Teacher Quality and Equity in Educational Opportunity: Findings and Policy Implications

June C. Rivers and William L. Sanders

In the following piece, Dr. June C. Rivers and her husband, Dr. William L. Sanders, both formerly of the University of Tennessee, describe how teachers can be evaluated based on the academic gains students make in their classrooms. Since 1992, these two statistical pioneers have annually compiled standardized test scores for approximately six million Tennessee children in a variety of academic fields. Along with test score information, data are collected on a wide range of student variables, including ethnicity.

This innovative approach allows student progress to be assessed from year to year, without controlling for external influences like poverty and family conditions. The longitudinal approach allows each student to serve as his or her own control, and empirical studies of the early results showed these analyses to be virtually unbiased by these external influences. Testing is sustained and consistent, which allows them to track what academic gains individual students make from one year to the next, regardless of the background of the children. Thus, teachers are considered as effective if they elicit appropriate gains for their students.

The findings have been provocative. Differences in teacher ability are substantial, and if students are assigned to consecutive ineffective teachers, the impact on student achievement in the short and long terms can be devastating. Most important, successful teachers can elicit significant gains from students of all ethnicities and income levels. This value-added model has potential to revolutionize how teacher quality is assessed, using objective measures to evaluate teachers.

INTRODUCTION

"Equity in education" is not a new phrase but one that has a diversity of meanings in educational policy circles. The different meanings of equity are accompanied by disagreements as to both the definition and the measurement of it. To some, equity in educational delivery will be achieved only when simple group averages across various demographic subgroups are equal. This simplistic definition of equity has resulted in huge debates and disagreements, ranging from accusations of ethnic biases in the measurement process (or instruments) to the argument that having all students at the same academic attainment level at the same time is an unrealistic expectation for educators. However, one definition of equity avoids much of this debate and allows for a more realistic measurement process to be put in place.

If true equity is defined as each student making appropriate academic growth each year, then expectations for *educators and students* can be set in terms of academic growth rates. The results of newer research indicate that the academic growth rate of student populations is primarily a function of the effectiveness of school districts, schools, and, most important, teachers. If appropriate rates of academic growth are sustained across grades, then *all* students' academic attainment will be ratcheted to higher levels. The achievement levels of second- and third-graders are not nearly as important as their attainment levels when they are eleventh- and twelfth-graders. Measurement methodology that separates educational influences from a multitude of possible confounding biases provides realistic diagnostic feedback for educators. Practice informed by appropriate measurement ensures that all students have opportunities to reach their full potential.

Tennessee Value-Added Assessment System

The Tennessee Value-Added Assessment System (TVAAS) is one statewide system that measures the impact that districts,

schools, and teachers have on academic growth rates of student populations. The TVAAS database contains approximately six million student achievement test records from 1991 to the present. The individual student information was linked to specific teachers in 1994, allowing estimation of teacher effectiveness.[1]

The TVAAS accommodates learning indicators from a variety of tests, both multiple-choice and those requiring open-ended responses. To be included, tests must have high repeatability and strong correlation with curricular objectives, and they must also allow for sufficient discrimination at the extremes of the achievement spectrum.[2]

The TVAAS applies statistical mixed model methodology to a longitudinal database that has been created from Tennessee's testing regime, which measures each student each year in five subjects. The educational influences on academic gain are estimated from a multivariate longitudinal model that uses all information for each student, no matter how sparse or complete. With this methodology, the TVAAS avoids many of the problems that have traditionally barred the use of achievement data in assessing effectiveness of schools and teachers; with the TVAAS, (1) exogenous influences are separated from test performance because students are allowed to serve as their own controls; (2) longitudinal analysis across years with repeated measures across subjects improves the efficiency of the estimates of the model parameters; (3) all available data are used and no imputation techniques are required; (4) at the teacher level, estimation of shrinkage protects against fortuitous misclassification. The TVAAS database allows exploration of the effect of the teacher's impact on student achievement.[3]

---

[1]William L. Sanders, A. M. Saxton, and S. P. Horn, "The Tennessee Value-Added Assessment System: A Quantitative, Outcomes-Based Approach to Educational Assessment" in *Grading Teachers, Grading Schools*, edited by J. Millman, Thousand Oaks, Calif.: Corwin Press, Inc., 1997.

[2]Ibid.

[3]Ibid.

Pertinent Findings from the University of Tennessee Value-Added Research and Assessment Center

Many research findings from the TVAAS, replicated by other researchers, are pertinent to the issue of teacher quality. The major findings summarized here may be useful for policy makers as they attempt to provide equitable opportunities for all students.

- The effect of teachers can be separated from ethnic, socioeconomic, and parental influences.[4]
- The variability of teacher effectiveness increases across grades and is most pronounced in mathematics.[5]
- In the extreme, fifth-grade students experiencing highly ineffective teachers in grades three through five scored about 50 percentile points below their peers of comparable previous achievement who were fortunate enough to experience highly effective teachers for those same grades.[6]
- A teacher's effect on student achievement is measurable at least four years after students have left the tutelage of that teacher.[7]

---

[4]D. A. Harville, "A Review of the Tennessee Value-Added Assessment System (TVAAS)" (manuscript, Iowa State University, 1995); H. R. Jordan, R. L. Mendro, and D. Weerasinghe, "Teacher Effects on Longitudinal Student Achievement" (paper presented at the National Evaluation Institute, Indianapolis, Ind., 1997); William L. Sanders and S. Horn, "Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for the Assessment of Educational Outcomes" in *Educational Policy Analysis Archives* 3, no. 6; Sanders, Saxton, and Horn, "Tennessee Value-Added Assessment System."

[5]University of Tennessee Value-Added Research and Assessment Center, *Graphical Summary of Educational Findings from the Tennessee Value-Added Assessment System (TVAAS) 1995* (Knoxville: University of Tennessee Value-Added Research and Assessment Center, 1995).

[6]Jordan, Mendro, and Weerasinghe, "Teacher Effects"; William L. Sanders and June C. Rivers, "Cumulative and Residual Effects of Teachers on Future Student Academic Achievement: Research Progress Report" (Knoxville: University of Tennessee Value-Added Research and Assessment Center, 1996).

[7]June C. Rivers-Sanders, "The Impact of Teacher Effect on Student Math Competency Achievement" (Ed.D. diss., University of Tennessee, 1999); Sanders and Rivers, *Cumulative and Residual Effects of Teachers.*

- When a student has experienced an ineffective teacher or a series of ineffective teachers, there is little evidence of a compensatory effect provided by experiencing more effective ones in later years.[8]
- Regardless of ethnicity, children of similar previous achievement levels tend to respond similarly to an individual teacher.[9]
- Within two Tennessee metropolitan districts, children of color were overrepresented in less effective teachers' classrooms by about 10 percent and underrepresented in highly effective teachers' classrooms by about 10 percent.[10]
- Teachers who are relatively ineffective tend to be ineffective with all student subgroups across the prior achievement spectrum, whereas teachers who are highly effective tend to be very effective with all student subgroups across the same spectrum.[11]
- The effect of the teacher far overshadows classroom variables, such as previous achievement level of students, class size as it is currently operationalized, heterogeneity of students, and the ethnic and socioeconomic makeup of the classroom.[12]
- In the extreme, for students scoring in the lowest quartile in fourth-grade math, the probability of passing an eighth-grade-level test (required for high school graduation) ranged from 15 to 60 percent as a function of the sequence of teachers and how effective they were. Students in this achievement group experiencing four teachers of average effectiveness had a 38 percent probability for passing the test.[13]

---

[8]Sanders and Rivers, *Cumulative and Residual Effects of Teachers.*

[9]Ibid.

[10]Ibid.

[11]Ibid.

[12]S. P. Wright, S. P. Horn, and William L. Sanders, "Teachers and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation" in *Journal of Personnel Evaluation in Education* 11, no. 1, 57–67.

[13]June C. Rivers-Sanders, "Impact of Teacher Effect on Math Achievement."

- In the extreme, students testing between the 25th percentile and the 50th percentile in the fourth grade who also experienced a series of highly effective teachers in grades five through eight could be expected to pass the high-stakes test with a probability of about 80 percent; their peers of comparable previous achievement unfortunate enough to have experienced four very ineffective teachers in the same grades could be expected to pass the same test with a probability of about 40 percent. A sequence of four teachers of average effectiveness offered students within this prior achievement level a probability of passing of about 60 percent.[14]

DISCUSSION

The cumulative and residual effects of teachers on the academic progress of students are huge. Collectively, these research findings paint a vivid picture of the extreme variability of teachers' effectiveness and the dramatic effect this variability has on student progress. The starkness of this picture suggests three critical areas of policy implication if school districts are to provide *equity in education:*

1. School districts must have a measurement system that allows them to monitor the variability among schools and teachers and an adequate means of communicating the measurement.
2. School districts must shrink the variability in effectiveness among existing teachers.
3. When school districts assign students to teachers, districts must minimize possible teacher effectiveness inequities for students.

Measuring Schooling Effects

Within their budgetary constraints, policy makers must determine the most effective way to provide robust and unbi-

[14]Ibid.

ased measurement. Unless educators receive meaningful measurement, they will continue to disavow measures of effectiveness that focus only on such convoluted and biased measures of student attainment as ethnicity, socioeconomic status, and parental influences. In the absence of unbiased measures of student progress, the disgruntled voice of all teachers, inappropriately evaluated, will continue to protect those teachers who are truly ineffective, making it very difficult to assist the latter in improving their classroom practices.

Based on what has been learned about teacher effectiveness and teacher effectiveness sequences, measurement and related reporting should serve two purposes, each requiring varying degrees of conservatism—accountability and diagnosis. Sophisticated statistical and computing techniques allow an integrated delivery of the two from the same testing regime. Much debate focuses on the types of testing instruments to be deployed. However, it appears that having measures of student achievement annually is more crucial than using a particular type of instrument,[15] provided that the three previously mentioned conditions are met (strong repeatability, high correlation between the test and curricular objectives, and sufficient stretch for discriminating on both ends of the achievement spectrum). Local districts have a range of choices in obtaining the desired test data, but economic constraints may determine what is locally optimal. Consider, for instance, this scenario: testing occurs *only* in grades three and eight in mathematics; students at the eighth-grade level do not reach the attainment level mandated by the local board of education. It is highly improbable that schools and teachers would be able to accurately assess where between grades four and eight the incidence or incidences of inadequate progress occurred. This testing regime that includes only data from grades three and eight does not have the diagnostic sensitivity of an annual testing program.

Reporting by TVAAS is an example of integrated delivery of accountability and diagnostic information from the same testing regime. Tennessee began testing students in grades

[15]Sanders and Horn, "Educational Assessment Reassessed."

two through eight in math, reading, language arts, science, and social studies in 1991. Data from these administrations; writing assessment data for grades four, seven, and eleven; and high school end-of-course math tests are analyzed to produce district, school, and teacher reporting. Multiple indicators for each student increase the sensitivity provided from any one measurement.

The official TVAAS reporting for accountability purposes provides districts and schools with three years of estimates of means and gain for each subject and grade served by the district and school, and this reporting provides teachers with three years of estimated student gain for each of their subject-grade combinations. The three years of data are averaged to provide the progress indicator used for accountability. Additionally, the district and school reporting provides an estimate of the percentage of progress the average student would be expected to make across grades when compared to the average student from the national norming sample. This estimate of percent of cumulative norm gain is the progress indicator for determining the effectiveness or ineffectiveness of either a school or a district. The minimal expectation for districts and schools is that enrolled children make at least enough progress to maintain their position relevant to their peers of comparable prior achievement. A school with a 100 percent cumulative norm gain is meeting this minimal target. Students in a 120th percentile school are progressing at a rate that will put them at an attainment level one grade beyond that of their peers of comparable prior achievement, when the 120th percentage rate is maintained across grades three through eight.

Additionally, a less conservative reporting series for districts and schools provides simple paired mean gains for various prior achievement subgroups. For this report, students are placed in prior achievement groupings based on their standing within the district. A student's standing is determined by averaging the two most recent years of data for a particular subject. The mean gains for each subgroup are

provided. This report is especially useful in monitoring the availability of equitable learning opportunities for students with various levels of prior achievement because the disaggregation by prior achievement allows the user to compare the subgroup mean gain to the target gain. Subgroup means below the target reflect less than adequate progress, whereas subgroup means at or above the target imply that students are progressing at more desirable rates.

Shrinking Teacher Variability

Another concern of policy makers is to determine what is needed in order to shrink the variability in teacher effectiveness. Decreasing the variability will help ensure that fewer children have an ineffective teacher. Ineffective teachers cause learning consequences for students that are compounded when the frequency of ineffective teaching increases. Although *most* students do not experience teacher quality sequences at either of the extremes, many children, unfortunately, lack the benefit of effective teaching at some point in their K–12 years. The effect is insidious, causing underachievement each year they encounter an ineffective teacher until the cumulative effect becomes extremely visible in later grades. For example, some students complete algebra in the eighth grade, whereas others of similar prior achievement levels in early grades struggle with anything beyond the most remedial mathematics at the conclusion of their elementary years.

Policies for decreasing the variability of teacher effectiveness must address two areas: (1) identification through measurement and (2) professional development. Many teachers do not recognize that they are ineffective until confronted with the objective evidence that their students are not making appropriate rates of gain. Once they recognize their students' lack of progress, many tend to self-correct their teaching practices. Those with pedagogical weaknesses and others may require assistance through strategically

planned professional development in order to learn to teach more effectively.

Minimizing the Impact of Ineffective Teachers

The existing policies for placing students with teachers deserve serious study to ensure that various subpopulations of prior achievement are not being subjected to systematic inequity across grades because they are assigned systematically to less effective teachers. For example, where do beginning teachers typically get assigned within a district? In larger districts, beginning teachers generally begin their employment in inner-city schools and seek transfers in assignment when they gain experience. Current TVAAS research suggests that teachers' effectiveness increases dramatically each year during the first ten years of teaching. If beginning teachers are disproportionately assigned to a school within a district, the children who have these teachers quite possibly are not receiving an opportunity to get a good education. Without some measurement of teacher effectiveness, this situation may be difficult to address.

Even though schools may be assisting less effective teachers to become more effective, principals should make a conscientious effort to avoid assigning students to multiple ineffective teachers in succession. Students unfortunate enough to encounter two or more ineffective teachers in sequence show measurably retarded academic growth. The effects on students' attitudes from having several ineffective teachers in a row have not been quantified; we need more research in this area. All children deserve to have highly effective teachers every year, but until something can be done to shrink the variability, no child deserves to experience two very ineffective teachers in a row.

CONCLUSION

Improving teacher quality is the mutual responsibility of educators and policy makers. Sophisticated measurement of teacher effectiveness is critical to this process because it ensures

that teachers are evaluated fairly and provides diagnostic information for improving teacher effectiveness. When teachers understand that they are evaluated on their ability to facilitate progress for their students, their perception of the fairness of this evaluation should increase the likelihood that they will explore and implement ways to improve their teaching practices. And the fairness issue aside, the sensitivity of more sophisticated measurement provides better diagnostic information on which to base programmatic decisions. It brings a focus to these efforts that less sensitive measures fail to provide. With this more finely tuned metric, teachers and administrators are in a better position to ensure that all students have appropriate learning opportunities. Improving teacher quality will help ensure that more students reach their potential because they benefited from effective teachers *every* year.