



## REVIEW OF *MEASURING THE IMPACTS OF TEACHERS*

*Reviewed By*

Moshe Adler

Columbia University and Empire State College, SUNY

April 2014

### Summary of Review

Can the quality of teachers be measured the way that a person's weight or height is measured? Some economists have tried, but the "value-added" they have attempted to measure has proven elusive. The results have not been consistent over tests or over time. Nevertheless, a two-part report by Raj Chetty and his colleagues claims that higher value-added scores for teachers lead to greater economic success for their students later in life. This review of the methods of Chetty *et al.* focuses on their most important result: that teacher value-added affects income in adulthood. Five key problems with the research emerge. First, their own results show that the calculation of teacher value-added is unreliable. Second, their own research also generated a result that contradicts their main claim—but the report pushed that inconvenient result aside. Third, the trumpeted result is based on an erroneous calculation. Fourth, the report incorrectly assumes that the (miscalculated) result holds across students' lifetimes despite the authors' own research indicating otherwise. Fifth, the report cites studies as support for the authors' methodology, even though they don't provide that support. Despite widespread references to this study in policy circles, the shortcomings and shaky extrapolations make this report misleading and unreliable for determining educational policy.

**Kevin Welner**

*Project Director*

**William Mathis**

*Managing Director*

**Erik Gunn**

*Managing Editor*

---

**National Education Policy Center**

School of Education, University of Colorado  
Boulder, CO 80309-0249  
Telephone: (802) 383-0058

Email: [NEPC@colorado.edu](mailto:NEPC@colorado.edu)  
<http://nepc.colorado.edu>

Publishing Director: Alex Molnar



This is one of a series of Think Twice think tank reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at <http://greatlakescenter.org>.

*This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at [nepc@colorado.edu](mailto:nepc@colorado.edu).*

# REVIEW OF *MEASURING THE IMPACTS OF TEACHERS*

*Moshe Adler, Columbia University and Empire State College, SUNY*

---

## I. Introduction<sup>1</sup>

The two-part National Bureau of Economic Research report, *Measuring the Impacts of Teachers*, by authors Raj Chetty, John N. Friedman and Jonah E. Rockoff, has been used by advocates as evidence that measuring teacher quality through the use of standardized test score gains (value-added) is worthwhile because students of teachers with high “quality” scores have substantially better economic success later in life as a result.<sup>2</sup>

Even though the two parts of the report have not been published in a peer-reviewed journal, they have been extraordinarily influential. A previous version of the report (that version had only one part) was covered on the front page of *The New York Times*, and President Obama relied on it in his 2012 State of the Union Address when he asserted “a great teacher can offer an escape from poverty to the child who dreams beyond his circumstance.”<sup>3</sup> The paper became a major bone of contention during the 2012 Chicago teachers’ strike. In New Zealand, the Treasury Department posted the paper on its website.<sup>4</sup> One of the authors, Raj Chetty, has submitted an expert opinion in a current court case that involves a dispute over the teachers’ contract in Los Angeles; his opinion is based on the report that is the subject of this review.<sup>5</sup> Chetty received a 2012 MacArthur Foundation “genius” award, and the foundation singled out this study in its citation.<sup>6</sup>

This review shows, however, that the authors’ methods fail in several significant ways: They neglected, in their revised report, to even mention a result that contradicts their conclusion; their claim that they did not have sufficient data for adults older than 28 is false; they used misleading calculations to achieve their results; and they claimed that evidence from other studies they cite supports their methods—when those other studies offer little or no such support. In addition, their own results show that the calculation of teacher value-added is unreliable. The only valid conclusion from this study is the opposite of what’s been reported and trumpeted: that teacher value-added scores have *not* been shown to have a long-term impact on income.

## II. Findings and Conclusions of the Reports

The first part of the report (NBER Working Paper No. 19423) reviewed here linked information about students and teachers in grades three through eight in New York City

(NYC), spanning the years 1989-2009. The research used this linked dataset for “value-added” (VA) calculations for individual teachers. The model used for the VA calculation controls for factors such as students’ prior achievement, parents’ income, and the performance of other students in the classroom. But none of these factors can entirely predict a child’s performance. After accounting for all known and easily measurable social and economic factors, a residue still remains. The report assumes that this residue is attributable to the teacher, and it calculates a teacher’s value-added by using the residues of his or her students.

The second part of the report then linked the incomes of adults with the value-added of their teachers when those earner-adults were students. Using this linked data set they found that a one unit (one standard deviation) increase in teacher value-added increases income at age 28 by \$286 per year or 1.34%. The study then assumes that this percentage increase in income will hold for a person’s entire working life, producing a cumulative lifetime increase of \$39,000 per student.

### **III. The Report’s Rationale for Its Findings and Conclusions**

The report first assumes that it is possible to measure the quality of a teacher by the standardized test results of his or her students and develops a statistical technique to conduct such measurements. The result of the measurement is called teacher “value-added,” and the report concludes with the finding that a higher value-added score for the teacher causes higher incomes in adulthood for his or her students. Relying on these assumptions and findings, the report suggests that low value-added teachers should be fired or otherwise removed and replaced by high value-added teachers.

### **IV. The Report’s Use of Research Literature**

The literature will be reviewed in the context of the following sections.

### **V. Review of the Report’s Methods and of the Validity of the Findings and Conclusions**

#### **Review of Part II: Teacher Value-Added and Student Outcomes in Adulthood**

Part II is reviewed first because it includes the most significant policy results. This part of the report investigates the effect of teacher value-added on the economic well-being of students in adulthood. Its main finding is that a one-unit (one standard deviation) increase in teacher value-added raises income by 1.34% at age 28. In a normal distribution

of teachers' value-added scores, 34.1% of those scores will fall between the mean and one standard deviation below the mean. That is, this one-unit increase can be thought of as moving from a teacher at the 16<sup>th</sup> percentile to the one at the 50<sup>th</sup> percentile. The report then assumes that this increase of 1.34% will persist throughout a person's working life, resulting in a cumulative lifetime increase of \$39,000 per student.

There are five problems with this part of the report, each of which is discussed in greater detail below:

1. An earlier version of the report found that an increase in teacher value-added has no effect on income at age 30, but this result is not mentioned in this revised version. Instead, the authors state that they did not have a sufficiently large sample to investigate the relationship between teacher value-added and income at any age after 28, but this claim is untrue. They had 220,000 observations (p. 15), which is a more than sufficiently large sample for their analysis.
2. The method used to calculate the 1.34% increase is misleading, since observations with no reported income were included in the analysis, while high earners were excluded. If done properly, it is possible that the effect of teacher value-added is to *decrease*, not increase, income at age 28 (or 30).
3. The increase in annual income at age 28 due to having a higher quality teacher "improved" dramatically from the first version of the report (\$182 per year, report of December, 2011) to the next (\$286 per year, report of September, 2013). Because the data sets are not identical, a slight discrepancy between estimates is to be expected. But since the discrepancy is so large, it suggests that the correlation between teacher value-added and income later in life is random.
4. In order to achieve its estimate of a \$39,000 income gain per student, the report makes the assumption that the 1.34% increase in income at age 28 will be repeated year after year. Because no increase in income was detected at age 30, and because 29.6% of the observations consisted of non-filers, this assumption is unjustified.
5. The effect of teacher value-added on test scores fades out rapidly. The report deals with this problem by citing two studies that it claims buttress the validity of its own results. This claim is both wrong and misleading.

*1. The existence of a crucial conflicting result was not mentioned in the revised report*

The statement by the authors that "28 is the oldest age at which we currently have a sufficiently large sample size to estimate earnings impacts" is contradicted by their own work. An earlier version of their report stated that the authors had 61,639 observations of 30-year-olds, that they tested the hypothesis that teacher value-added affect income at age 30, and that the result of the test was that teacher value-added does not have a statistically significant effect on income at that age.<sup>7</sup>

In the earlier version the authors used unconventional and peculiar language to describe their findings. They stated: "Column 2 shows the effect on wages at age 30. The point estimate is slightly larger than that at age 28, but because the sample is only one-sixth the size, the 95% confidence interval for the estimate is very wide. We therefore focus on

earnings impacts up to age 28 for the remainder of our analysis.”<sup>8</sup> Thus, they appear to have ignored inconvenient findings.

The fact that the sample for 30-year-olds is smaller than the sample for 28-year-olds is irrelevant to the decision of whether the result of this statistical test can be ignored. The only relevant issue is whether a sample with almost 61,639 observations in the previous version, or 220,000 in this version, is sufficiently large. The authors did not discuss this question at all. They not only rejected the result of their own test without examining

*The high volatility of teacher value-added scores indicates that they do not measure true quality of teachers.*

whether the sample was sufficiently large, they also stopped mentioning its existence. (In fact, the sample is sufficiently large.<sup>9</sup>)

While the authors stated that they were focusing on the result at age 28, in fact they did much more than that. They made the assumption that their results for 28 -

year-olds (an increase in income of 0.9%) would have also been the result for any subsequent age group, certainly including 30-year-olds.<sup>10</sup> This version of the report goes even further: it does not mention the 30-year-olds result and assumes that an even larger increase in income persists year after year throughout a person’s life. Given their own evidence that the result for 28-year-olds does not hold for 30-year-olds, the assumption that the increase in income for 28-year-olds will hold steady and fixed for all subsequent years is unjustified.

## *2. The main result distorts the effect of teacher value-added (VA) on income.*

The report found that a 1 standard deviation increase in teacher VA raises wages at age 28 by 1.34% (p. 37). But this result is based on a sample that includes 29.6% of observations with *no* reported wages who were assigned a value of zero income, and it excludes 1.3% of the observations with a wage above \$100,000. Both the inclusion and the exclusion distort the calculation in very fundamental ways.

The reported increase in income was \$286 a year. This constitutes 1.34% of average income, but only because the calculation of average income includes the workers with zero income. An increase of 1.34% in the income of a person who earns an income is meaningful. But to somebody who does not earn anything this number is meaningless. Observations with zero income should not be included in the calculation of the percentage increase of income because they inflate the increase while at the same time rendering it meaningless.

In addition, the inclusion of zero income observations means that the result of the analysis is random. Consider the following example. Suppose there are three workers; two who had a low value-added teacher and one who had a high value-added teacher. Of the first two, one does not report any income and the other earns \$11 per hour. The third worker who had a high value-added teacher earns \$9 per hour. The correct conclusion from these data is that those who had a high value-added teacher had (a) a lower probability of not reporting income and (b) a roughly 20% *lower* wage. Because we expect that a high (or

low) value-added teacher would have a similar effect on all workers, we would estimate that among non-reporters also, the incomes of those who had low value-added teachers were 20% higher than the incomes of those who had high value-added teachers. But including non-reporters in the calculation and attributing zero income to them yields a very different result. It yields a finding that a good teacher *increases* a worker's wage by 53%.

The reason that including people who do not report earning a wage has such a dramatic effect on the calculation is that the distance between zero and most of the positive wages—even for non-filers—are far greater than the differences between the positive wages. The large differences dominate the calculations. Of course, the example above depends on the non-filers having had low value-added teachers. Otherwise the result that value-added has a positive effect on income would have meant that among the employed workers those who had higher value-added teachers earned more. But this is precisely the problem with including non-filers: It hinges on the random proportion of low and high value-added scores among their teachers.

The study excluded workers who earned more than \$100,000, and this too is unjustified. What if high earners had low value-added teachers? Excluding them distorts the result in much the same way that the inclusion of no-filers does.

*3. A result “improved” dramatically from one version to the next, indicating that the correlation between teacher value-added and income later in life may be random.*

In the previous version of the report, the authors found that an increase of 1 standard deviation in teacher VA raises earnings at age 28 by 0.9% or \$182 a year (p. 39). In the latest version the increase in earnings is 1.3%, or \$286 (see note to Figure 8 and Table 3 column 2). The increase from \$182 to \$286 between versions is dramatic—more than 57%. Because the data sets are not identical, a slight discrepancy between estimates is to be expected. But since the discrepancy is large, it suggests that the link between teacher value-added and income at age 28 may be random, entirely dependent on the data set that one happens to use.

*4. The wage growth assumption is not consistent with the evidence.*

The calculation of an increase of 1.34% in income at age 28 is based on the inclusion of 29.6% of the observations of non-filers. Can it be assumed that this proportion would continue throughout the life of the cohort? In addition, as was already discussed, the authors' own test revealed that there was no statistically significant effect of teacher value-added on income at age 30. For both reasons, it cannot be assumed that the increase they found at age 28 would hold forever.

*5. The report cites two studies that it claims buttress the validity of its results. This claim is both wrong and misleading.*

Value-added teacher effects fade rapidly. Jacob *et al.* found that only 25% of the first year teacher effect remains in the second year.<sup>11</sup> Kane and Staiger found that 50% remains in

the second year and 25% remains in the third, and Rothstein found that 30% remains in the second year.<sup>12</sup> Furthermore, Rothstein found that the correlation between the first-year effect and the third-year effect is only .4, leading him to conclude: “A teacher’s first-year effect is a poor proxy for his or her longer-run impact.”<sup>13</sup> If the teacher’s effect does not last even through elementary school, why would it last a lifetime?

The report does not address this question. Instead it cites two studies that show verified long-term effects despite a fadeout in test scores. Only it does not tell the reader that these studies come from a very different area of inquiry. The studies by James Heckman *et al.* of the Perry High Scope program and by David Deming of Head Start demonstrate that these are successful programs (particularly the former). But they are programs that distinctly do not focus on cognition; instead they focus on child and parent psychology, nutrition and health.<sup>14</sup> Furthermore, these programs are beneficial not because they involve the hiring and firing of teachers but because they increase the resources that are available to children and their families.

Heckman *et al.* draw attention to the distinction between the concerns of education economists and the concerns of the early education program. They explain:

The literature in the economics of education assumes the primacy of cognitive ability in producing successful lifetime outcomes. . . From this perspective, the success of the Perry program is puzzling. Although the program initially boosted the IQs of participants, this effect soon faded. . . Consistent with this evidence, we show negligible effects of increases in IQ in producing program treatment effects. Although Perry did not produce long run gains in IQ, it did create persistent improvements in personality skills. The Perry program substantially improved Externalizing Behaviors (aggressive, antisocial, and rule-breaking behaviors), which, in turn, improved a number of labor market outcomes, health behaviors, and criminal activities.<sup>15</sup>

While it is easy to understand why the Perry program and Head Start lead to success in adulthood despite the fade out in test scores, it does not follow that short-duration improvements in elementary school test scores would lead to economic success in adulthood. Citing the success of early childhood programs as evidence for the efficacy of the value-added method of evaluating teachers is a misleading scaffold that cannot bear the weight.

## **Review of Part I: Evaluating Bias in Teacher Value-Added Estimates.**

This part of the report addresses how teacher value-added is measured. There are four problems with this section, each of which is discussed in greater detail below:

1. Value-added scores in this report and in general are unstable from year to year and from test to test, but the report ignores this instability.
2. The report inflates the effect of teacher value-added by assuming that a child’s test scores can be improved endlessly.

3. The procedure that the report develops for calculating teacher value-added varies greatly between subjects within school levels (math or English in elementary/high school) and between schools within subjects (elementary or middle school math/English), indicating that the calculated teacher value-added may be random.
4. The commonly used method for determining how well a model predicts results is through correlations and illustrated through scatterplot graphs. The report does not present either the calculations or graphs and instead invents its own novel graph to show that the measurements of value-added produce good predictions of future teacher performance. But this is misleading. Notwithstanding the graph, it is possible that the quality of predictions in the report was poor.

*1. Value-added scores in this report and in general are unstable from year to year and from test to test, but the report ignores this instability.*

This study, like all other studies of teacher value-added, found that value-added scores vary widely from one year to the next. A measure of the variability is autocorrelation, which indicates whether teachers with high value-added scores one year have high value-added scores in the following year. The highest value auto correlation can attain is 1. The autocorrelations in this study range from .23 to .48, which is low.

Of course, a teacher may have a bad year or a good year. But autocorrelation uses observations on many teachers, and while some of these teachers may experience irregular performance in a particular year, most do not.

A low autocorrelation with a large number of observations means that the high volatility of teacher value-added scores is not an exception but the rule. For example, economists McCaffrey, Sass and Lockwood found low autocorrelations for teacher value-added in four large counties in Florida, ranging from .008 to .36.<sup>16</sup>

Autocorrelation is just one way to investigate the stability of a teacher value-added. McCaffrey *et al.* used their Florida data, and Koedel and Betts used data from San Diego, to examine the stability of value-added scores in an alternative way.<sup>17</sup> Their results varied from place to place, but the average result was that 13% of teachers who were at the bottom 20% of the VA scale in one year were at the top 20% the following year, and 29% of those at the bottom 20% were at the top 40% the following year. Similar results held at the top. Twenty-six percent of teachers who were in the top quintile in one year were in the bottom 40% the following year. Only 28% percent of teachers who were at the top one year stayed at the top the following year.

VA scores are also not consistent across tests. The state of Florida uses two different standardized tests, and McCaffrey and his co-authors found that the two test yielded different results. Of the teachers who were in the bottom 20% of the VA score according to one test, 5% were at the top 20% and 16% were at the top 40% according to another test. McCaffrey *et al.* dealt with tests that were issued by different institutions, but even tests that are issued by the same institution differ, and these differences among tests may be part of the reason that VA measurements vary from year to year.

The high volatility of teacher value-added scores indicates that they do not measure true quality of teachers.

*2. The study inflates the effect of teacher value-added by assuming that a child's test scores can be improved endlessly.*

The study makes the assumption that a child's test scores can be improved endlessly (p. 38). It calculates the effect of having a high VA teacher in every year as the sum of the effects of having a high VA teacher in just one year. This assumption inflates the effect of having high VA teachers and is unrealistic.

*3. The procedure that the report develops for calculating teacher value-added varies greatly between subjects within school levels (math or English in elementary/ high school) and between schools within subjects (elementary or middle school math/English), indicating that the calculated teacher value-added may be random.*

According to the report to calculate a teacher's value-added score in NYC, if the teacher is an English teacher in elementary school, a greater weight is placed on average residuals of her or his students from ten years ago than the average residuals of the students of six years ago. But if it is an English teacher in middle school, lower weights are placed on the residuals of both ten years ago and six years ago, but these weights are equal. If it is a math teacher in elementary school, a smaller weight is placed on the residuals of ten years ago than on the residuals of nine years ago, but if it is a math teacher in middle school, the opposite holds: A higher weight is placed on the residuals of 10 years ago than on the residuals of nine years ago. A similar procedure applies for each grade from 1-10.<sup>18</sup> If this all appears random, it is.

Why would a teacher's performance from 10 years earlier be included when calculating the quality of her teaching today if her performance over the last two or three years is known? And why would a teacher's performance 10 years earlier matter more than their performance six years earlier when they teach elementary school English, while when they teach English in middle school the two would have the same impact? Why would the performance of 10 years ago be more important than the performance of nine years ago in elementary school, while in middle school the reverse is true? And why would the importance within the same year vary so much from one school to the next or one subject to the next? All in all, the consistency (and therefore the validity) of the measure is left in question.

*4. The commonly used method for determining how well a model predicts results is through correlations and illustrated through scatterplot graphs. The report does not present either the calculations or graphs and instead invents its own novel graph to show that the measurements of value-added produce good predictions of future teacher performance. But this is misleading.*

In the report, Figure 2a shows that the *averages* of actual scores of teacher value-added are very close to the average values that are predicted by the model. But this is misleading because the actual individual scores can be far from the predicted individual scores. In

other words, notwithstanding Figure 2a, it is possible that teachers with low predicted value-added scores had high actual scores and that teachers with high predicted scores had low actual scores. The report does not provide readers with this information.

## Summation

The key finding of this two-part report—that teacher value-added affects income in adulthood—finds little support in the evidence. Seven weaknesses and contradictions are described above:

1. Teacher value-added was not shown to have an effect on income at age 30;
2. This result was not even mentioned in the revised report;
3. The calculation of the effect of teacher value-added on income at age 28 is distorted by the data choices;
4. The value-added scores are unstable and fade out;
5. Studies cited as validating the report's methodology provide no such support;
6. A novel and problematic method is used to claim that the model is sound; and
7. An unrealistic assumption that is contradicted by the report itself is used to inflate the results.

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

The two-part report under review here uses questionable techniques to reach conclusions that are not supported by the data. These problems render the report of no value in guiding educational policy.

## Notes and References

---

- 1 I wish to thank Henry Braun for his guidance and encouragement and Ellen Adler, Bill Mathis, Kevin Welner, and an additional (blinded) reviewer for their editorial assistance.
- 2 Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013). *Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates* (Working Paper No. 19423). Cambridge, MA: National Bureau of Economic Research. Retrieved March 1, 2014, from <http://www.nber.org/papers/w19423>;  
  
Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013). *Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood* (Working Paper No. 19424). Cambridge, MA: National Bureau of Economic Research. Retrieved March 1, 2014, from <http://www.nber.org/papers/w19424>.
- 3 Lowery, A. (2012, January 6). Big study links good teachers to lasting gain. *The New York Times*, A1;  
  
Massari, P. (2012, February 21). Superstar teachers. *The Harvard Gazette*. Retrieved March 1, 2014, from <http://news.harvard.edu/gazette/story/2012/02/superstar-teachers/>.
- 4 Presentation outline by Chetty and his coauthors posted by the New Zealand Department of the Treasury. Retrieved March 1, 2014, from <http://www.treasury.govt.nz/downloads/pdfs/gen-conf-14dec11-chetty.pdf>.
- 5 Chetty, R. & Olssen, A. (2013, December). The impacts of last-in, first-out vs. effectiveness-based teacher layoff policies. *Beatrix Vergara v. State of Ca., Case No. BC484642*, Exhibit 236.
- 6 The MacArthur Foundation (2012, October 2). *Meet the Class of 2012: Raj Chetty*. Retrieved March 1, 2014, from <http://www.macfound.org/fellows/861/>.
- 7 Chetty, R., Friedman, J. N. & Rockoff, J. E. (2011, December). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (Working Paper No. 1769), Table 6. Cambridge, MA: National Bureau of Economic Research. Retrieved March 1, 2014, from [http://obs.rc.fas.harvard.edu/chetty/value\\_added.pdf](http://obs.rc.fas.harvard.edu/chetty/value_added.pdf).
- 8 Chetty, R., Friedman, J. N. & Rockoff, J. E. (2011, December). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (Working Paper No. 1769), 39. Cambridge, MA: National Bureau of Economic Research. Retrieved March 1, 2014, from [http://obs.rc.fas.harvard.edu/chetty/value\\_added.pdf](http://obs.rc.fas.harvard.edu/chetty/value_added.pdf).
- 9 I analyzed this issue in a recent peer-reviewed article:  
  
Adler, M. (2013, February 1). Findings vs. interpretation in “The long-term impacts of teachers by Chetty *et al.*” *Education Policy Analysis Archive*, 21 (10). Retrieved March 1, 2014, from <http://epaa.asu.edu/ojs/article/view/1264/1033>.
- 10 Chetty, R., Friedman, J. N. & Rockoff, J. E. (2011, December). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (Working Paper No. 1769), 39. Cambridge, MA: National Bureau of Economic Research. Retrieved March 1, 2014, from [http://obs.rc.fas.harvard.edu/chetty/value\\_added.pdf](http://obs.rc.fas.harvard.edu/chetty/value_added.pdf).
- 11 Jacob, B., Lefgren, J. B. & Sims, D. (2010). The persistence of teacher-induced learning gains. *Journal of Human Resources*, 45(4), 915-943.
- 12 Kane, T. J. & Staiger, D. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (Working Paper No. 14607). Cambridge, MA: National Bureau of Economic Research. Retrieved March 1, 2014, from <http://www.nber.org/papers/w14607>;

- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125(1), 175-214.
- 13 Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125(1), 175-214; 209.
- 14 Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics* 2009, 1 (3), 111–134. Retrieved March 1, 2014, from [http://www.people.fas.harvard.edu/~deming/papers/Deming\\_HeadStart.pdf](http://www.people.fas.harvard.edu/~deming/papers/Deming_HeadStart.pdf);
- Heckman, J., Pinto, R., & Savelyev, P. (2012). *Understanding the mechanisms through which an influential early childhood program boosted adult outcomes*. Retrieved March 1, 2014, from [https://economics.sas.upenn.edu/sites/economics.sas.upenn.edu/files/u21/o\\_PerryFactorPaper\\_AER\\_2012-09-07\\_sjs.pdf](https://economics.sas.upenn.edu/sites/economics.sas.upenn.edu/files/u21/o_PerryFactorPaper_AER_2012-09-07_sjs.pdf).
- 15 Heckman, J., Pinto, R., & Savelyev, P. (2012). *Understanding the mechanisms through which an influential early childhood program boosted adult outcomes*, 3. Retrieved March 1, 2014, from [https://economics.sas.upenn.edu/sites/economics.sas.upenn.edu/files/u21/o\\_PerryFactorPaper\\_AER\\_2012-09-07\\_sjs.pdf](https://economics.sas.upenn.edu/sites/economics.sas.upenn.edu/files/u21/o_PerryFactorPaper_AER_2012-09-07_sjs.pdf).
- 16 Sass, T. (2008, November). *The stability of value-added measures of teacher quality and implications for teacher compensation policy*. Washington, DC: The Urban Institute. Retrieved March 1, 2014, from <http://files.eric.ed.gov/fulltext/ED508273.pdf>.
- 17 Sass, T. (2008, November). *The stability of value-added measures of teacher quality and implications for teacher compensation policy*. Washington, DC: The Urban Institute. Retrieved March 1, 2014, from <http://files.eric.ed.gov/fulltext/ED508273.pdf>.
- 18 See also:
- Ballou, D. (2011). *Review of “The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood.”* Boulder, CO: National Education Policy Center. Retrieved March 1, 2014, from <http://nepc.colorado.edu/thinktank/review-long-term-impacts>.

DOCUMENTS REVIEWED:

**Measuring the Impacts of Teachers  
I: Evaluating Bias in Teacher Value-  
Added Estimates;  
II: Teacher Value-Added and  
Student Outcomes in Adulthood**

AUTHORS:

Raj Chetty, John N. Friedman,  
and Jonah E. Rockoff

PUBLISHER:

National Bureau of Economic Research  
(NBER Working Papers Nos. 19423  
and 19424)

DOCUMENT RELEASE DATE:

September 2013

REVIEW DATE:

April 10, 2014

REVIEWER:

Moshe Adler, Columbia University  
and Empire State College, SUNY

E-MAIL ADDRESS:

ma820@columbia.edu

PHONE NUMBER:

(917) 453-4921

SUGGESTED CITATION: Adler, M. (2014). *Review of “Measuring the Impact of Teachers.”* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-measuring-impact-of-teachers>.