

Goals and Aims of Value-Added Modeling: A Chicago Perspective

John Q. Easton

Consortium on Chicago School Research at the University of Chicago

October 2008

The goal of this paper is to reflect on several questions about the goals and uses of value-added modeling from my vantage as an external researcher and active commentator and participant in the school reform process in the Chicago Public Schools. Though there are probably many factors that distinguish CPS from other large school districts and state education agencies, one factor that CPS has in common with its peers is a great interest in VAM accompanied by perhaps a naively wishful and hopeful view that VAM will answer many questions and solve multiple problems facing the district. I want to look at some of these issues in this paper, and as best as I can lay out what see as current thinking in Chicago on how VAM can contribute to school reform and school improvement. Perhaps this can be considered an incomplete “case study” of VAM in Chicago and will provide some insights that may be applicable elsewhere.¹

Though CPS is governed by the mayor of Chicago, who appoints the CEO and the board members, there are still a wide range of stakeholders with strong influence over CPS policy and practice. At the very least, these include foundations (both local and national), the Chicago Teachers Union, external partners and service providers including universities, and the education research community. Each of these constituencies has

¹ I should acknowledge right away the possibility (likelihood) of my personal biases entering into this description. I am not fully aware of all conversations about VAM methodologies across Chicago, have my own beliefs and values, and am heavily identified with my organization and institution, which has been involved in this topic.

articulated a voice on VAM in Chicago in the past few years, sometimes in response to a request from the school district and sometimes unsolicited. Since he became CEO in 2001, Arne Duncan has been eager to develop and employ VAM (or similar gains-based methods) in CPS, given his belief that these will provide a fairer and more accurate measure of student learning than traditional metrics like those used for judging AYP under No Child Left Behind.

CPS has some experience and history with VAM and gains-based measures of student achievement. Throughout the 1990's Tony Bryk lead a team of researchers here at the Consortium on Chicago School Research in developing a longitudinal gains-based method of measuring elementary school "academic productivity."² This work resulted in a 1998 study showing trends in math and reading productivity across the city and sparked a great deal of discussion about the need for fair and accurate measures of school improvement at a time when the district was solely focused on the "percent of students at or above grade level" as a measure of academic strength and uninterested in other methods. Because of this strong test-based accountability focus and "keep it simple" approach within the school district the value added approach gained little traction inside the system leadership.

More recently and with more receptive school leadership, Bryk and Stephn Ponisciak (with assistance from Steve Raudenbush) developed a second, more advanced VAM methodology and analyzed CPS elementary test score data during the early

² Bryk, Anthony S., Yeow Meng Thum, John Q. Easton, and Stuart Luppescu (1998). Academic Productivity of Chicago Public Elementary Schools. Chicago: Consortium on Chicago School Research at the University of Chicago.

2000's.³ School-level VAM estimates from this methodology were provided to CPS but never used—for reasons discussed subsequently in this paper. Currently, with significant support from local foundations, CPS has engaged the Value-Added Research Center (VARC) at the University of Wisconsin, directed by Rob Meyer, to create a sophisticated value-added measurement system for elementary schools, ultimately creating not only school-level and grade level VAM estimates, but also classroom and teacher level estimates.⁴ There are very high expectations for this work among many prospective users both within and outside of the school district. We will look at some of these expectations later in this paper.

There are also many less visible applications of VAM methods in Chicago by the research and evaluation community. We at CCSR use a variety of VAM procedures to develop outcome measures for our research. Long-term research of ours on the “five essential supports for school improvement” is predicated on the academic productivity measure mentioned above and has been replicated using the more recent VAM procedures. Our studies show the positive impact of strengths in these school organizational capacities on long-term improvements in student learning⁵. As a result of this work, CPS has adopted these essential supports as their “five fundamentals for school success.” Using other VAM-like measures with high school test scores from the EPAS system from ACT, Consortium researcher Elaine Allensworth recently studied test prep activities in Chicago high schools and found that test prep is negatively related to ACT

³ Ponisciak, Stephen and Anthony Bryk (2005). Value Added Analysis of the Chicago Public Schools: An Application of Hierarchical Models. In Robert Lissitz, (Ed.), *Value Added Models in Education: Theory and Applications*. Maple Grove, MN: JAM Press.

⁴ See <http://varc.wceruw.org/>

⁵ See http://www.stratplan.cps.k12.il.us/school_success.shtml

scores.⁶ These few examples are not meant to catalog the current uses of VAM in research on Chicago schools, but rather to suggest that up to now, VAM methods have been more productive in research applications than in more direct applications around school district use for accountability or school improvement. The remainder of the paper is about exploring this assertion.

In developing the specifications for the VAM approach to be used in Chicago, the VARC team consulted widely across stakeholders about the most appropriate variables to include in their statistical models. They considered all of the usual suspects, including prior achievement, race, gender, special education status, language status, school mobility, and economic status as measured by eligibility for free or reduced cost lunch.⁷ Following wide discussion within the school district leadership and with an external advisory group (of which I am a member), it was decided not to use student race in the school-level statistical models. The reasoning behind this decision is well known – that we should have uniform expectations for all students and not make comparisons solely within racial and ethnic groups but across them. The practical implications are clear though. Predominantly African American schools will have uniformly lower value-added estimates than schools and classrooms with students of other races/ethnicities because of the consistent achievement gaps across the school district. As the analysis moves to the teacher/classroom level in the upcoming years, we will see the same effect – a systematic bias against predominantly African American classrooms.

⁶ Allensworth, Elaine, Macarena Correa, and Steve Ponisciak (2008). *From High School to the Future: ACT Preparation—Too Much, Too Late* Chicago: Consortium on Chicago School Research at the University of Chicago.

⁷ See <http://research.cps.k12.il.us/cps/accountweb/Research/ValueAdded.html>

The statistical models do include the free/reduced lunch indicator, however, and one can argue that this picks up poverty, which is the real culprit behind the achievement gap. Unfortunately, FRL is such a crude indicator that it provide little information, especially since CPS is about 85 percent FRL. Given the depth of urban poverty, the bar for FRL is set relatively high, so it doesn't differentiate among degrees of disadvantage. Broadly speaking, FRL combines the "working poor" with the "truly disadvantaged" in CPS, and the former group is more likely to include Latino students and schools and the latter group is more likely to include African American students and schools. This issue, not widely understood or appreciated, has had negative repercussions for judging Chicago schools in the past.⁸

I noted earlier that expectations are high in Chicago about the potential uses for the new VAM results from VARC. These run across the multiple stakeholders groups in Chicago, with probably the greatest hopes from the foundation community that is supporting the work and from other external partners working with CPS. These hopes are probably also for the most contentious uses of the VAM results, especially around evaluating teachers. The problems here are enormous, and not the least of them is a basic technical problem of accurately matching teachers to their students. Until recently, CPS did not have the technological capacity to make this match. Though the student

⁸ One quick example – the Broad Foundation analyzes district-wide student achievement results across the nation to determine eligibility for their prestigious Broad awards. Using FRL as a control variable, they compare achievement test results among school districts in the same state and ask whether the test scores are high or lower than those in "similar" districts. CPS does not compare well against other Illinois school districts with comparable FRLs because they are all predominantly Latino districts, and CPS has much higher African American enrollment. A more discriminating measure of economic status would show CPS to be much poorer—and have its achievement test scores look comparatively better. When CPS reading test scores are analyzed by race and compared across the state, CPS students outperform their counterparts in all racial/ethnic groups in the higher elementary grades (see Easton, John Q., Stuart Luppescu and Todd Rosenkranz (2007). 2006 ISAT Reading and Math Scores in Chicago and the Rest of the State. Chicago: Consortium on Chicago School Research at the University of Chicago.)

information system grouped students to classrooms, there was no corresponding technology to link teachers to classrooms and students. CPS has both a brand new student information system and a new teacher personnel system that were both intended to capture the link between students and teachers. Now in the early implementation phase of these new systems, though, it is unclear that the systems are accurately capturing this information. In one pilot study that I am involved in as an external evaluator, CPS is manually collecting school-based rosters of students and teachers because of doubts about the new technology. Though this is merely a technical problem, it is a formidable one. Erroneous attributions of students to teachers could severely undermine this effort.⁹

Though probably not yet a pervasive view, there are certainly voices in Chicago that look forward to seeing the VAM measures used to reward teacher performance, to remove underperforming teachers, and to make tenure decisions. With a substantial federal grant CPS conducts a Teacher Advancement Program (Chicago TAP), which in its early years is using VAM only at the school level to reward all staff for positive results. In my view, CPS has a very long way to go before they can credibly use teacher level VAM estimates for any of the purposes noted here. This will be an especially great problem if race isn't specifically included in the statistical models. Teachers of poor African American students will be unlikely to attain high value-added results. This problem is somewhat related to the selection effects discussed in the VAM literature.

⁹ An even greater problem for this entire endeavor is the basic **accuracy** of the test scores themselves. The state testing system for elementary grades, the Illinois Standards Achievement Test (ISAT), has had numerous recent problems related to scoring and equating of forms. The 2008 tests were rescored (after re-equating) when districts complained that scores were too low. Many question the credibility of these scores. The Illinois State Board of Education is seriously under-resourced and does not have rigorous quality control procedures in place around its student assessment program.

Because teachers are not randomly assigned to either schools or students it is difficult to disentangle the teacher and student effects.

VAM results are also being proposed as part of the CPS school accountability system that provides both rewards and sanctions for school performance. In the early stages of the new value-added system, these results will carry relatively little weight in the accountability system, but presumably with additional experiences and over time, they may carry greater weightings. There have been times in the past when gains-based measures have been part of the school accountability system, so these ideas are not new to CPS. This is the first time, however, that a formal VAM metric will be used.

I wonder how schools will respond to a metric that they don't understand and what it will mean to include such a complex metric in the accountability system. CPS has a long history of holding schools accountable for "the percent of students at or above the national norms" (now the percent of students who meet or exceed state standards) on standardized achievement tests. It was widely accepted in CPS that one strategy to move this indicator was to concentrate on the "bubble kids" – those just below the cut score. There is no reason to believe that this doesn't still occur under NCLB. This is not the kind of behavior envisioned when the indicator was first included in the accountability system as a broad measure of student achievement. But it is an obvious strategy to game the system to move the indicator but not what the indicator was intended to measure (percent above the cut scores vs. improved students learning). But what is the strategy to move an indicator that isn't well understood? The indicators and the accountability system they are part of should be promoting stronger instructional practices that will benefit all students as a VAM indicator measures. We have many examples from the past

where the indicators create incentives for bad practice as a result of gaming of the system. How will people respond to an indicator that they don't fully understand?

CPS had access to complete school level value-added achievement estimates several years ago and made little use of them in spite of the interest and demand for them. Although there were probably multiple reasons that I am not aware of why CPS did not use this information it appeared to me that the problem was lack of internal capacity to use it profitably. Such profitable use puts a large demand on staff and leadership to fully understand the outcome data largely by careful, thorough, and repeated analysis and “interrogation.” This requires understanding patterns and trends in the VAM outcomes and the relationships among VAM outcomes and other test score metric (average scale scores, percentage of students meeting state standards, simple one year gains, etc.). Data analysts routinely do this work but building knowledge and coherence about how the VAM indicators behave is more demanding and requires considerable thought and attention. This also requires a set of communication skills that are rarer still. The school district and broader research and reform community must explain to a broad constituency—especially to schools and their leaders and teachers – what the VAM indicators are saying about school performance and how and why the information is better or different. CPS needs to build greater capacity in this arena in order to fully exploit the inherent possibilities within VAM indicators.

I have noted either interest in or intention to use VAM metrics in CPS for teacher and school evaluation and accountability purposes and raised a few red flags about what I see as current potentially problematic issues. On the other hand, I see much greater upsides in using VAM metrics in school research and evaluation studies and in fact think

that these are essential first steps in preparing for wider acceptance and greater buy-in for using these same metrics for accountability purposes. The more widely circulated research using VAM metrics as outcomes there is, the more understanding there will be about they can be used most successfully and what their limits are. The research should help us understand more about the properties of the metrics, their distribution and variability across school and grades (and ultimately classrooms and teachers) and more importantly, about the validity of the VAM indicators. What is being proposed now places the new VAM indicators as the criterion in validation studies, when in fact the VAM indicators themselves need to be validated, especially in relation to other constructs and related measures. This kind of research and evaluation with VAM outcome metrics will help build frameworks to encourage good school-wide and classroom based practices. As this research builds coherence and credibility among the many stakeholders there will be greater civic capacity around the use of VAM metrics for multiple purposes. Finally, the greater use of VAM metrics in Chicago should and can lead to better policy and practice as knowledge accumulates on what leads to higher VAM measures. In sum, I think that for us to see great potential benefits of VAM measures that we need to solve all the technical problems around the indicators, we need to build evidence of the construct validity of the measure and communicate this research broadly. Then the VAM indicators can play a positive role in improving classrooms and schools.