

Tom Loveless's Response to Gregory Camilli's Review of "High Achieving Students in the Era of NCLB."

I thank Greg Camilli for his review of "High Achieving Students in the Era of NCLB" (View his full review [here](#)). The review contains many insightful comments and areas on which we agree. The review also contains numerous factual errors and misstatements. I will confine most of my comments to Camilli's review of Part 1 of the report, an analysis that I authored of National Assessment of Educational Progress (NAEP) data. Part 2 of the report, a survey of teachers, was authored by Steve Farkas and Ann Duffett. A foreword to the two studies was authored by Chester E. Finn, Jr. and Michael J. Petrilli.

I respond to four topics: purpose of the study, literature review, NAEP data, and policy recommendations.

Purpose of the Study

Camilli claims that an economic rationale for educating gifted children permeates the report. He first quotes from the forward by Finn and Petrilli, who argue that giving high achievers a good education is essential if the United States is to remain competitive in the world. This argument sounds reasonable to me, but it is deduced from—and not a description of—the findings of the studies. Camilli attempts to show that the "economic rationale" runs through the report.

To document, as Camilli puts it, that the "economic concern is echoed in Part 1,"¹ he cites a quotation by Susan Goodkin: "By forcing schools to focus their time and funding almost entirely on bringing low proficiency students up to proficiency, NCLB sacrifices the education of the gifted students who will become our future biomedical researchers, computer engineers, and other scientific leaders."²

One problem. The Goodkin quotation was not endorsed. Nor was it provided to support an "economic rationale." It was provided as an example of the argument that NCLB creates a Robin Hood effect—that incentives focusing on low achievers take away from high achievers. In the original report, the sentences immediately following the Goodkin passage declare the agnostic position taken by the study: "Are these concerns well founded? Do the incentives of NCLB create a Robin Hood effect, yielding gains for low achieving students but at the expense of high achievers? That's what we set out to investigate."³

What makes taking this quote out of context particularly egregious is that the study shows that a Robin Hood effect is not confirmed by NAEP data. Thus, Camilli uses a quotation that is largely refuted by our study to explain to readers what our study is about. He fails to quote sentences that declare precisely what the study is about. Having built a straw man, Camilli spends a lengthy paragraph in his review tearing it down, showing that many factors go into national economic prosperity. The exercise is not only banal; it is irrelevant.

Literature Review

Camilli criticizes the literature review for not including all of the research on accountability. That would be an apt criticism if the study sought to analyze the comprehensive effects of

NCLB and accountability systems. But it didn't. As just noted, this study was focused on a narrower topic, whether there is evidence of a Robin Hood effect under NCLB. The study therefore searched for literature on that topic, specifically, whether "NCLB-style" accountability systems foster educational triage—the shifting of resources towards low achieving students in response to incentives in accountability systems.

The search produced three high quality studies investigating that question. Camilli does not discuss any of them. He instead calls the review "cursory" and says "the three studies that are discussed are provided to show mixed evidence of achievement growth rather than to provide a balanced survey of the research." This is strange wording indeed. It conveys the impression that instead of a literature search, three studies were selected to conform to a pre-conceived idea of what the literature says. No, three high quality studies were found on triage, and they offer mixed findings on whether triage occurs under accountability systems. Camilli cites no additional studies on triage that should be included nor explains how he would resolve the mixed findings.

What would constitute a "balanced" survey of the literature? Camilli claims that the paper should have acknowledged the "national accountability research" and cites Amrein and Berliner (2002). That study examines the effects of states adopting high stakes, high school exit exams. Camilli does not explain how a study of high school exit exams would inform our study. The Amrein and Berliner study does not investigate triage. It does not discuss effects on high achievers. High school exit exams are not part of NCLB. High school students are not the subjects of our study.⁴

Several curious problems arise as Camilli's critique discusses a study by Carnoy and Loeb.⁵ First he gets the publication date wrong. The study was published in 2002, not 2003. A trivial typo perhaps, but an ironic error in a discussion of proper scholarship. More importantly, he confuses how the Carnoy and Loeb study was used. It was used to code whether states had accountability systems, not to support a particular interpretation of NAEP trends (see page 24, High Achieving Students).

Camilli creates the false impression of a conflict, that Carnoy and Loeb conducted a later study "in a more comprehensive manner"⁶ with findings that refute the NAEP trends reported in the paper. The study is not cited. His documentation for this assertion is a second-hand account, a reporter's description of an unpublished study's findings (see note 8, Camilli's Review). Again, an odd modeling of rigorous scholarship. The quote from the reporter's story, that gains in math scores from the late 1990s "tapered off from 2000 to 2003" refutes absolutely nothing stated in our study⁷.

In sum, in his critique of our study's literature review, Camilli does not discuss the literature that was actually reviewed, gets a publication date wrong, incorrectly depicts how one study was used, and asserts that the review was not "balanced" without either documenting the assertion or defining what "balanced" means. In making suggestions on how to improve the literature review, he cites a second-hand account of an unpublished study and argues that a study of a policy irrelevant to our study's research questions should have been included.

The NAEP Data

The discussion of NAEP data is confused. Camilli repeatedly asserts that the findings concerning the 90th-10th percentile gaps are based on state NAEP data (see page 6, Camilli Review). That is not true. The main NAEP is administered to national and state samples of students. The findings of the study come from both sets of data (see Figures 2a-2d on page 21, High Achieving Students, for national scores).

The NAEP also administers another test, the long term trend. Our study did not use data from the long term trend. Camilli argues that the long term trend casts doubt on the findings: “the state and long term data sets give different results, and this problem needs to be addressed before gap statistics can be used confidently to describe the effects of accountability policies.”⁸ The statement is awkward. I think Camilli means data from the main NAEP, not “state” data. Moreover, readers of our study are cautioned more than once about the limitations of cross-sectional achievement data. No NAEP data can be used “confidently to describe the effects” of any policy.

Which of the two NAEP tests—main or long term trend—is more appropriate for a study of high achievers in the NCLB era? The long term trend NAEP is administered to an age-based sample and has only been given once since 2000—in 2004. The main NAEP was administered five times from 2000-2007 in 4th grade reading and four times in 4th grade math, 8th grade reading, and 8th grade math. Only national scores are available on the long term trend, no state level scores. In addition, the main NAEP was launched to reflect current curricula and so is probably more in synch with the assessments that states use in accountability systems. To examine how test scores behave during accountability regimes, the main NAEP is more appropriate than the long term trend.

Another benefit of using the main NAEP is that the state and national assessments produce independent sets of data. The same test is given in the same grades, yet results are derived from different samples of students, often collected in different years. These two data sets support the main findings of the study:

- 1) Scores on the National Assessment of Educational Progress (NAEP) rose for both 90th percentile and 10th percentile students from 2000-2007.
- 2) During this period, more growth was recorded by 10th percentile than 90th percentile students, leading to a narrowing of the gaps between them.
- 3) The gap narrowing was not evident in NAEP scores during the 1990s. The 1990's present a mixed picture, with one important exception: from 1996-2002, states with accountability systems show more improvement at the 10th percentile than the 90th percentile.

Camilli presents the following table of scores from the mathematics test at age 9 on the long term trend.

Table 1: Gap between low- and high-achievers in NAEP mathematics data (age 9)⁹

| Year | 10th % | 90th % | Gap |
|------|--------|--------|-----|
| 1978 | 171 | 264 | 93 |
| 1982 | 173 | 263 | 90 |
| 1986 | 177 | 264 | 87 |
| 1990 | 186 | 271 | 85 |
| 1992 | 185 | 271 | 86 |
| 1994 | 187 | 272 | 85 |
| 1996 | 187 | 274 | 87 |
| 1999 | 187 | 275 | 88 |
| 2004 | 197 | 282 | 85 |

Camilli concludes from this table: “Here, there is no evidence of a shrinking gap at age 9 in mathematics, except, perhaps, from 1978-1986, well before NCLB. There has been more growth at the 10th than at the 90th percentile since 1978, but most of this differential growth occurred prior to 1990. For other age-subject combinations in the long-term data, evidence of a shrinking gap consistent with accountability influences is similarly weak.”¹⁰

Let us examine these data in light of the three findings of our study. Consider our study’s conclusion that the 1990’s national data “present a mixed picture.” Nothing in this table contradicts that finding. The data show the gap in math among nine year olds expanding by 3 scale score points in the 1990s. If Camilli had also shown reading scores for nine year olds in the 1990s, we would see the gap contracting by 15 points. A mixed picture.

Consider the finding that state NAEP data show narrower gaps favored from 1996-2002 in states that had adopted accountability systems compared to states that had no such accountability systems. I say “narrower gaps favored” instead of “narrowing gaps” because in one grade-subject combination, 8th grade math, the gap expanded in both groups of states but expanded less in accountability states. The data Camilli presents are national scores and cannot confirm nor reject any trends detected by comparing states with different policies. Only 16 states had accountability systems at the time so comparing their test scores to those of non-accountability states is more informative than merely examining national scores.

Our third finding is that the gap shrank from 2000 to 2007, the NCLB era, on both the state and national main NAEPs. This table shows that the achievement gap contracted by 3 scale score points from 1999 to 2004 on the long term trend. That time period only overlaps with a portion of 2000 to 2007. Nevertheless, what Camilli’s data show—the gap shrinking—is in line with our findings. Indeed, if Camilli had shown all four age-subject combinations from the long term trend, readers would see the gap shrinking in three out of the four combinations from 1999-2004.¹¹

Camilli calculates effect sizes for the long term trend and the main NAEP. He says the latter statistic is an effect size for the “state” NAEP but, again, this is surely wrong. To make sense of the following passage, “4th NAEP” refers to 4th grade NAEP and instead of “NAEP state data” substitute “national data from the main NAEP:”

For example, consider 4th NAEP mathematics. Using the NAEP state data, the gap shrinks from 2000 to 2005 by 8 points, an effect size of about .25. For the long-term NAEP data, however, the 1999-2004 gap at age 9 (about 4th grade) decreases by only 3 points, an effect size of less than .10.¹²

On one level this is a trite point: that a different test with a different sample may detect a different magnitude of gap narrowing over a different time period. But Camilli is on to something important. The two NAEPs may indeed offer different conclusions about gap narrowing since 2000. They have before. In a previous study, I contrasted differences in the performance of high and low achievers on the long term trend and main NAEP tests in the 1990s. The main NAEP showed a widening gap in 4th grade reading; the long term trend showed a shrinking gap in reading scores for 9 year olds.¹³

But it is premature to decide that gap trends for the NCLB era differ between the two NAEP tests. Only one long term trend test has been given since 1999. The long term trend is being given in 2008 with data scheduled for release in 2009. Camilli and I are in agreement that comparing the two NAEP tests may shed light on trends in achievement gaps, but that endeavor needs to wait for more data.

Policy Recommendations

Camilli mentions three policy recommendations from the study: 1) web-based courses for high achieving students who attend schools that cannot offer advanced math courses, 2) an experiment featuring incentives for schools to boost the achievement of high achievers, and 3) tracking students into homogenous ability groups. He does not have much to say about the first two recommendations and opposes the third, to which he devotes a great deal of attention refuting.

Here is something strange for a review: the third recommendation was not made. As Camilli admits, “To be sure, tracking isn’t explicitly recommended. However, Loveless has recently authored a work favorable to tracking practices...” Camilli cites my 1999 book, *The Tracking Wars*. But it didn’t recommend tracking either. In short, Camilli criticizes a recommendation that I did not make. Then to show that I might be capable of making such a recommendation, he cites a book in which I did not make the recommendation either.¹⁴

Conclusion

Camilli praises the study for its use of student-level NAEP data to describe high achieving students, and in particular, high achievers who are students of color or who live in poverty. On the recommendation for an experiment using incentives targeting high achievers, perhaps Camilli and I can also find common ground. Camilli notes, quite accurately, that the data in the study show correlation, not causation, and declares, “To formulate evidence-based policies, however, the same degree of rigor is required as one would use in estimating a causal effect in a randomized experiment.”¹⁵

Yes, indeed, and a pattern of correlations across multiple data sets can be valuable in generating hypotheses for such an experiment. Our study uncovers just such a pattern. It uses the best data available, reviews relevant literature, and makes recommendations appropriate to the findings. I thank Greg Camilli for his review and urge readers to read the original study in its entirety.

¹ Gregory Camilli, *Review of High Achieving Students in the Era of NCLB*, Great Lakes Center for Education Research and Practice, 2008, page 2.

² Susan Goodkin, "Leave No Gifted Child Behind," *Washington Post*, December 27, 2005, A25.

³ Loveless, T., Farkas, S., and Duffett, A. (2008). *High Achieving Students in the Era of NCLB*. (Washington, DC: Thomas B. Fordham Institute). Page 14.

⁴ I review the research on accountability, including two studies by Amrein and Berliner, in "Test Based Accountability: The Promise and the Perils," *Brookings Papers on Education Policy, 2005* (edited by Diane Ravitch, Brookings Institution Press, 2005). Pages 7-45.

⁵ Martin Carnoy and Susanna Loeb, "Does External Accountability Affect Student Outcomes? A Cross-State Analysis," *Educational Evaluation and Policy Analysis*, vol. 24, no 4 (2002): 305-331.

⁶ Camilli, page 4.

⁷ Viadero, D. (2005). "Second Look at Tougher Accountability Yields New Results," *Education Week* vol. 24, issue 36, page 12 (May 11, 2005). Bethesda, MD: Editorial Projects in Education, Inc.

⁸ Camilli, page 6.

⁹ Camilli does not provide a source for the data in Table 1. They can be found in Perie, Moran, and Lutkus (2005).

¹⁰ Camilli, page 6.

¹¹ See Perie, M., Moran, R., and Lutkus, A.D. (2005). *NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics* (NCES 2005-464). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.

¹² Camilli, page 6.

¹³ Loveless, T. (2001). *The 2001 Brown Center Report on American Education*, (Washington, DC: Brookings Institution Press), pp. 12-15.

¹⁴ Loveless, T. (1999). *The Tracking Wars: State Reform Meets School Policy*. (Washington, DC: Brookings Institution Press).

¹⁵ Camilli, page 6.