REVIEW OF *THE LONG-TERM IMPACTS OF TEACHERS*

Reviewed By

Dale Ballou

Vanderbilt University
February 2012

Summary of Review

A new report titled *The Long-Term Impacts of Teachers* concludes that teachers whose **students tend to show high gains on their test scores (called "high value-added teachers")** also contribute to later student success in young adulthood, as indicated by outcomes such as college attendance and future earnings. To support this claim, it is not sufficient for researchers to show an observed association between teacher value-added and later outcomes in young adulthood. It is also necessary to rule out plausible alternative explanations—for example, that parents who **did the most to promote their offspring's** long-term success also endeavored to secure high value-added teachers for their children. This review explains that, for the most part, the evidence needed to rule out these alternatives is missing from the report. Thus, policy-makers should tread cautiously in their reaction: the case has not been proved.

REVIEW OF *THE LONG-TERM IMPACTS OF TEACHERS*

Dale Ballou, Vanderbilt University

I. Introduction

Value-added assessment of teachers is an attempt to measure the contribution of teachers to student learning, as gauged by progress on standardized tests. As a procedure for evaluating teachers, it remains controversial and widely unpopular with teachers. Part of this controversy involves concerns that a policy of high-stakes teacher evaluation linked to **students' test scores would intensify practices such as the narrowing of curriculum and** teaching to the test. A second part involves questions of whether value-added models are reliable and valid. This review focuses on the latter, modeling and measurement questions, setting aside for the moment those other elements of the policy debate.

Within the measurement discussion, there are concerns that the statistical methods do not sufficiently control for other factors that influence test-score gains. But this is not all. The focus on testing is deemed unsatisfactory, in light of the following: (1) Tests are often poorly aligned with the curriculum and fail to reflect what is taught; (2) Test performance is noisy, measuring student ability with considerable measurement error; (3) Other **important parts of a teacher's job (e.g., building character, teaching citizenship) are not captured by performance on tests. Thus, evaluating teachers on the basis of students'** performance on a single test (an examination that frequently holds no consequences for students) can yield a reading on teacher performance that is incomplete at best, and possibly seriously misleading.

A new report, *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood*, by Raj Chetty, John N. Friedman, and Jonah E. Rockoff, adds important new evidence that helps address these concerns. Using data from a large urban school system, the report asks whether teachers who are effective a**t raising students' test scores also have positive impacts on students' subsequent transition from school to young** adulthood.[1] Among these indicators are the probability of going to college, earnings once students enter the workforce, the probability that female students give birth as teenagers, the quality of the neighborhood in which students live at age 25, and the establishment of a retirement savings account.

The study finds a broad pattern of positive effects associated with teachers who are effective at raising student test scores. Given these observed positive effects, the report

goes on to consider how public policy might enhance students' opportunities to learn from such instructors.

II. Findings and Conclusions of the Report

The report estimates the long-term impact of reading/language arts and mathematics **teachers in grades 4 through 8. Teachers who are effective at raising students' test scores** in either of these subjects also have a positive effect on other outcomes several years later. Specifically, increasing by one standard deviation the value-added of a reading or math teacher at one of these grade levels has the following beneficial consequences:

- It increases the probability that a student is in college at age 20 by half a percentage point (relative to an average probability of 37.8% in this sample). Though greatest at younger ages, there continues to be a positive impact on college attendance through age 25 (0.28 percentage points).

- It increases the quality of the college attended (as measured by average annual earnings **of that college's graduates) by $164. This estimate combines the impact on students who** would not have gone to college otherwise (in which case the increase is measured from the average earnings of those who do not attend college) and the impact on students who attend a better college than they otherwise would have. Both components are positive.

- It increases annual earnings at age 28 by $182, nearly a full percentage point gain over the sample mean of $20,912. Assuming this 1% gain persists over a working life, the estimated impact on lifetime income has a net present value of $4,600. This may well be an understatement of the effect, inasmuch as the students of high value-added teachers are more likely to attend college. They **reach age 28 with fewer years' experience in the** labor market, but on a higher growth trajectory than those who did not attend college. Calculations taking into account these factors suggest that the lifetime gain might be nearer $5,700.

- It reduces the probability that a female student gives birth while a teenager by 0.099 percentage points (relative to a mean of 8%).

- It raises the probability of living in a high-SES neighborhood at age 25 (measured as the percentage of residents who are college graduates) by 0.063 percentage points, an impact that more than doubles at age 28.

- There is no positive impact on the probability that a student contributes to a 401(k) retirement savings account at age 25, though this may reflect the fact that the students of higher value-added teachers are more likely to have attended college and therefore are less likely to have found a job in which they begin to save for retirement by this age.

To be clear, the foregoing estimates represent the impact of an increase of one standard deviation in teacher value-added in *one* subject in *one* grade (loosely speaking, the effect of having one really good reading or math teacher at some point in grades 4 through 8). It is also important to stress that according to the report, these are not mere measures of association, reflecting the fact that students who have had high value-added teachers are

also more likely to go on to college, obtain higher-paying jobs, etc. Such associations might arise for a variety of reasons other than the quality of the teacher. (For example, parents who care enough to get their children into the classrooms of the best teachers might also **do a better job of preparing them for life's later challenges). The report claims that these** positive outcomes are the *result* of having had higher value-added teachers, above and beyond any association that might arise for other reasons. The methods used to support these causal claims are described in Section V below.

Finally, it should be borne in mind that the foregoing findings are estimates and subject to error. However, even allowing for estimation error, the evidence is strong enough to reject with a high degree of confidence the hypothesis that the true impact of high-value-added teachers on these long-term indicators is zero.

The report also investigates whether the long-term impact of high value-added teachers varies over student subgroups: female and male, low and high income, and minority and non-minority. For each of these subgroups, only one long-term outcome is examined: the

The report is highly persuasive on a key point: teacher value-added has been measured free of significant bias. However, measuring value-added is only a step along the way to the report's larger findings: that high value-added teachers improve students' long-term outcomes.

quality of the college attended at age 20. The impact of teacher value-added is greater for females, high-income students, and non-minorities, suggesting the presence of complementarities between teacher value-added and family inputs. Students more likely to experience better outcomes anyway appear to be more receptive to the impact of these high value-added teachers. However, the impact is positive in all of these groups. The estimated effect of a high-value-added teacher is also greater when that teacher is encountered in middle school rather than in elementary school, though the impact, again, is positive at both levels.

Given the broad pattern of positive long-term effects, the report goes on to consider implications for public **policy: what can be done to enhance students' opportunities to** learn from such instructors. Bonuses might be paid to effective teachers, for instance, to increase the probability that they remain in the profession. However, the authors conclude that such an approach is not cost-effective, since most of these bonuses would be paid to **teachers who would remain in the classroom anyway. The net gain in students' future** incomes would barely exceed the cost of financing the bonuses.[2]

A second option described in the report explores whether replacing teachers who have the lowest value-added scores with teachers of average effectiveness has much larger net gains. Applied to teachers in the bottom 5% of the value-added distribution, such a policy generates substantial economic benefits. The expected gain increases the longer one waits to make the replacement, because more years of data are then available to determine

whether such a teacher is truly in the bottom 5%. However, waiting imposes a still greater cost, as these teachers—most of whom would likely remain low value-added teachers—remain in the classroom longer. Assuming a class size of 28 students, the report contends that a policy of replacing a teacher who turns up in the bottom 5% of the distribution in her first year is expected to raise the lifetime earnings of that class by $135,000 (in present value). Given the number of years such a teacher might otherwise remain on the job, the total gains are potentially substantial.

III. The Report's Rationale for Its Findings and Conclusions

The report tackles its research question in two stages.[3] First, estimates of teacher value-added are obtained using data from a large urban school system over the years 1989 to 2009. (School years are identified by the year of the spring semester.) The data include more than 18 million test scores in reading/ELA and math for students in grades 3-8. Students were linked to their instructors in these subjects. In the lower grades these were often the same individual for both reading and math. Whether this was the case or not, separate estimates of teacher value-added were obtained for each subject.

A key goal in estimating teacher value-added is to isolate the effect of the teacher from other influences on student achievement. The model used in this study controlled for a **number of such influences, including each student's achievement in the prior academic** year. There were also controls for a variety of student, classroom, and school characteristics. These included student ethnicity, gender, age, limited English proficiency, and whether the student received special education services. Additional classroom-level variables included classroom type (honors, remedial) and size. Classroom- and school-level means of these same student characteristics were also included in the model. Controls were also introduced to capture variations in average test performance by grade and by year.

The purpose of such models is to estimate the expected test performance of students as a function of the aforementioned variables. In effect, such a model predicts what a student with a given set of individual, class, and school characteristics is expected to score if assigned a teacher of average ability.[4] To the extent that students of a given teacher tend to score above or below this expectation, the difference, averaged over those students, is **attributed to the instructor's influence: it becomes the measure of teacher value**-added.

Inevitably, concern arises that the model has not controlled for enough other factors: that something omitted becomes confounded with the effect of the teacher. The mere fact that the model does not control for every influence on student achievement is not automatically a source of bias. Such influences need to vary systematically by teacher—that is, some teachers are normally assigned students who, for reasons not observed, score higher than expected, while other teachers are normally assigned students who do worse. This imparts a positive bias to the estimated value-added of the former and a negative bias to the value-added of the latter. The report contains several tests examining whether important omitted

factors have biased estimates of teacher value-added. The report concludes on the basis of these tests that there are no significant biases.

The second stage of the analysis investigates whether the measures of teacher value-added **obtained in the first stage are related to students' long**-term outcomes. Information on these outcomes is obtained from tax records, which are merged with school district student records. College attended is identified from 1098-T forms filed by postsecondary institutions on tuition payments received for every student. Neighborhood quality is derived from zip codes on 1040 forms. Teenage births are detected based on whether a **young woman claims a new dependent on her tax returns. Information on students'** parents is obtained from the earliest 1040 form on which the student was claimed as a dependent. The use of tax records means some outcomes are not measured with perfect accuracy (for example, teenage births to women not filing tax returns). However, the overall match rate is above 80% and the amount of information obtained in this manner is impressive.

To ascertain whether teacher value-added has affected these outcomes, it is again necessary to estimate a model. The mere fact of an association between teacher value-added and a subsequent outcome, such as earnings, does not in itself establish a causal relation. It is also necessary to rule out the possibility that other factors, related to both value-added and earnings, is responsible for the observed association—a point to which we will return in Section V.

IV. The Report's Use of Research Literature

The report makes effective use of previous teacher value-added research. The models estimated are informed by prior work. The additional tests conducted to establish the validity of the causal inferences, while innovative, build on the ideas of other researchers. On the basis of objections that have been raised to earlier studies, the report anticipates **many of its readers' concerns. Findings that deviate from those of the most prominent** prior reports are noted along with the most likely explanations.

V. Review of the Report's Methods

Establishing the validity of the report's conclusions requires two things. First, it must be shown that teacher value-added has been measured free of bias. If the contrary turns out to be true—if, for example, teachers with high measured value-added are systematically assigned students who are more likely to have above-average test-score gains, say because of parental inputs—then the teacher value-added derived from the models measures something other than the contribution of the teacher, and inferences about the long-term effect of teachers would break down.

However, establishing that value-added measures are free of bias is only half the battle. Even if we could be assured that the value-added model perfectly measures a teach**er's** contribution to student knowledge, it would remain problematic to attribute long-term

outcomes to the impact of teachers. Once again, the principal difficulty arises with regard to the way students are assigned to teachers. To the extent that these assignments can be affected by parents, it may be that parents with a high concern for their offspring's long-term success take steps to place them in the classrooms of the most effective (i.e., high value-added) teachers. Or they may be the passive beneficiaries of school policies such as ability grouping that make such assignments more likely. That the offspring of these same parents then enjoy greater success in pursuing higher education, in obtaining jobs, etc., may say more about their parents than the quality of their instructors. An association between high value-added teachers and subsequent life success would be observed, but it would not be a casual relation, but the result of good parenting. Thus, the second task is to establish that no such selection mechanism has been at work: that high value-added teachers are not more likely to have been assigned students who were for other reasons (e.g., parents) destined for greater long-term success.

This section reviews the methods employed in the report to establish these two claims. To anticipate, the report contains compelling evidence supporting the first: that measured value-added is free of bias. However, the report contains much less evidence supporting the second: that students on a path to greater life success were not more likely to have been placed in classes taught by high value-added teachers. The same methods used so persuasively to establish the first point are not consistently applied to establish the second, though there is no apparent reason why they could not have been.

Establishing that estimates of value-added are unbiased

One way to investigate bias is to examine the kinds of students assigned high value-added teachers. The student characteristics used for this purpose cannot be the same characteristics that were controlled for when teacher value-added was estimated (e.g., ethnicity, gender, limited English proficiency, and prior test score). Since those teacher estimates are based on residual achievement gains, by construction there will be no association between measured value-added and those controls. However, other student characteristics not controlled for can be used for this purpose. This study is rich in having available a large set of such variables, drawn from parental tax returns, including **information on household income, mother's age at childbirth, financial assets and home** ownership, and marital status. These variables are strongly correlated with student test scores. Indeed, they are strongly correlated with test results even after controlling for the variables (such as prior achievement) that were used to estimate teacher value-added. Thus, if there exist factors simultaneously influencing test-score gains and placement in the classrooms of high value-added teachers, this set of variables represents a likely place to look for them.

No such association is found. High value-added teachers are not more likely to have been assigned students primed to make large test-score gains by virtue of the values of these other factors. Of course, even this set of variables does not contain all potentially relevant information about student and family background. Some as yet unobserved factors could cause value-added to be measured with bias. Hence the report presents the results of a

second kind of test, based on quasi-experimental variation in the assignment of teachers to students: specifically, changes in average value-added caused by teachers moving in and out of a particular grade level within a school. (These measures of value-added are based **on teachers' performance in other years, not the year of the move.) For the most part,** parents do not have time to respond to such moves by changing schools, nor is it likely most would do so based on the movement of a small number of teachers. Of course, **parents can still attempt to pick and choose among the teachers at their child's grade level.** For this reason, the quasi-experimental test relies not on the value-added of a given **student's teacher. Instead, it relies on mean value**-added of the teachers providing **instruction in that student's grade. Sorting of students into particular classrooms within** the grade does not affect this mean. Thus, if there is an association between mean teacher value-added and subsequent test scores for students in that grade, it cannot be due to the way students were assigned to teachers within that grade, but only to the movement of teachers in and out of that grade. In other words, the association would be insulated from the factors that assign certain students to certain teachers in such a way as to bias value-added estimates.

Two questions are now asked. First, do test scores increase in the year of the move when teachers coming into a grade have higher value-added than the teachers who left? Second, can we rule out the occurrence of other suspicious changes at the same time? For example, are there simultaneous changes in parental characteristics that would predict improved scores? Are there changes in test results in subjects where mean teacher value-added has not changed, or in adjacent grades that are unaffected by teacher moves?

The answer to the first question is yes. An increase in mean teacher value-added at the grade level is highly predictive of an increase in average year-end test results. This cannot be attributed to the favorable matching of students to teachers, because it applies to the entire grade: it is the mean change. The answer to the second question is no. There is no detectable change in the make-up of the students attending the school that would have predicted this increase. The placebo tests (looking for higher scores in subjects and grades not affected by teacher moves) also show no change. Thus, it is unlikely that the change in scores is attributable to anything other than the improvement in teacher quality resulting from teacher movement, as measured by teacher value-added.

It is worth noting that these results could not have been taken for granted. The movement of teachers among schools is not the same as a controlled experiment. Depending on district policies, the configuration of neighborhoods, etc., there may be a tendency for better teachers to seek and obtain jobs in better schools. Or teachers may be moving along with students as attendance zones change and as schools open and close. Thus it is important to show that the movement of a high-value-added teacher into a particular school was not accompanied by other changes that would have produced higher test scores. The fact that no such changes were detected strengthens the conclusion that the subsequent improvement in test scores had no other cause than the arrival of high value-added teachers.

Establishing that estimates of long-term effects are unbiased

Estimating teachers' long-term effects is subject to the same potential biases as the estimation of teacher value-added itself: the presence of other factors (most likely home and family background) that lead high value-added teachers to be assigned students with above-average long-term prospects. Thus the same two sorts of tests could be used to test for bias.

First, one could test whether students whose family backgrounds are conducive to better long-term outcomes more frequently turn up in the classrooms of high-value-added teachers. One may wonder if this was not already done in testing whether value-added was estimated without bias. It was not. The tests described in the preceding subsection asked whether family background factors that predict **high test scores** are distributed in such a way as to systematically favor teachers with high value-added estimates. What is needed

It must be established that the observed association between value-added and later earnings, college attendance, etc., is not the result of some third factor, such as good parenting.

now is to test whether family background factors that predict long-term success in **employment, college attendance, etc.,** are distributed in such a way as to favor high-value-added teachers. It should not be assumed that the same mix of factors that predicts high test scores necessarily predicts the others. While one would expect some overlap, it is likely that success in finding a job or in avoiding a teen pregnancy also depends on the values that parents impart and on their ability to shape character, factors that may not be of equal importance in predicting cognitive function on a standardized achievement test. Thus, the tests for bias need to be run again, allowing for the mix of family background factors to change depending on which long-term outcome is under examination. Indeed, assuming that all we need do is test whether high-value-added teachers have benefitted by being assigned students whose family backgrounds predict high test scores is to beg the question that is the focus of this report: do the same factors that predict high scores also predict these other outcomes?

Second, a set of quasi-experimental tests could be conducted, relying once again on variation in teacher value-added that arises when teachers move between grades or schools. The same questions should be asked as before. When high-value-added teachers move into a grade, do we see improved long-term outcomes? Do we see improved long-term outcomes where they are not expected (e.g., for students in an adjacent grade that did not experience a change in the mix of teachers)? The reason for conducting these quasi-experimental tests is the same as the reason given in the preceding paragraph. The fact that similar tests have already validated the estimates of teacher value-added does not imply that they are not needed to validate inferences about the impact of teachers on long-term outcomes. Indeed, it is probably even more important that these tests be conducted with respect to outcomes like earnings and the avoidance of teen pregnancy. Data available

from tax returns are not likely to distinguish well between families that nurture the development of character and families that are much less successful in this task. Unobserved differences between families are likely to be very important. The way to test whether high-value-added teachers have been systematically assigned more students whose families are richer with respect to these unobservable factors is to conduct the quasi-experimental tests described above. Absent that, we will not know whether such differences were present and were the underlying reason for the observed association between teacher value-**added and students' long**-term success.

The report contains the results of few such tests. There are no tests of the first kind, using additional data on parents from tax records to see whether the inclusion of this information in models predicting long-term outcomes takes away from the explanatory power of teacher value-added. As for quasi-experimental tests, only two are reported. An increase in mean teacher value-added caused by teacher movement in and out of a grade raises the probability of enrolling years later in college and improves college quality. However, no tests of this kind are reported for the other long-term outcomes investigated: earnings, residential neighborhood quality, the probability of a teenage birth, and contribution to a retirement savings account.

VI. Review of the Validity of the Findings and Conclusions

As just noted, the report is highly persuasive on a key point: teacher value-added has been measured free of significant bias. However, measuring value-added is only a step along the **way to the report's larger findings: that high value-added teachers improve students' long**-term outcomes. Clearly value-added must be measured accurately in order to make the latter claim, but that is not sufficient. It must be established that the observed association between value-added and later earnings, college attendance, etc., is not the result of some third factor, such as good parenting. On this key point the report falls short. Much more evidence could have been presented to support this claim. The same kind of tests conducted to establish that value-added was estimated free of bias could have been applied to test the larger and more significant claim of this report: that high-value-added teachers improve life outcomes many years after students have left their classrooms. In the absence **of such evidence, it is premature to conclude that the report's central conclusions are** correct

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

The report offers an impressive set of data and analyses that add substantially to our knowledge base. The report includes several persuasive tests that substantiate the claim that teachers with high measured value-**added are in fact raising students' test scores: it is not merely an observed association, but a causal impact. However, the report's findings with regard to students' long**-term success do not include adequate tests of this sort. Thus, **the report's key findings linking teacher value**-added scores to outcomes such as later

earnings are not sufficiently validated; more evidence on that point needs to be presented before policy and practice are shaped in response to this report.

The report **raises a question that is both important and timely. If the report's conclusions** regarding long-term effects can be substantiated, it would strongly suggest that high value-added teachers do more than simply raise test scores. To at least a modest extent, these teachers would be shown to be transformative, changing students for the better in ways that do not show up for years to come. Given the number of students with whom a teacher comes into contact over the course of a career, these modest impacts could have a large cumulative effect.

Notes and References

1 Chetty, R., Friedman, J. N. & Rockoff, J. E. (2011). *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood*. NBER Working Paper No. 17699. Retrieved January 29, 2012, from http://obs.rc.fas.harvard.edu/chetty/value_added.pdf.

2 The report does not estimate the impact of bonuses on the probability of a teacher**'s remaining in the classroom,** but it uses an estimate from a study of North Carolina teachers, finding that a $1,800 bonus reduced teacher attrition by 17%. See:

Clotfelter, C., Glennie, E. Ladd, H., & Vigdor, J. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics 92*, 1352-70.

Whether this estimate would hold good in the urban system considered in this study is open to question. The bonus program examined in the NC study was also short-lived, raising doubts about the long-term versus the short-term response. It is possible that the long-term response to such a bonus differs from the short-term response: teachers might defer a decision to quit in order to obtain a bonus, but ultimately they would leave anyway. If 17% overstates the decline in attrition in response to a bonus of $1,800, such a policy would be still less cost-effective.

3 The description in this review omits many technical details. Omission of a particular issue should not be taken as an indication that the report likewise fails to mention it.

4 The term *teacher of average ability* must be qualified. It is the teacher of average ability for students with of the same characteristics. If there are differences in average teacher quality based on student, class, and school characteristics, these show up in the estimated effect of those variables on achievement, and are not part of the residual variation in student achievement that is the basis of teacher value-added.