10-1-2015

Value-Added Results for Public Virtual Schools in California

Richard Ford
California State University

Kerry Rice
Boise State University

Ford, R., & Rice, K. (2015). Value-Added Results for Public Virtual Schools in California. *Educational Technology & Society, 18* (4), 412–423.

Value-Added Results for Public Virtual Schools in California

Richard Ford[1] and Kerry Rice[2*]
[1]Department of Mathematics and Statistics, California State University, Chico, CA, USA // [2]Department of Educational Technology, Boise State University, Boise, ID, USA // RFord@csuchico.edu // krice@boisestate.edu
[*]Corresponding author

(Submitted September 4, 2014; Revised January 19, 2015; Accepted February 16, 2015)

ABSTRACT

The objective of this paper is to present value-added calculation methods that were applied to determine whether online schools performed at the same or different levels relative to standardized testing. This study includes information on how we approached our value added model development and the results for 32 online public high schools in California. Student level California Standards Test results in English Language Arts and Mathematics for over 5,000 online students were analyzed. Mean value added metrics for each school were calculated for 8 courses held during the 2010-2011 academic year. We found that schools of distinction existed in 7 of the 8 course categories.

Keywords

Online learning, Value-added, Virtual schools, K-12, English, Mathematics

Introduction

In 2000 approximately 45,000 K-12 students nationwide engaged in some type of formal online learning course or activity. By 2010 that number had grown to over 4 million (Staker, 2011). The accelerating growth in public online coursework at the K-12 levels elevates the importance of research into the efficiency and effectiveness of online education. Nationally we have a critical and pressing need to expand our knowledge base to facilitate the identification of what works best in online learning environments (Means et al., 2010).

The purpose of this study was to determine whether the virtual schools operating in California in the 2010-2011 academic year produced equivalent or different value added results on standardized tests in English language arts and mathematics. To be precise, we looked at eight specific courses, four in math and four in English language arts, offered at the 32 identified virtual schools whose student level test score data for successive years was provided by the California Department of Education (DOE). (The research was conducted independently by the authors and was not supported nor endorsed by the California DOE.) For this study, virtual schools were defined as those schools in which instruction was delivered entirely or primarily through online methods. The California DOE provides a service that identifies schools that deliver a minimum of 30% of content online. This threshold was too low for our purposes so further identification of schools for this study was accomplished through a comprehensive review of all listed charter school websites for information on their primary delivery method. In particular, we looked for schools whose names reflected some online or electronic component and schools that specifically designated themselves as online in their program descriptions. Given the challenges in defining and categorizing online schools, one limitation of the study is that the schools subsequently included in this report likely do not represent a complete sample of all online schools in California.

Method

The research objective was to identify, in each of the eight separate courses, schools that produced statistically superior value added metrics. The initial data pool consisted of all students who took a math or English language arts California Standards Test (CST) in the spring, 2011 at any one of the 32 identified public online schools. This initial pool was back-mapped by the California Department of Education to retrieve corresponding CST test scores for 2010, regardless of which public school generated the pretest score. Thus any student from the initial pool who was also tested anywhere in California in 2010 would remain in the pool. Those students for whom no pretest score could be retrieved were then eliminated from the pool. In English language arts approximately 82% of the initial pool was retained for the study. In mathematics approximately 77% of the initial pool was retained for the study. Each student record provided by the California DOE contained a scrambled student id number, school id, test and grade level information, and scaled scores for 2010 and 2011. The initial data set for the 32 virtual schools consisted of 5,666 records. Several records could not be used due to missing test scores or test id information in either English or mathematics. Students from cohorts of pretest-posttest pairings of fewer than 36 students were also excluded. The number of usable records for English totaled 5085. The number of usable records for mathematics was 4147. The mathematics number was significantly lower due to the fact that students taking the California summative exam were excluded from the mathematics portion of the study. The California mathematics summative exam is given to all students who have completed algebra 2 prior to the current academic year regardless of whether or not they are currently enrolled in any math class. Therefore the summative exam cannot be associated with any specific course. Four different courses in each subject area were studied for each of the 32 online schools. A determination of value added was calculated for each course at each school. For each course, the lower boundaries of one-tailed ninety-five percent confidence intervals were established for each school. Those schools whose lower bound of the confidence interval was above the overall mean became designated as “distinguished” for that course.

Overview of value added methods

The use of value added methods (VAMs) by schools, districts, and states, now dates back over twenty years. The Tennessee Value Added Assessment Model (TVAAS), a layered mixed effects model, developed by William Sanders and Robert McLean of the University of Tennessee, has been in use since 1991 (Sanders & Horn, 1994). Simply stated, value added methods are a way to measure changes in student performance over time. They continue to be strongly encouraged nationally and are even required for states to be competitive for Race To The Top funding (Corcoran, 2010). As a result there has been tremendous growth in the research base providing analysis of the benefits as well as drawbacks of these new methods. In general these methods are very complex and highly technical and there are concerns that they may be used inappropriately (Condie, Lefgren, & Sims, 2014). In large-scale studies, value added methods “have proved valuable for looking at a range of factors affecting achievement and measuring the effects of programs or interventions” (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012, p. 8). However, they may not be an accurate measure of teacher effectiveness given the wide variety of factors that can affect individual student performance. Indeed, even when applying the similar value-added techniques to the same data sets, different researchers can sometimes generate different results (Briggs & Dominigue, 2011). One example is a study conducted in 2004 in which a team of researchers led by Carmen Tekwe compared four similar, but different value-added approaches; hierarchical linear mixed models (HLM) with and without student covariates, layered mixed effects models (LMEM), and simple fixed effects models (SFEM). The team claimed to show that the results of the LMEM and SFEM models were different, but highly correlated and concluded the much simpler SFEM model was more desirable (Tekwe, Carter, Ma, Algina, Lucas, Roth, Ariet, Fisher, & Resnick, 2004). This claim was reviewed and disputed by other researchers who stated that the Tekwe et al. (2004) study “relied on a narrow data structure, which may have seriously limited its conclusions” (Doran & Fleischman, 2005, p. 85). The fact is that education data is influenced by an endless variety of factors and will always remain noisy, particularly at the teacher level, no matter how sophisticated and complex the method is. Still the search for ever better models that are fair, comprehensible, and provide reproducible results should certainly continue. As such, one of our long-term goals is to determine if the results from appropriately applied value added calculations could be relevant and reliable at a program level. By identifying those programs or schools that produce exceptional results would it then be possible to tease out the reasons why they were successful?

One of the simplest models we found was used by the United Kingdom to calculate value-added for their schools between 2000 and 2004. This UK method did not take into consideration the wide variety of student, school and other confounding characteristics, which might influence performance. Instead they assumed that, on average, those characteristics are randomly distributed in each of two distinct school classifications, “mainstream” and “special” schools. Rather than using linear regression, they establish a “natural median line” which consists of the set of all points, (x, y) where x is a particular pretest score range and y is the median of all posttests scores from students with x in the pretest range (DfES Analytical Services, 2004). The value added score for any student is the difference between their posttest score and the corresponding median score for their specific pretest value. The value added for a school is then calculated as the mean of the value added scores for all the students in the school, plus 1,000. Thus a school with a score of 995 is below average and one with a score of 1006 is above average. This system was very attractive to policy makers due to its simplicity. It is quite likely that use at the teacher level would not be particularly reliable due to the wide variety of student characteristics that do not distribute evenly at the classroom level. Since our goal was not to evaluate performance at the teacher level, but rather at the program or school level, we concluded that this level of analysis was similar to what we might need. We note that this model has since been replaced in the UK with a more sophisticated and complex model (Evans, 2008).

Complex model issues

The objective of ever more complex value-added models is to control for variables that contribute to student advancement that are unrelated to the teacher or school education inputs. The models we studied (Amrein-Beardsley & Collins, 2012; Atteberry, 2012; DfES Analytical Services, 2004; Evans, 2008; Goe, 2008; Isenberg & Hock, 2011; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Raudenbush, 2004; Sanders & Horn, 1994; Tekwe, et al., 2004; Value-Added Research Center, 2015; Wright, White, Sanders, & Rivers, 2010) typically employed some type of multiple linear regression to accomplish this control. One way to explain how the control is accomplished is in terms of expected posttest outcomes. Suppose in a large population of students it turns out that students with some specific characteristic, say left-handedness, produce generally greater growth from year to year in a particular subject. We recognize that teachers and schools have nothing to do with whether or not a student is left-handed and therefore wish to make our model fairer by controlling for that variable. In addition, we want our expected posttest prediction to be as accurate as possible, and knowing whether or not a student is left-handed would be information that should help. Linear regression that includes this factor essentially improves the prediction of posttest score by including the average effect differential between right and left-handedness. The end result, simply stated, is that the model will produce an *expected* score for a left-handed student that is appropriately more than one who is right-handed with all other factors equal. When value-added is defined as the residual, i.e. the difference between the *expected* and *actual* posttest scores, we see that the left-handed student would be assigned lower growth than if the factor were not included in the model. In this way when value-added scores are compiled for schools or teachers that have disproportionate numbers of right or left handed students, they will not be rewarded nor penalized for something unrelated to the educational input provided.

The example above helps to illuminate important education issues related to complex value-added models. The model by itself makes no attempt to explain *why* our lefties perform better. In fact, by including the factor in the model, motivation by schools, teachers, and administrators to study causal factors is reduced because they are not held accountable for that characteristic. It could be as simple as the fact that our student desks at which the tests are taken are all designed for lefties. We should want to study and mitigate the right-left discrepancies, but instead, accounting for it in our model removes the incentive because the lower relative growth for the right-handed group doesn’t lower the value-added calculation for the teacher or school.

Controlling for some factors may also encourage inappropriate adjustments. Suppose we discover that students with tattoos tend to do more poorly on average than the general population. We decide to control for that variable resulting in adjusting value-added scores slightly higher for those students with tattoos. It’s certainly a bit silly, but perhaps a serious administrator decides that it will help his school’s overall value-added score by asking all his students to get tattoos. The example is far-fetched, but currently many models control for free and reduced-price lunch. This means that if the school could qualify more of its existing students, value-added scores would rise slightly only due to the mathematical calculation adjustments.

These issues surrounding ever more complex value-added models need to be understood and discussed by educators and political leaders. Controlling for variables unrelated to education input can and likely will result in unintended consequences. In our choice of a simpler method, we are making the case that we do not want nor were we able, given the data set, to control for extraneous factors – we give an honest depiction of relative performance given the student population and data set available to us.

Value added model development

In the end, the struggle to determine the best VAM to apply in this study was essentially decided for us. Due to the fact that only pre and posttest data were provided whereas demographic, individual student characteristics and other possibly confounding data were not provided, our choices were limited. As stated earlier, the pretest data was generated from California Standards Testing (CST) exams given in the spring of 2010. The posttest data was generated from CST exams given in the spring of 2011. Our original plan was to follow an established procedure using normal curve equivalents (NCEs). This is a process used in early SAS EVAAS analysis where value added was based, essentially, on the change in normal curve z-scores from year to year, using the reference distribution for the full test-taking population in the state (Wright, White, Sanders, & Rivers, 2010). We learned by examining our data that students who all took the same posttest took differing pretests. For example, from the group whose posttest consisted of Algebra 1, students took any one of four different pretests; 7th grade math, 8th-9th general math, Algebra 1 (repeated) or even Geometry. In this report we define a "cohort" to be any collection of students in the study whose pretest-posttest pair is the same. Thus all those students who took the Geometry exam in 2010 and then took the Algebra 1 exam in 2011 form a single cohort. With this definition we see that the Algebra 1 posttest group includes four significant cohorts. Cohorts consisting of fewer than 36 students were excluded from the analysis due to the low correlation coefficients in the linear regression. Appendix A contains technical data for the cohorts included in the study.

California testing policy required that students enrolled in and attending a specific math or English course in academic year 2010-2011 must take the associated CST exam in the spring of 2011. Therefore the exam taken by a student informed us of exactly which course the student was enrolled in. For example, all students who took the Algebra 1 exam in 2011 were also enrolled in the Algebra 1 course at their virtual school during the academic year 2010-2011. Our objective was to establish a value-added score for each *course* at each school. This meant we needed to develop some way to pool value-added assessments from the various cohorts taking the same posttest. Excluded cohorts consisted of pretest-posttest pairs that represented unusual course sequencing. For example, one excluded cohort consisted of students whose tests indicated they took Algebra 2 and subsequently took Algebra 1.

It took some time to understand the implications of multiple cohorts taking a single posttest, but after some analysis we realized that the typical SAS EVAAS approach using NCEs would not provide a true picture of value-added. The normal process of equating value-added with the z-score changes based on the respective reference distributions simply doesn't work when you have multiple cohorts. The NCE approach is essentially equivalent to redefining student test score values as the corresponding z-scores earned on each test relative to the distributions for the full populations taking those exams. One reason this method breaks down in our situation is because the cohorts are not random distributions of the pretest population. For example, one cohort consists of students who took the Algebra 1 exam in *both* years. We would normally expect a student to repeat Algebra 1 only if they performed below expectations on the pretest. So the mean of the pretest scores for this restricted cohort will certainly be much lower than the reference distribution mean for the full population whose pretest was Algebra 1. Similarly, the English language arts cohort of those students taking the 9th grade ELA exam as their posttest and the 7th grade ELA exam as their pretest would generally consist of those students who did exceptionally well on the pretest and skipped 8th grade. For this cohort we would expect a much higher mean on the pretest than the reference distribution. These kinds of variations in cohort pretest averages unacceptably distort the meaning of value added based on NCEs using reference distributions for the full test taking populations.

Based on the above considerations, we selected a standard linear regression method encouraged by the Value Added Research Center (Value-Added Research Center, 2015). We first established expected posttest scores based on linear regression of the known data in each individual cohort. Residuals then formed the value added score for each student. This method is similar to that used early on by the United Kingdom in establishing value-added measures for their schools. Given our research goals and given the data set we had access to, we were confident the method would produce meaningful distinctions between our identified online schools that could also be digested and duplicated by a wide audience.

The model used to establish expected posttest scores was the following:

$$\hat{y}_{ijk} = \alpha_{jk} + \beta_{jk} x_{ijk} + \varepsilon_{jk}$$

Here *j,k* represents the pretest-posttest pair and *i* indicates the student. The α and β are regression coefficients, $\hat{y}_{ijk}$ is the expected posttest score, and ε is normally distributed random error term with mean 0 and constant variance. Value added is interpreted as the residual or difference between the expected posttest score, $\hat{y}_{ijk}$ and the actual scaled posttest score earned, y_{ijk}. Since the value added is the residual of the regression, the mean of the value added scores will necessarily be 0 in each cohort. To combine the value added scores for all cohorts with the same posttest we first convert all the residuals to z-scores based on the distributions of the residuals within the appropriate cohort. This may amplify or dampen the value-added within specific cohorts, but is consistent with the idea that fluctuations in variances between cohorts are primarily artifacts of the varying scaled score magnitudes. We then take the z-score equivalent of the residual as the value added for the particular student. The pooled values-added for the cohorts for a specific posttest were then sorted by school. The mean of these scores represents the value-added for the associated course at the particular school. Mathematically, we calculated that value added as follows:

First, we established value added, z_{ijk}, for student *i* for cohort *j,k*: (pretest *j*, posttest *k*):

$$z_{ijk} = (y_{ijk} - \hat{y}_{ijk}) / \sigma_{jk}$$

where σ_{ijk} is the standard deviation of the set of residuals, $\left\{ y_{ijk} - \hat{y}_{ijk} \right\}$ over the (*j*, *k*) cohort. Then for each school, the value-added in a specific course associated with posttest *k* is the mean of the values-added for all students in the study who took the k^{th} posttest at the school.

Standard one-sided 95% confidence intervals were then established for each (course, school) pair. We identified “distinguished schools” for a specific course as any school whose value-added confidence interval was entirely positive. (Note the overall course means will also be zero since each cohort mean is zero.) Our interpretation is that any school that is designated distinguished for a course is above average with 95% certainty.

Data and results

The eight courses studied included the following: English language arts for grades 8, 9, 10 and 11, General Mathematics, Algebra 1, Geometry, and Algebra 2. The distinguished schools, courses, and relevant data are presented below:

English language arts

iHigh Virtual Academy in San Diego had an outstanding 10^{th} grade class in AY 2010-2011. Their adjusted value added was .765 standard deviations above the mean giving 95% confidence that they performed at least .355 standard deviations above the mean on average. They had 18 students, which most likely represents a single class. CA Virtual Academy at Los Angeles performed above average with 95% confidence in both 9^{th} and 10^{th} grades. Their student count is quite high, 262 and 239 respectively. The same was true for CA Virtual Academy at San Diego who performed above average in both 8^{th} and 9^{th} grades.

Distinguished schools performing above average with 95% confidence

ELA 8
CA Virtual Academy, San Diego

ELA 9
CA Virtual Academy, Kings
CA Virtual Academy, LA
CA Virtual Academy, San Diego

ELA 10
Capistrano Connections Academy
CA Virtual Academy, Kings

CA Virtual Academy, Los Angeles
CA Virtual Academy, San Joaquin
EDUHSD Virtual Academy at Shenandoah (El Dorado)
iHigh Virtual Academy - San Diego

ELA 11
iHigh Virtual Academy - San Diego
Riverside Virtual

ELA data and results for high performing schools in the study are provided in Appendix B. A complete data set is available by request.

Mathematics

Most notable in mathematics was the general mathematics course at the iQ Academy in Los Angeles. Their mean value added was .776 standard deviations above the mean producing a minimum of .230 standard deviations above the mean with 95% confidence. The number of students was small, 11, and likely represents a single class with a single teacher. We note that due primarily to small individual classes in nearly all of the online algebra 2 courses, no single school performed at the distinguished level. The best school in this category was probably the Choice 2000 Online School in Riverside County with a mean value added of .396 standard deviations above the mean. The data covered only 18 students so the lower end of the 95% confidence interval extended down to -.120. Therefore this school does not meet our definition of distinguished.

Distinguished schools performing above average with 95% confidence

General Mathematics
Capistrano Connections Academy
iQ Academy LA

Algebra 1
CA Virtual Academy, Los Angeles
La Entrada Yorba Linda

Geometry
CA Virtual Academy, San Mateo
CA Virtual Academy, Los Angeles

Algebra 2
NONE

Mathematics data and results for high performing schools in the study are provided in Appendix C. A complete data set is available by request.

Other notable performances

By examining 90% confidence intervals we identified additional *notable* schools that were close to meeting our standard for distinction. We observed that a small number of schools performed well in multiple mathematics courses. Those schools were RAI Online Charter, performing *notably* in General Mathematics, Algebra 1, and Geometry. CA Virtual Academy at Los Angeles performed with distinction in both Algebra 1 and in Geometry. Capistrano Connections Academy performed with distinction in General Mathematics, but also performed notably in Geometry.

Limitations of the study

It must be emphasized that the primary objective of this initial research was to apply a value added approach to identify schools that indicated greater growth from one year to the next. The second phase of study would be to further investigate why this is the case. We would explore whether or not there was something in particular that these higher performing schools and programs were doing that resulted in greater growth over time. For this initial study, a simple value added model was applied that identified distinguished schools solely on the basis of the residuals of linear regression applied to cohorts of students that took the same pretest-posttest pair. It is becoming widely accepted among educators and researchers that proper identification of the true contribution of the educational experience to test performance must take into consideration a variety of additional factors. The data available for this analysis did not include additional factors and represents a limitation to the study. Another limitation to this study is the small number of students involved relative to the test-taking population of California. A few specific pretest-posttest cohorts that have large numbers statewide were excluded from this study because the number of students in our pool fell below 36. The online populations, while growing quite rapidly, still represent a very small percentage of the full population. As the online population continues to grow, opportunities for analysis that takes into consideration all test–pair cohorts and multiple years of performance will develop. These future studies will improve our ability to identify distinguished schools with greater certainty. Finally, this study is limited by the restraints of the current standardized testing system. Standardized tests are only one measure of student learning and represent a very narrow range of overall learning outcomes. In addition, their overall quality is limited to the types of subjects that lend themselves well to standardized tests.

Discussion and summary

We are witnessing tremendous growth in the number of public school students choosing to receive their education from authorized public online and hybrid schools. School leaders are being pressed to expand the number of authorized online schools. However, there is very little, if any research evaluating online schools or programs using value-added measures. The objective of this research was to illustrate how value-added methods can be used to identify online schools in California that perform with distinction compared with their counterparts. A simple value-added model was explained and applied to standardized testing results. The model measured educational growth differences at 32 schools in 8 subjects during the 2010-2011 academic year. The growth was based on California Standards Test data in successive years 2010 and 2011 for each student included in the study. The student pool consisted of those who were enrolled in one of 32 identified online or hybrid public schools at the time of the 2011 testing and whose corresponding test scores for 2010 were available.

This report included a review of the various value-added models. A brief discussion was provided pointing out the need for mindful policy development to avoid unintended consequences that may arise due to the control of non-educational variables in these models. A very basic value-added model was described and applied to the CST data. In this model no student characteristics were controlled. We avoided any distortion to the value-added calculations and did not account for variables such as free and reduced lunch densities, or other socio-economic or racial or any other factors. The underlying assumption in the selection of such a simple model is the idea that all students have similar capacities to learn. Despite our choice, we do not rule out that the control of some variables might need to occur to properly understand the relative quality of program outcomes. The selection of which variables to control requires significant discussion, well beyond the scope of this discussion. With the use of this model, the subsequent value-added results indicated the existence of distinguished online schools that perform above average with 95% statistical confidence in seven of the eight course categories analyzed.

Implications for further study

Educational effectiveness is multifaceted and any investigation into effectiveness should consider a multifaceted approach. Value-added measures only inform us about one dimension of the entire educational process (Condie, Lefgren, & Sims, 2014; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Konstantopoulos, 2014; Polikoff & Porter, 2014; Sanders, 2000) and should be viewed as one measure in a complex school or program improvement process. This holistic approach to school improvement would support future research examining the quality of program level interventions, resources, supports and curriculum. Based on this rationale, the next phase of study would evaluate:

- The quality of program facilities and resources. Do students have an opportunity to collaborate and ask for help? If needed, are students provided additional resources and assistance such as access to online tutors or additional online content and practice for example?
- The quality of program content, knowledge and skill development (curriculum). Is the curriculum aligned with content area standards and to state standardized tests? Is it of sufficient rigor? Does it allow for teacher input and/or adaptation? Are extension activities built into the curriculum? Is there evidence of quality in online course design?
- The quality of program supports for students. Does the program offer extended-time learning opportunities? Are there math labs or tutorial sessions for struggling students for example? Is there a staffed help line or open lab hours? Are students provided multiple pathways for learning?
- The quality of program supports for teachers. Are teachers offered regular and consistent professional development in online teaching methods? Is peer coaching or mentoring integrated into professional practice?

In addition, because there is some evidence that increased certainty in estimated value-added scores have been shown over time (Cocoran, 2010; Ferrão & Couto, 2013), research using value-added measures should adopt a longitudinal approach. Future research would include applying the same value-added model of investigation to subsequent yearly cohorts.

Finally, online schools and programs offer a fairly unique opportunity in value-added investigations, in that results from these investigations can be linked directly to student and teacher behavioral data stored in server logs. Future studies using educational data mining combined with value-added measures may provide another avenue to further our knowledge and increase the confidence in value-added approaches to program evaluation.

Value added analysis of student performance over time can be a valuable tool when used appropriately. The results of our analysis should in no way be used as a judgment about the overall quality of any one school, course or teacher but rather as an initial large scale study, using aggregate data as part of a long term integrated analysis. Assessment and identification of best practices in online education is a growing national imperative and our intent was to focus on identifying high performing online schools and in so doing lay the foundation for further investigation into the what these identified “distinguished” schools are doing to promote long-term growth in student outcomes.

References

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS[R] EVAAS[R]) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives, 20*(12).

Atteberry, A. (2012, March). *The Misattribution of summers in teacher value-added*. Paper presented at Society for Research on Educational Effectiveness Spring 2012 Conference: Understanding Variation in Treament Effects, Washington, DC. Retrieved from https://www.sree.org/conferences/2012s/program/downloads/abstracts/580.pdf

Briggs, D., & Dominigue, B. (2011). *Due dilligence and the evaluation of teachers*. Boulder, CO: National Education Policy Center. Retreived from http://nepc.colorado.edu/publication/due-diligence

Condie, S., Lefgren, L. & Sims, D. (2014). Teacher heterogeneity, value-added, and education policy. *Economics of Education Review, 40*, 76-92.

Corcoran, S. (2010). The Use of value-added measures of teacher effectiveness in policy and practice. In *Education Policy for Action Series: Education Challenges Facing New York City*. Providence, RI: Annenberg Institute for School Reform at Brown University. Retrieved from http://annenberginstitute.org/pdf/valueaddedreport.pdf

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. H., & Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappan, 93*(6), 8-15.

DfES Analytical Services. (2004). Value added technical information [Report]. Retrieved from http://www.education.gov.uk/schools/performance/archive/schools_05/sec9.shtml

Doran, H. C., & Fleischman, S. (2005). Challenges of value-added assessment. *Educational Leadership, 63*(3), 85-87.

Evans, H. (2008). *Value-added in English schools.* London, UK: Department for Children, Schools, Families. Retrieved from http://www.wcer.wisc.edu/news/events/VAM%20Conference%20Final%20Papers/VAMinEnglishSchools_HEvens.pdf

Ferrão, M. E., & Couto, A. P. (2013). The Use of school value-added model for educational improvement: A Case study from the Portuguese primary education system. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 25*(1), 174-190.

Goe, L. (2008). *Using value-added models to identify and support highly effective teachers*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/UsingValueAddedModels.pdf

Isenberg, E., & Hock, H. (2011). *Design of value-added models for IMPACT and TEAM in DC public schools, 2010-2011 school year*. Washington, DC: Mathematica Policy Research Inc. Retrieved from http://files.eric.ed.gov/fulltext/ED521120.pdf

Konstantopoulos, S. (2014). Teacher effects, value-added models and accountability. *Teachers College Record, 116*(1), 1-21.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A Meta-analysis and review of online learning studies*. Washington, DC: US Department of Education. Retrieved from http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

McCaffrey, D., Lockwood, J. R., Koretz, D., Louis, T., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*(1), 67-101.

Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis, 36*(4), 399-416. doi:10.3102/0162373714531851

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice?. *Journal of Educational and Behaviorial Statistics*, *29*(1), 121-129.

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education, 14*(4), 329-339.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299-311.

Staker, H. (2011). The Rise of K-12 blended learning, profiles of immerging models. *Innosight Institute.* Retrieved from http://www.christenseninstitute.org/wp-content/uploads/2013/04/The-rise-of-K-12-blended-learning.emerging-models.pdf

Tekwe, C. D., Carter, R. L., Ma, C. X., Algina, J., Lucas, M. E., Roth, J., Ariet, M., Fisher, T., & Resnick, M. B. (2004). An Empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29*(1), 11-35.

Value-Added Research Center. (2015). *Tutorials* [Web page]. Retrieved from http://varc.wceruw.org/tutorials/

Wright, S. P, White. J. T., Sanders, W. L., & Rivers. J. C. (2010). *SAS EVAAS Statistical Models.* Cary, NC: SAS Institute. Retreived from http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf

Appendix A

Technical data

Value Added Regression Model: $\hat{y}_{ijk} = \alpha_{jk} + \beta_{jk} x_{ijk} + \varepsilon_{jk}$

English Language Arts Cohort Data

Test Pair	α	β	Correlation	*N*	VA ST DEV
07 -> 08	35.2028863	0.898920487	0.837081373	1349	33.63325091
08 -> 09	75.3121477	0.794646791	0.819974315	1096	57.68691759
09 -> 09	82.50142079	0.762449557	0.753896931	40	44.95626846
09 -> 10	37.98750984	0.852815283	0.823073882	1298	32.44465709
10 -> 10	136.8518591	0.588405325	0.601939595	37	45.9677917
09 -> 11	-38.10792065	1.013138697	0.858618204	39	31.13760807
10 -> 11	49.91924363	0.847915759	0.802559329	1178	36.1823348
11 -> 11	85.21475599	0.762701545	0.704078496	48	40.31095084

Note.VA ST DEV = standard deviation of cohort scaled score residuals. Example: 07-> 08 pair is ELA grade 7 pretest, ELA grad 8 posttest pair.

Mathematics cohort data

Test Pair	α	β	Correlation	*N*	VA ST DEV
0->1	86.33706271	0.731949056	0.683467932	255	38.81331348
1->1	65.6262479	0.80254344	0.763661159	190	32.07149261
3->1	127.7321388	0.68344329	0.618302607	92	46.80170421
0->3	75.91702932	0.628227061	0.684321574	1031	42.57601403
1->3	130.7451863	0.443840181	0.597067193	447	33.55203361
3->3	99.34208976	0.656159185	0.691095793	557	32.140072
5->3	129.1694387	0.59520695	0.597078508	46	37.98014106
1->5	103.8829867	0.531670478	0.713017217	129	35.48511869
3->5	123.9689371	0.520891358	0.62430379	827	39.9145496
5->5	84.37919208	0.713123308	0.64708855	108	31.80480223
3->7	-32.84972536	0.911098088	0.797665759	36	46.07957427
5->7	72.81724677	0.650672789	0.679836044	429	36.0494139

Note. Test Codes: 0 = 7th Grade Math; 1 = 8th-9th General Math; 3 = Algebra; 1, 5 = Geometry; 7 = Algebra 2.

Appendix B

English Language Arts

Course: ELA Grade 8

	ELA 8		95% Confidence	90% Confidence
	N	Mean VA	VA Min	VA Min
CA Virtual Academy, San Diego	234	0.135	0.027*	0.051**
Capistrano Connections Academy	108	0.147	-0.011	0.023**

Note. Per California CDE policy data is deleted when N is below 11 students. VA = value added, VA Min = lower limited of the confidence interval, *N* = student count. *indicates entirely positive 95% confidence intervals; **indicates entirely positive 90% confidence intervals.

Course: ELA Grade 9

	ELA 9		95% Confidence	90% Confidence
	N	Mean VA	VA Min	VA Min
CA Virtual Academy, Kings	38	0.286	0.019*	0.077**
CA Virtual Academy, Los Angeles	258	0.197	0.094*	0.117**
CA Virtual Academy, San Diego	156	0.183	0.051*	0.080**

Course: ELA Grade 10

	ELA 10		95% Confidence	90% Confidence
	N	Mean VA	VA Min	VA Min
CA Virtual Academy, Kings	39	0.302	0.038*	0.096**
CA Virtual Academy, Los Angeles	239	0.195	0.088*	0.111**
CA Virtual Academy, San Joaquin	33	0.338	0.051*	0.114**
Capistrano Connections Academy	113	0.238	0.082*	0.116**
EDUHSD Virtual Academy at Shenandoah (El Dorado)	29	0.360	0.045*	0.116**
iHigh Virtual Academy - San Diego	18	0.765	0.355*	0.451**

Course: ELA Grade 11

	ELA 11		95% Confidence	90% Confidence
	N	Mean VA	VA Min	VA Min
CA Virtual Academy, Los Angeles	235	0.085	-0.023	0.001**
CA Virtual Academy, San Diego	148	0.123	-0.013	0.017**
CA Virtual Academy, San Mateo	77	0.163	-0.025	0.016**
iHigh Virtual Academy - San Diego	16	0.525	0.087*	0.190**
Riverside Virtual	17	0.603	0.180*	0.279**

Appendix C

Mathematics

Course: General Mathematics

	General math		95% Confidence	90% Confidence
	N	Mean VA	Mean Min	Mean Min
Capistrano Connections Academy	96	0.224	0.056*	0.092**
iQ Academy LA	11	0.776	0.230*	0.363**
RAI Online Charter - San Diego	23	0.322	-0.036	0.046**

Course: Algebra 1

	Algebra 1		95% Confidence	90% Confidence
	N	Mean VA	Mean Mininum	Mean Mininum
CA Virtual Academy, Kings	79	-0.124	-0.310	-0.269
CA Virtual Academy, Los Angeles	581	0.115	0.047*	0.062**
La Entrada Yorba Linda	11	0.632	0.086*	0.219**

Course: Geometry

	Geometry		95% Confidence	90% Confidence
	N	Mean VA	Mean Mininum	Mean Mininum
CA Virtual Academy, Los Angeles	264	0.151	0.050*	0.072**
CA Virtual Academy, San Mateo	84	0.250	0.070*	0.110**
CA Virtual Academy/Jamestown - Tuolumne	12	0.406	-0.112	0.013**
Capistrano Connections Academy	60	0.191	-0.022	0.024**
RAI Online Charter - San Diego	11	0.530	-0.017	0.116**