

Multiple Measures of Teacher Effectiveness in Hillsborough County Public Schools: Implementing Value-added Measures

Hillsborough County Public Schools (HCPS) was a leader in designing, implementing, and managing a comprehensive value-added measurement system to determine its teachers' individual impact on student achievement. Every one of the Tampa, FL-based district's 13,000 K-12 teachers was evaluated using value-added measures – statistical tools that measure the quantity of student learning attributable to a particular teacher. Art, physical education, career and technical, and special education teachers were all included in the student achievement measure, which made up 40 percent of a teacher's annual evaluation. Implementing value-added measures on such a large scale required more than just finding reliable and valid pre- and post-measures of student learning. The new system forced HCPS to strengthen its data management and security systems, communicate a complicated mathematical concept in a clear and compelling manner, and accept some future uncertainty in the link between teacher behaviors and student performance. Implementation was ongoing and there were many more lessons to be learned. But, as the 2011-2012 school year started, all K-12 teachers in the district would know how much learning they were personally responsible for in their students, which would be reflected in their performance evaluation.¹ That knowledge, combined with peer and principal observations offered powerful insights into how to improve instruction.

What Are Value-added Measures?

The search for a teacher's unique contribution to a student's learning has been elusive. The challenge for educational leaders has been to account for the effects individual characteristics and varying experiences have on students' learning, while also measuring growth in a specific subject or content area over a distinct time period. Value-added measures offer a potential solution to the problems inherent in measuring individual teacher performance. Many states and school districts have incorporated value-added measures as components of teacher evaluations. Value-added measures are used statewide in Tennessee, Pennsylvania, and Ohio. Hundreds of school districts in 21 states also use value-added assessment including Chicago, Dallas, Milwaukee, New York City, Pittsburgh, and Seattle.² Each value-added model was different, but all incorporated similar core components.

Pioneered in education by William Sanders in his study of teacher effectiveness using the Tennessee Value-Added Assessment System, value-added measures have become prevalent in educational research, program evaluation, and teacher evaluation.³ Researchers such as Thomas Kane at Harvard, Rob Meyer at the University of Wisconsin, and Douglas Staiger at Dartmouth have applied value-added methods to analyze the effects of different types of teacher certification⁴ (traditional vs. alternative certification) and schools⁵ (charter or pilot vs. regular public schools). Applied to the teacher evaluation, value-added measures combine students' adjusted pre- and post-test scores to estimate the learning growth over a specific time period.

The value-added calculation produced a score for every teacher that compared their students' growth to that of a typical teacher.⁶ Scores ranged from negative to positive; negative scores indicated that the teachers' students learned at a slower rate than an average grade or subject teacher in the district. Positive scores suggested students had faster growth. Specific student, classroom, and school characteristics were used as "controls" in the value-added model. That way, teachers working with different groups of students in different types of schools could be compared with each other. For example, student-level controls might include English fluency level, parental education, gender, and absenteeism rate. Classroom size, peer performance, and school demographics such as percent minority and free or reduced-price lunch were other characteristics typically controlled for in value-added measurements.

Value-added measures had generated some controversy as they became more broadly used to allocate pay to teachers. Some researchers critiqued the theoretical assumptions underlying value-added models,⁷ while others questioned the effects of linking compensation to student achievement.⁸ At the same time, there was evidence that value-added measures correlated to other measures of teacher quality, such as student perceptions of teacher effectiveness and specific pedagogical practices.⁹ In addition, principal perceptions of teacher effectiveness and teachers' value-added scores were positively correlated.¹⁰ In general, educational leaders and researchers agree that value-added measures can be a valuable measure of teacher effectiveness when used appropriately. However, most caution against using it as the sole factor in determining teacher performance; rather, it should be used as one component of comprehensive teacher evaluation system that incorporates other forms of data.¹¹

Laying the Groundwork with MAP and End-of-Course Exams

Linking student achievement to teacher evaluation was not a new concept for teachers in HCPS. Following a Florida statute, beginning with the 2002-2003 school year, HCPS offered a voluntary performance bonus plan based partly on student achievement gains. Teachers had the option to earn a five percent supplement in addition to their regular salary by demonstrating outstanding performance in 12 teaching practices.¹² While significant student learning gains were the primary criterion to obtain an outstanding rating, the documentation and description of the gains were left to the teachers' discretion. Implementing the program had also been a challenge. It was unfunded by the state, so the amount of money available for bonuses varied year-to-year. In addition, the vaguely-defined rating system led to many more teachers being rated outstanding than the initial target of 10 percent. By 2006, Florida crafted a more ambitious program that linked teacher bonuses to performance called the Special Teachers are Rewarded (STAR). However, STAR was soon eclipsed six months later by the Merit Award Program.

HCPS was one of the very first to sign on to Florida's Merit Award Program (MAP), which started in the 2006-2007 school year. The state offered top teachers bonuses of five percent of the average annual salary in the district. Student achievement growth had to account for at least 60 percent in determining the high performing teachers to receive bonuses. Superintendent MaryEllen Elia partnered with the teachers' union, the Hillsborough Classroom Teachers Association (HCTA) to build a MAP-approved system that included nearly all teachers. It used student achievement data for 60 percent and principal observations for 40 percent of the effectiveness rating. Teachers were placed into instructional-based groups and ranked based on their overall evaluation ratings. Generally, those falling above a designated percentile within their instructional group were awarded bonuses.

Implementing Value-added Measures

In its first year of implementation around 2,740 HCPS teachers received \$2,100 bonuses, representing five percent of the average teacher salary in the district.¹³ However, three-fourths of the teachers receiving bonuses taught in high-income schools.¹⁴ The student growth measurement used in the evaluation had not adequately accounted for between school differences. Undeterred by the setback, the district, with HCTA's support split the teacher comparison groups into those teaching in low-income and high-income schools using Title I designation as a criterion. The result was over fifty different teacher comparison groups based on instructional categories (e.g. grade level or subject area taught) and Title I status. In the following years, teachers were ranked within each of the comparison groups to determine bonus-eligibility.

Constructing and validating the hundreds of pre- and post-test instruments necessary to conduct value-added calculations could have been overwhelming. Fortunately, HCPS also had a head-start on this aspect of incorporating student achievement in teacher evaluations. It had pioneered end-of-the-course exams for all core subjects in the 1980s. Through the years, HCPS built on this early work by including elective subjects and elementary grades. Beginning-of-course tests were soon added in nearly all courses. Importantly, the test questions and format for non-core subjects were developed by teachers. The department of assessment then evaluated each test's reliability and item validity. Invalid test items were removed each year and content area teachers designed new items.

In the process of implementing the MAP, HCPS designed a system to link teachers to the individual students they taught in particular courses. Logistically, this was one of the most difficult challenges in soundly implementing value-added measures. With MAP, HCPS needed to link every class a student attended to the student's pre- and post-test scores and his or her class teacher. The district accomplished this first by assigning every student, course, and teacher a unique identification number. A database was then constructed that linked each of these numbers with specific test score and student demographic data. The resulting data system set the stage for a comprehensive integration of student achievement data and teacher evaluation.

Building a Robust Value-add Measure

When HCPS started looking into using multiple measures of teacher effectiveness, it was already well-positioned to incorporate a value-add measure as one performance metric. The district had in place pre- and post-tests for most subject areas and all teachers, students, and courses had unique identification numbers, which could then be used to link individual students to teachers. Perhaps more importantly, teachers were familiar with the idea of connecting student learning growth to evaluations. At the same time, there were problems with the MAP student achievement bonus system. Teachers were skeptical of the "black box" nature of the MAP calculations and wanted to know the specific students and scores used in the calculations.¹⁵ In addition, MAP relied on single pre- and post-tests to determine a teacher's individual contribution to student learning. There were also question raised about test security and data management.

One of the first steps HCPS took was to find an external partner with the capacity and expertise to conduct large scale value-added measures for over 13,000 teachers. With input from the teachers' union, the district eventually selected the University of Wisconsin's Value-Added Research Center (VARC) to design and calculate the value-added measures. VARC had extensive experience designing and implementing value-added measures in large school districts such as Chicago Public Schools and Milwaukee Public Schools. In addition, the director of VARC, Rob Meyer, had been working with states and districts for more than 15 years on applying statistical methods to measure teacher impact. Finally, VARC was one of the only groups that agreed to consider all teachers in

value-added calculations, not just those in core content areas. As Anna Brown, director for assessment and performance management, explained:

One of the biggest critical criteria for us was someone who was willing to branch out in to the inclusion of all subject areas, the inclusion of all courses that we measure. There were some organizations that absolutely would not go beyond core subjects of reading and math state-tested data. That's a critical element for us because we had to have a measure for everyone because we've had that kind of measure for the last four years.

Working with VARC, HCPS constructed a course testing map that described every pre- and post-measure used for all courses in the district. These included state tests, the PSAT, AP exams, IB exams, technical certification tests, and district-developed assessments. Data from each of these were then tabulated and transferred to an electronic data warehouse.

From there, the assessment data could be tied directly to individual students and teachers. A strong link between teacher, course, and students was especially important. Furthermore, it was necessary to conduct a close analysis of what courses were connected to a teacher. The class load for teachers needed to be balanced with courses that had accurate data measures. For example, administering in-school suspension was a duty for some teachers that replaced a course. If a teacher was assigned to teach in-school suspensions all day long, there would not be accurate data associated with that teacher. Thus, implementing value-added measures with fidelity forced HCPS to reexamine some of its methods for assigning courses to teachers.

While assessments existed for most of the subjects and courses in HCPS, one of the biggest challenges has been developing objective tests that can be scanned into the database and used multiple times. Another issue has been how to treat small-sample sized courses such as special dropout prevention programs or specialized classes like diesel mechanics. To address these and other difficulties, HCPS relied on teachers to devise solutions. Brown commented on the teachers' role in the process:

Teachers are involved at every step in our effort. It's teachers who help us find some of these unique things that are happening, but definitely we have teachers heavily involved in our exam writing process. We have teachers involved in our exam review process. We have teachers involved once we do item analysis and statistics on every exam in every administration. Then we have to rewrite exams, or eliminate items that have been proven to be a poor item, et cetera. So they are involved, and then also we're involving teachers as we're looking at how these unique situations impact our ability to give them an accurate effectiveness measure.

One Component of a Coherent Evaluation System

The value-added measure played a very important role in the new evaluation system. It was the single largest contributor determining teacher performance in HCPS. At the same time, the 40 percent contribution was still less than the 60 percent student achievement growth had played in the MAP bonus eligibility. Also, in contrast to MAP, three years of student results would eventually be used to determine a teacher's value-added score. This ensured greater reliability as an abnormal fluctuation in student achievement one year would be counterbalanced by two years of data. But, the most important feature of the value-added measure was that it was integrated into a multiple-measure evaluation system that incorporated peer and principal observations.

In this way, value-added measures in HCPS were more than a static score of a teacher's performance. Teachers could evaluate year-to-year whether a particular instructional approach or

Implementing Value-added Measures

curriculum was effective for their classes as a whole and specific subgroups of students. At a broader level, schools and the district could use aggregated value-added measures to determine the efficacy of a particular program, curriculum, or intervention. In effect, value-added measures were seen as more than a ruler used to measure teacher performance; they contributed to the development of better curricular, scheduling, and instructional decision-making. The value-added measures were also integrated into facets of school leadership in HCPS. Five percent of a principal's evaluation was based on the degree to which his or her own teacher evaluations aligned with the value-added measures. In sum, the integrated value-added measures in HCPS offer another tool for creating the best possible instructional environment for student learning.

Endnotes

- ¹ Value-add measures for PreK and adult education teachers would be rolled out over the following years.
- ² Center for Greater Philadelphia, University of Pennsylvania, "Value-Added Assessment." Retrieved January 2011 from http://www.cgp.upenn.edu/ope_value.html.
- ³ William Sanders and S Horn, "Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research." *Journal of Personnel Evaluation in Education*, 12(3), 2008: 247–256.
- ⁴ Thomas Kane, Jonah Rockoff, and Douglas Staiger "What Does Certification Tell Us about Teacher Effectiveness?: Evidence from New York City" *Economics of Education Review* 27(6), 2008: 615-31.
- ⁵ Atila Abdulkadiroglu, "Informing the Debate: Comparing Boston's Charter, Pilot, and Traditional Schools," Boston Foundation, 2009. Retrieved January 2011 from http://www.tbf.org/uploadedFiles/tbforg/Utility_Navigation/Multimedia_Library/Reports/InformingTheDebate_Final.pdf
- ⁶ Henry Braun, "Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models," ETS Policy Information Perspective, 2005. Retrieved January 2011 from <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>
- ⁷ Jesse Rothstein, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *The Quarterly Journal of Economics*, MIT Press, vol. 125(1), February 2010: 175-214.
- ⁸ Steven Glazerman and Allison Seifullah, "An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year Two Impact," *Mathematica Policy Research* 6319-520, May 17, 2010. Retrieved January 2011 from http://www.mathematica-mpr.com/publications/pdfs/education/tap_yr2_rpt.pdf; Matt Spring et al., "Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching," National Center on Performance Incentives, September 21, 2010. Retrieved January 2011 from http://www.performanceincentives.org/data/files/pages/POINT%20REPORT_9.21.10.pdf
- ⁹ Thomas Kane et al., "Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project," MET Project Research Paper, Bill & Melinda Gates Foundation.
- ¹⁰ Jonah Rockoff, Douglas Staiger, Thomas Kane, and Eric Taylor, "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools," Working Paper, July 2010. Retrieved January 2011 from <http://www0.gsb.columbia.edu/faculty/jrockoff/papers/Information%20and%20Evaluation%20RSKT%20July%202010.pdf>
- ¹¹ Daniel McCaffrey, J. R. Lockwood, Daniel M. Koretz, and Laura S. Hamilton, "Evaluating Value-Added Models for Teacher Accountability," Report, RAND., 2003. Retrieved January 2011 from www.rand.org/pubs/monographs/2004/RAND_MG158.pdf; Henry Braun, "Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models," ETS Policy Information Perspective, 2005. Retrieved January 2011 from <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>
- ¹² Hillsborough Classroom Teachers Association and School Board of Hillsborough County Contract, August 4, 2005. Retrieved January 2011 from <http://www.nctq.org/contracts/8-05.pdf>.
- ¹³ Letitia Stein, "Q&A: The Merit Award Program," *St. Petersburg Times*, February 24, 2008. Retrieved January 2011 from http://www.sptimes.com/2008/02/24/Hillsborough/QA__The_Merit_Award_P.shtml
- ¹⁴ Letitia Stein, "Hillsborough's Merit Pay Experiment Benefits Affluent Schools," *St. Petersburg Times*, February 24, 2008. Retrieved January 2011 from http://www.sptimes.com/2008/02/24/Hillsborough/Hillsborough_s_merit_.shtml

Implementing Value-added Measures

¹⁵ Hillsborough County Public Schools, "Empowering Effective Teachers USP Progress Report Year 1," 2010.