 2009–002). Washington, DC: National Center for Education Statistics.

Pollack, J., Atkins-Burnett, S., Rock, D. and Weiss, M. (2005). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Psychometric Report for the Third Grade* (NCES 2005–062). Washington, DC: National Center for Education Statistics

11 The IRT-estimated number-right scores for reading and mathematics are estimates of the number of items students would have answered correctly had they taken the entire exam and responded to all items in the test item pool.

12 The NAEP standard is used here instead of the state standard, because state standards tend to vary dramatically. Based on linear linking between NAEP and ECLS-K scales, NAEP 8th grade reading and math achievement scores that correspond to "Proficient" levels (cut score of 281 in reading and cut score of 299 in math) are equivalent to ECLS-K 8th grade scale scores of 185 in reading and 152 in math respectively. For the NAEP-ECLS linkage, the equating samples were composed of the main NAEP (2007) national sample of 8th graders and the ECLS-K 2007 8th grade cohort members. The following linear conversion formula was used for converting scores on test *X* (NAEP) to the scale of test *Y* (ECLS-K):

$$y = \frac{\sigma(Y)}{\sigma(X)}x + \left[\mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\mu(X)\right]$$

13 For ECLS-K, special adaptive testing procedures were designed to minimize both floor and ceiling effects that typically distort gain scores. The analysis of ECLS-K scale scores reveals slightly negative skewness (e.g., skewness = -.30 for reading and -.25 for math in ECLS-K grade 3), but a skewness value between ± 1 is considered acceptable for most psychometric purposes.

14 Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74,* 68-80.

15 The second panel of this figure has been corrected from a previously posted version.

16 Consistent with previous research, there are vast individual differences in student learning rates, and classroom or school variance components are small in comparison to student variance. Yet both teacher and school effects are important and significant.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*, 237–257.

Scheerens, J. & Bosker, R. J. (1997). *The Foundations of Educational Effectiveness*. Oxford, UK: Pergamon.

Palardy, G., Rumberger, R. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*(2), 111-140.

17 Hout, M. & Elliott, S. W. (Eds.) (2011). *Incentives and Test-Based Accountability in Public Education*. Committee on Incentives and Test-Based Accountability in Public Education; National Research Council. Washington, DC: The National Academies Press.

DOCUMENT REVIEWED:	**Do High Flyers Maintain Their Altitude?**
AUTHORS:	Yun Xiang, Michael Dahlin, John Cronin, Robert Theaker, & Sarah Durant
PUBLISHERS/THINK TANKS:	The Thomas B. Fordham Institute/Northwest Evaluation AssociationPublisher
DOCUMENT RELEASE DATE:	September 2011
REVIEW DATE:	October 13, 2011
REVIEWER:	Jaekyung Lee, University at Buffalo, SUNY
E-MAIL ADDRESS:	JL224@buffalo.edu
PHONE NUMBER:	716-645-1132

SUGGESTED CITATION:

Lee, J. (2011). *Review of "Do High Flyers Maintain Their Altitude?"* Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/thinktank/review-high-flyers.