Technical Report on the Wisconsin School-Level Value-Added Model Academic Year 2015-2016

Prepared by
Value-Added Research Center
Wisconsin Center for Education Research
University of Wisconsin-Madison

DOCUMENT CONTROL

Title	Technical Report on the Wisconsin School-Level Value-Added Model Academic Year 2015-2016
Revision	4
Issue Date	11/3/2016
Security Level	Public
Filename	WI DPI School VA Technical Report.pdf
Changes	Revision 2: • Added document control section. • On page 12 changed “effect of economic disadvantage on student growth” to “effect of female gender on student growth”. Revision 3: • Added detail on the conversion of posttest standard deviations into posttest scale score units (page 3-4 and page 13). Revision 4: • Revised tables: sample summary statistics, coefficient estimates, and correlations based on updated economic disadvantaged data.

INTRODUCTION

This report describes the value-added model used by the Value-Added Research Center (VARC) of the Wisconsin Center for Education Research at the University of Wisconsin to measure the productivity or effectiveness of Wisconsin public schools using Forward and Badger test score data. The report is in four parts. The first part describes the data set used to produce the value-added estimates. The second part describes the model used to estimate value-added for schools in Wisconsin. The third part presents some properties of the value-added results. Finally, the fourth part identifies areas for continuous improvement in the development and production of value-added measures.

Conceptually, value-added analysis is the use of statistical techniques to isolate the component of measured student knowledge that is attributable to schools from other factors such as prior knowledge and student characteristics. In practice, value-added models focus on the improvement students make on annual assessments from one year to the next. Value-added models often control for measurable student characteristics using available data such as economic disadvantage and disability, to help isolate the impact of schooling. The model used in Wisconsin uses the available set of student characteristics to identify the extent to which schools contribute to the improvement of student achievement outcomes.

ANALYSIS DATA SET

Before estimation can take place, a substantial amount of work is required to assemble the analysis data sets used to produce the value-added estimates. A separate analysis data set is produced for each grade, subject, and test. In total, 10 analysis data sets are produced, covering grades 4 through 8 for Forward English language arts (ELA) and math in 2015-16.

Each analysis data set includes students who have a posttest in the grade and subject being considered, pretests in both ELA and math, had full academic year (FAY) status in their school or district, and were tested in consecutive grades.

Student-level variables

Posttest and pretest variables

The test scores used are from 2014-15 Badger and 2015-16 Forward assessments. The value-added system produces school-level measures for grades 4 through 8 in ELA and math based on performance on the Forward assessment. Value-added in ELA and math is defined by its usage of an ELA or math test as a posttest. All value-added models include pretests in both ELA and math. Since the Badger and Forward assessments were scored using different test scales (with very different means and standard deviations), in the value-added statistical analyses all test scores were linearly transformed to the z-statistic scale with means equal to zero and standard deviations equal to 1 in each grade and subject. Thus, in the value-added analyses, all test scores were measured relative to the state means and in the units of the statewide standard deviation of test scores in given grades and subjects. Use of this transformation is common when posttest and pretest test scores are measured on different scales. The transformation is used to make it easier to interpret estimates of the value-added models, but it does not affect the statistical power of the model or the ranking of estimated school effects.

Reliability of pretest variables

The reliability estimates of math and ELA pretest scores are available in the technical manual for the Badger exam prepared by Educational Testing Service (ETS). They range from 0.87 to 0.91 across grades and subjects. These reliability estimates are used for a correction for measurement error in the pretests.

Gender, race/ethnicity, economic disadvantage, and migrancy

Gender, race/ethnicity, economic disadvantage, and migrancy are drawn from the Forward student test score dataset. In the analysis data set, students are assigned the gender, race/ethnicity, low-income status, and migrant status reported in the posttest year. Gender categories are male and female. Race categories are American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, Black/African American, Hispanic/Latino, White, and multi-racial. The analysis employs an indicator for economically disadvantaged students and an indicator for migrant students.

English as a second language (ESL) classification

There are six indicators for English-language proficiency included in the analysis dataset, based on students' ESL classification into seven categories. Students with ESL classifications of 1 through 5 are considered to be English-language learners. Students with an ESL status of 6 are those that were formerly classified as having limited English proficiency. Students with an ESL classification of 7 are considered to be proficient in English and form the omitted group.

Disability

The analysis includes five indicators for students with disabilities according to their primary disability code. There are separate indicators for emotional/behavioral disability (EBD), learning or intellectual disability (LD/ID), autism (A), and speech/language disability (SL). All other disability codes are grouped into a single indicator for other disabilities.

School enrollment

Students that have FAY status at a single school are assigned to that school using the school enrollment data. Some students have FAY status in a single district but not at a single school because of mobility within the district. These students are assigned to a placeholder school within their district.

In previous years, mobile students would have contributed to the value-added estimate of each enrolled school proportionally to the fraction of the school year that they were at the school. This type of value-added analysis is sometimes called a 'dosage' model. In the current year, we do not use a dosage model to accommodate mobile students. A student only contributes to a school or district if they have FAY status at that school or district.

Descriptive statistics of analysis samples

The following tables describe the sample used for the 2016 year:

Forward Math Sample

Grade Level	**4**	**5**	**6**	**7**	**8**
Number of Students	55437	55320	55770	54986	54611
Posttest (Forward) Mean	576.686	602.322	615.488	630.647	644.190
Math Pretest (Badger) Mean	2440.335	2479.693	2509.183	2531.043	2549.745
ELA Pretest (Badger) Mean	2437.616	2472.513	2515.806	2524.298	2558.042
Posttest Standard Deviation	54.673	49.240	52.071	55.989	55.676
Math Pretest Standard Deviation	72.638	78.492	88.276	96.317	103.308
ELA Pretest Standard Deviation	81.632	89.054	89.223	94.586	96.503
Proportion in ESL Level 1	0.001	0.001	0.001	0.001	0.001
Proportion in ESL Level 2	0.003	0.003	0.002	0.002	0.003
Proportion in ESL Level 3	0.015	0.010	0.009	0.012	0.012
Proportion in ESL Level 4	0.032	0.024	0.018	0.018	0.016
Proportion in ESL Level 5	0.023	0.013	0.011	0.006	0.007
Proportion in ESL Level 6	0.018	0.037	0.043	0.044	0.042
Proportion Female	0.489	0.489	0.488	0.486	0.489
Proportion Asian	0.039	0.039	0.038	0.037	0.036
Proportion African-American	0.089	0.085	0.083	0.081	0.082
Proportion Hispanic	0.120	0.113	0.110	0.108	0.104
Proportion Native American	0.012	0.012	0.012	0.012	0.012
Proportion Native Hawaiian or Other Pacific Islander	0.001	0.001	0.000	0.001	0.000
Proportion Two or More Races	0.033	0.030	0.029	0.025	0.025
Proportion Special Education LD/ID	0.034	0.041	0.048	0.051	0.055
Proportion Special Education EBD	0.014	0.015	0.015	0.016	0.016
Proportion Special Education A	0.011	0.011	0.011	0.010	0.010
Proportion Special Education SL	0.038	0.026	0.015	0.012	0.008
Proportion Special Education Other	0.028	0.029	0.032	0.031	0.032
Proportion with Economic Disadvantage	0.439	0.418	0.404	0.388	0.382
Proportion Migrant	0.000	0.000	0.000	0.000	0.001

Forward ELA

Grade Level	**4**	**5**	**6**	**7**	**8**
Number of Students	55457	55349	55789	55022	54632
Posttest (Forward) Mean	585.140	602.080	612.561	626.324	639.739
Math Pretest (Badger) Mean	2440.318	2479.644	2509.144	2530.999	2549.728
ELA Pretest (Badger) Mean	2437.590	2472.472	2515.774	2524.262	2558.053
Posttest Standard Deviation	48.716	50.195	51.397	54.072	56.410
Math Pretest Standard Deviation	72.640	78.524	88.304	96.334	103.301
ELA Pretest Standard Deviation	81.643	89.064	89.243	94.613	96.466
Proportion in ESL Level 1	0.001	0.001	0.001	0.001	0.001
Proportion in ESL Level 2	0.003	0.003	0.002	0.002	0.003
Proportion in ESL Level 3	0.015	0.010	0.009	0.012	0.012
Proportion in ESL Level 4	0.032	0.024	0.018	0.018	0.017
Proportion in ESL Level 5	0.023	0.013	0.011	0.006	0.007
Proportion in ESL Level 6	0.018	0.037	0.043	0.044	0.042
Proportion Female	0.489	0.489	0.488	0.486	0.489
Proportion Asian	0.039	0.039	0.038	0.037	0.036
Proportion African-American	0.089	0.085	0.083	0.081	0.082
Proportion Hispanic	0.120	0.113	0.110	0.108	0.104
Proportion Native American	0.012	0.012	0.012	0.012	0.012
Proportion Native Hawaiian or Other Pacific Islander	0.001	0.001	0.000	0.001	0.000
Proportion Two or More Races	0.033	0.030	0.029	0.025	0.025
Proportion Special Education LD/ID	0.034	0.041	0.049	0.051	0.055
Proportion Special Education EBD	0.014	0.015	0.015	0.016	0.016
Proportion Special Education A	0.011	0.011	0.011	0.010	0.010
Proportion Special Education SL	0.038	0.026	0.015	0.012	0.008
Proportion Special Education Other	0.028	0.029	0.032	0.031	0.032
Proportion with Economic Disadvantage	0.439	0.418	0.404	0.388	0.382
Proportion Migrant	0.000	0.000	0.000	0.000	0.001

VALUE-ADDED MODEL

For the Wisconsin school level model, value-added is measured in math and ELA in grades four through eight for the Forward assessment. Schools are assigned single-year value-added measures that reflect student growth in 2015-16.

The model, in brief

The value-added model is defined by four equations: a "best linear predictor" value-added model defined in terms of true student post and prior achievement and three measurement error models for observed post and prior achievement:

Student achievement: $y_{1i} = \zeta + \lambda y_{0i} + \lambda^{alt} y_{0i}^{alt} + \beta' X_i + \alpha' S_i + e_i$ (1)

Posttest measurement error: $Y_{1i} = y_{1i} + v_{1i}$ (2)

Same-subject pretest measurement error: $Y_{0i} = y_{0i} + v_{0i}$ (3)

Other-subject pretest measurement error: $Y_{0i}^{alt} = y_{0i}^{alt} + v_{0i}^{alt}$ (4)

where:

- y_{1i} is true post achievement;
- y_{0i} and y_{0i}^{alt} are true prior achievement in the same subject and in the other subject (math in the ELA model, ELA in the math model), with slope parameters λ and λ^{alt};
- X_i is a vector of characteristics of student i, with slope parameter vector β;
- S_i is a vector of indicators for school;
- α is a vector of school effects;
- e_i is the error in predicting post achievement given the explanatory variables included in the model;
- Y_{1i} is measured post achievement;
- v_{1i} is measurement error in post achievement;
- Y_{0i} and Y_{0i}^{alt} are measured prior achievement; and
- v_{0i} and v_{0i}^{alt} are measurement error in prior achievement.

Substituting the measurement error equations (2), (3), and (4) into the student achievement equation (1) yields an equation defined in terms of measured student achievement:

Measured achievement: $Y_{1i} = \zeta + \lambda Y_{0i} + \lambda^{alt} Y_{0i}^{alt} + \beta' X_i + \alpha' S_i + \varepsilon_i$ (5)

where the error term ε_i includes both the original error component and the measurement error components:

Error in measured achievement: $\varepsilon_i = e_i + v_{1i} - \lambda v_{0i} - \lambda^{alt} v_{0i}^{alt}$ (6)

Estimating the measured student achievement equation (5) without controlling for pretest measurement error yields biased estimates of all parameters, including the value-added effects. This bias stems from the fact that measurement error in prior achievement causes the error term (6), which includes the measurement error components v_{0i} and v_{0i}^{alt}, to be correlated with measured prior achievement. The desired parameters, as defined in equation (1), can be estimated consistently if external information is available on the variance of measurement error for prior achievement; approaches for consistent estimation in the presence of measurement error are described in detail in Wayne Fuller, *Measurement Error Models* (Wiley, 1987). Information about the variance of test measurement error is obtained from the reliability estimates reported in the technical manual for the 2014-15 Badger exam assessment.

A shrinkage approach is employed to ensure that schools with fewer students are not overrepresented among the highest- and lowest-value-added cases due to randomness. The approach, Empirical Bayes shrinkage, is described in J. N. K. Rao, *Small Area Estimation* (Wiley, 2003).

The variables in the model

In addition to posttest and pretest scores, the student-level variables included in the model (the X variables in equation 1) are gender, race/ethnicity, ESL category, economic disadvantage, disability code, and migrancy.

Value-added regression

The value-added model is estimated using a least-squares regression approach that corrects for measurement error in the pretest variables. It estimates the coefficients λ, β, and α by regressing posttest on same-subject pretest, other-subject pretest, other student-level variables, and a full set of school fixed effects. This can be expressed mathematically using equation (5) above:

Measured achievement: $Y_{1i} = \zeta + \lambda Y_{0i} + \lambda^{alt} Y_{0i}^{\ alt} + \beta' X_i + \alpha' S_i + \varepsilon_i$ (5)

This regression is estimated using an approach that accounts for measurement error in the pretests Y_{0i} and Y_{0i}^{alt}. Recall from equation (6) above that the measurement error components of Y_{0i} and Y_{0i}^{alt}, v_{0i} and v_{0i}^{alt}, are part of the error term ε_i. As a result, estimating the regression using ordinary least squares will lead to biased estimates. The regression approach employed accounts for measurement error by removing the variance in the pretests that is attributable to measurement error. To illustrate the measurement error corrected regression, re-cast the above value-added regression equation into vector form:

$$Y_t = Y_{t-1}\lambda + W\delta + \varepsilon$$

where Y_t is an N × 1 vector of post-test scores, $Y_{t\text{-}1}$ is an N × 2 vector of same-subject and other-subject pre-test scores $Y_{t\text{-}1}$ and $Y_{t\text{-}1}^{alt}$, λ is a 2 × 1 vector made up of λ and λ^{alt}, W is an N × K vector of the X demographic variables, δ is a K × 1 vector of the β and α^{*} coefficients, and ε is an N × 1 vector of error terms. The biased ordinary-least-squares estimates of the coefficients in λ and δ are equal to:

$$\begin{bmatrix}\hat{\lambda}_{OLS}\\ \hat{\delta}_{OLS}\end{bmatrix} = \begin{bmatrix} Y'_{t-1}Y_{t-1} & Y'_{t-1}W \\ W'Y_{t-1} & W'W \end{bmatrix}^{-1} \begin{bmatrix} Y'_{t-1}Y_{t} \\ W'Y_{t} \end{bmatrix}$$

The measurement-error-corrected estimates of the coefficients in λ and δ are equal to:

$$\begin{bmatrix}\hat{\lambda}_{EIV}\\ \hat{\delta}_{EIV}\end{bmatrix} = \begin{bmatrix} Y'_{t-1}Y_{t-1} - \sum_{i}^{N} sem_{it-1} & Y'_{t-1}W \\ W'Y_{t-1} & W'W \end{bmatrix}^{-1} \begin{bmatrix} Y'_{t-1}Y_{t} \\ W'Y_{t} \end{bmatrix}$$

where $sem_{it\text{-}1}$ is a 2 × 2 variance-covariance matrix of the errors of measurement of $Y_{it\text{-}1}$ and $Y_{it\text{-}1}^{alt}$ for student i. This model is described in section 2.2 of Wayne Fuller, *Measurement Error Models* (Wiley, 1987).

Aggregation to multiple-grade value-added

The value-added regression to obtain unshrunk school value-added is performed separately for each combination of grade and subject. For schools that have results for more than one grade level, these estimates are averaged across grades, using the number of students attributed to the school as weights, to produce unshrunk multiple-grade value-added estimates. Before aggregation, value-added measures by grade are normalized in order to be on similar scales (i.e. with a mean of 0 and a true standard deviation of 1) across grades. This normalization is made by dividing the measures by an estimate of the standard deviation of value-added within grade.

Shrinkage of value-added

At all levels, the unshrunk value-added estimates are shrunk using an Empirical Bayes univariate shrinkage technique described in J. N. K. Rao, *Small Area Estimation* (Wiley, 2003). This procedure is employed to bring value-added estimates based on smaller sample sizes closer to the state average, so that schools with fewer students are not overrepresented among the highest- and lowest-value-added cases simply due to randomness. This is estimated by multiplying each value-added measure by its reliability:

$$\alpha_{\text{shrunk}} = (\omega^2 / (\omega^2 + \sigma^2))\alpha_{\text{unshrunk}}$$

where α_{unshrunk} is an unshrunk value-added estimate for a given school; σ^2 is the squared standard error of α_{unshrunk}; and ω^2 is the variance of value-added across schools within subject, test, and grade(s). The standard error of the shrunk value-added estimate is equal to

$$\text{s.e.}\ (\alpha_{\text{shrunk}}) = \text{sqrt}\ [\omega^2\sigma^2 / (\omega^2 + \sigma^2)]$$

The variance measure ω^2 is estimated by computing the variance of the unshrunk value-added estimates, then subtracting from that the average squared standard error of the unshrunk value-added estimates. This variance measure is an estimate of the variance of the underlying value-added measures, excluding variance due to randomness in the value-added estimates. The square root of this variance measure is also used for normalizing value-added measures by grade before aggregation to multiple-grade measures.

Student group value-added

Value-added is also measured by student groups defined by certain demographic characteristics. Specifically, we calculated differential value-added effects for the seven race/ethnicity groups, for students with disabilities, for economically disadvantaged students, and for English-language learners.

To produce the group results by school, we regress the estimate of the sum of the school effects and the residual, $\alpha' S_i + \varepsilon_i$, on a vector of school indicators to produce a new residual, which we will refer to as ε_i^*. This residual is the component of student achievement that cannot be explained with the pretest scores, demographics, or any overall school effect. These residuals are then regressed on interactions between school indicators and the group variable demeaned within school:

$$\varepsilon_i^* = \sum_{j=1}^{J} \theta_j I_{ij}(x_i - x_j) + u_i$$

where I_{ij} is an indicator that equals 1 if student i is associated with school j, x_i is an indicator variable that indicates whether a student is part of the group, and x_j is the proportion of students in school j who are in the group. This yields a slope estimate θ_j for each school, which is shrunk using Empirical Bayes shrinkage. Value-added for students in the group for school j is set to $\alpha_j + \theta_j(1 - x_j)$, where α_j is overall value-added for school j. For students outside the group, value-added is set to $\alpha_j - \theta_j x_j$. This description is for a case of a binary student group variable, such as disability (where X_i can only equal 0 or 1).

In the case of race/ethnicity groups, of which there are seven (American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, Black/African American, Hispanic/Latino, White, and multi-racial), this approach is generalized to use six x variables, one for each race/ethnicity group with White as the omitted group. Student group value-added measures that cover multiple grades are computed by averaging the unshrunk slopes θ_j across grades and years, shrinking them using Empirical Bayes shrinkage, and combining them with overall multiple-grade value-added measures in the same way as for single-grade measures.

PROPERTIES OF THE VALUE-ADDED RESULTS

Coefficients on student-level variables in the model

The coefficients estimated in the value-added model are presented on the next page. To interpret these coefficients, note that both pretest and posttest are measured using standardized scores; therefore all coefficients are measured in the posttest standard deviation scale. For example, note that the coefficient on female gender is -0.083 in grade 4 Math. The posttest standard deviation for grade 4 Math is 54.673. This implies that male students improved 0.083 standard deviations or about 4.538 scale score points more on the grade 4 Math test from spring to spring than otherwise similar female students.

It is important to keep in mind the standard errors of the coefficients when interpreting them. A span of 1.96 standard errors in both the positive and negative directions provides a 95 percent confidence range for a coefficient. Continuing with the example of the coefficient on female gender in grade 4 Math, note that the standard error of this coefficient estimate is 0.005 in posttest SD units or 0.273 in scale score points. This means that, while our best estimate of the difference in growth between female and male students is -4.538 scale score points, a 95 percent confidence interval for the difference ranges from -5.085 to -3.991 scale score points.

Coefficients on student-level variables, 2015-16 Forward Math

Variable	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Math Pretest	0.756	0.008	0.749	0.006	0.682	0.007	0.794	0.007	0.753	0.009
ELA Pretest	0.031	0.008	0.105	0.006	0.190	0.007	0.086	0.007	0.098	0.009
ESL Level 1	-0.305	0.076	0.335	0.093	-0.195	0.094	-0.033	0.093	0.189	0.084
ESL Level 2	-0.348	0.044	0.023	0.047	-0.301	0.050	0.063	0.053	-0.098	0.051
ESL Level 3	-0.233	0.022	-0.019	0.026	-0.133	0.025	-0.073	0.024	-0.067	0.025
ESL Level 4	-0.061	0.016	0.028	0.018	-0.011	0.018	-0.043	0.020	0.039	0.022
ESL Level 5	0.045	0.018	0.081	0.023	0.072	0.023	0.070	0.032	0.079	0.031
ESL Level 6	0.066	0.020	0.054	0.015	0.049	0.013	0.056	0.014	0.043	0.015
Female	-0.083	0.005	0.012	0.005	-0.007	0.005	-0.027	0.005	0.011	0.006
Asian	0.016	0.015	0.071	0.015	0.102	0.014	0.022	0.015	0.055	0.016
African-American	-0.129	0.013	-0.055	0.013	-0.100	0.011	-0.081	0.013	-0.142	0.013
Hispanic	-0.053	0.010	-0.028	0.011	-0.022	0.010	-0.032	0.011	-0.026	0.011
Native American	-0.095	0.025	-0.039	0.025	-0.037	0.023	-0.014	0.025	-0.027	0.025
Native Hawaiian or Other Pacific Islander	-0.160	0.092	0.063	0.083	-0.041	0.101	0.064	0.103	0.089	0.115
Two or More Races	-0.045	0.014	-0.059	0.015	-0.033	0.014	-0.028	0.016	-0.023	0.016
Special Education LD/ID	-0.296	0.014	-0.049	0.013	-0.231	0.011	-0.059	0.012	-0.080	0.012
Special Education EBD	-0.173	0.021	-0.142	0.021	-0.182	0.019	-0.101	0.020	-0.153	0.021
Special Education A	-0.124	0.024	-0.057	0.024	-0.161	0.021	0.020	0.024	-0.021	0.026
Special Education SL	-0.057	0.013	-0.026	0.015	-0.055	0.018	-0.022	0.022	-0.061	0.028
Special Education Other	-0.194	0.015	-0.093	0.015	-0.236	0.013	-0.084	0.015	-0.114	0.015
Economic Disadvantage	-0.038	0.006	-0.056	0.006	-0.048	0.005	-0.026	0.006	-0.041	0.006
Migrancy	0.057	0.144	-0.007	0.141	0.045	0.120	0.249	0.141	0.151	0.101

Coefficients on student-level variables, 2015-16 Forward ELA

Variable	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Math Pretest	0.133	0.008	0.247	0.006	-0.006	0.008	0.178	0.007	0.105	0.009
ELA Pretest	0.733	0.008	0.582	0.006	0.867	0.008	0.706	0.007	0.819	0.009
ESL Level 1	-0.058	0.077	-0.209	0.093	-0.101	0.104	-0.259	0.093	0.010	0.082
ESL Level 2	-0.204	0.045	-0.318	0.047	-0.308	0.056	-0.366	0.054	0.098	0.050
ESL Level 3	-0.101	0.023	-0.142	0.026	-0.052	0.027	-0.114	0.024	0.044	0.025
ESL Level 4	-0.061	0.016	-0.054	0.018	-0.027	0.020	0.002	0.020	0.025	0.021
ESL Level 5	-0.008	0.018	-0.020	0.023	0.009	0.025	0.052	0.032	0.016	0.030
ESL Level 6	0.050	0.020	0.067	0.015	-0.008	0.015	0.036	0.014	-0.011	0.015
Female	0.034	0.005	0.151	0.005	0.061	0.005	0.076	0.005	0.050	0.006
Asian	0.040	0.016	0.009	0.015	0.051	0.015	0.139	0.015	0.080	0.016
African-American	-0.040	0.013	-0.046	0.013	-0.063	0.013	-0.041	0.013	-0.047	0.013
Hispanic	-0.003	0.010	0.005	0.011	-0.009	0.011	0.035	0.011	-0.007	0.011
Native American	-0.037	0.026	-0.045	0.025	-0.032	0.025	-0.010	0.025	-0.033	0.024
Native Hawaiian or Other Pacific Islander	0.108	0.093	0.005	0.083	-0.013	0.112	0.287	0.102	0.123	0.111
Two or More Races	-0.014	0.014	-0.003	0.015	-0.017	0.015	0.024	0.016	0.008	0.016
Special Education LD/ID	-0.212	0.014	-0.228	0.013	-0.202	0.012	-0.149	0.012	0.000	0.012
Special Education EBD	-0.169	0.022	-0.141	0.021	-0.157	0.021	-0.122	0.020	-0.018	0.020
Special Education A	-0.138	0.024	-0.154	0.024	-0.159	0.024	-0.065	0.024	0.032	0.025
Special Education SL	-0.034	0.013	-0.107	0.015	-0.054	0.020	-0.042	0.022	0.005	0.027
Special Education Other	-0.151	0.016	-0.192	0.015	-0.171	0.015	-0.123	0.015	-0.007	0.015
Economic Disadvantage	-0.049	0.006	-0.056	0.006	-0.045	0.006	-0.015	0.006	-0.012	0.006
Migrancy	-0.098	0.146	-0.012	0.141	-0.101	0.134	0.143	0.142	0.081	0.098

Correlation with average prior proficiency

Results show a very low correlation between average prior proficiency--a measure of average performance in the previous year--and value-added. In general, schools were not more or less likely to have a low value-added score than a high score if their students began the year with low pretest scores rather than high scores.

Correlations between Prior Attainment and Value-Added						
Subject	***Grade 4***	***Grade 5***	***Grade 6***	***Grade 7***	***Grade 8***	***School***
ELA	0.03	0.06	0.11	0.00	-0.02	0.03
Math	0.20	0.00	0.03	0.00	0.15	0.13

Correlation between Math and ELA value-added

There were also substantive positive correlations between math and ELA value-added within each school. Schools that were high value-added in math were also more often than not high value-added in ELA.

Correlations between Subjects						
Correlation	***Grade 4***	***Grade 5***	***Grade 6***	***Grade 7***	***Grade 8***	***School***
2015 Math and ELA	0.46	0.47	0.50	0.42	0.36	0.44

CONTINUOUS IMPROVEMENT

VARC has identified and proposed to DPI several options for the continuous improvement of value-added models to enable effective use of these results.

1. Develop and implement protocol to ensure highest accuracy of results
   Such a protocol would involve parallel team replication of all steps involved in producing value-added results and a process for reconciliation of possible differences in data and results. Several school districts and states have developed similar protocols in order to reduce the likelihood of errors under extremely compressed time frames.

2. Off-season review of 2015-2016 model and results
   As in previous years, VARC would resume an off-season review with DPI to examine the effects, if any, of using the Badger exam for pretest and Forward assessment for posttest (see also #5); differential value-added effects by demographic characteristics; model performance for small and large districts; and options for model improvement.

3. ACT Value-added model

VARC would explore development of value-added model using ACT or ACT Aspire test data for high school students in grades 9, 10 and 11 as the outcome measure and prior grade test data from ACT Aspire or Forward as the measure of prior achievement. This work would include model options to address the impact, if any, of students missing from the growth analysis due to dropping out of school.

4. Non-neutrality in school value-added with respect to school size

   VARC would explore evidence of non-neutrality in Wisconsin value-added data with respect to school size, and consider options for addressing non-neutrality as well as model selection. See Appendix for technical discussion of this issue.

5. Aggregation of measures over time

   The 2015-2016 Wisconsin model uses the Forward assessment for posttest and the Badger assessment for pretest; the 2016-2017 Wisconsin model uses Forward assessment for both posttest and pretest. VARC would develop method options for aggregating value-added results from two different posttest-on-pretest models across multiple school years. These aggregation methods would also work for combining three or more years of value-added measures. VARC would work with DPI to select the most appropriate method for Wisconsin. See Appendix for technical discussion of this issue.

6. Value-added for student groups

   The current method for estimating value-added measures for student groups relies on having sufficient sample size in the omitted or reference group. This is not restrictive in the case of binary demographic variables such as students with or without disability and students with or without economic disadvantage, where the reference group is well-represented in all schools. However, in the case of demographic variables with several categories such as race/ethnicity, we can encounter schools with very few or no students in the reference group resulting in noisy estimates of value-added for all student groups compared to the reference group. VARC would explore options that take into account contexts with insufficient sample size for the reference group in each category.

7. Test opt-out and implications for VA production

   VARC would examine and document incidences of test opt-out; explore the effects, if any, on value-added measures; explore method options for addressing effects of test opt-out, particularly with respect to using value-added measures in school report cards for accountability; and work with DPI to select the most appropriate method for Wisconsin.

8. School report card review

   A school report card review is necessary to ensure the accuracy of the value-added measures included in these reports (see also #4 and #5). VARC would also review the use of statistical shrinkage methods for controlling estimation error due to differences in the number of students across schools and for controlling non-neutrality with respect to the number of teachers per school (see #4 above).

9. Assessment analytics

   VARC would explore the effects, if any, of the change in assessments from the Badger test in 2014-15 to the Forward test in 2015-16 on student attainment and growth. VARC would provide method options for addressing test differences, if significant, and work with DPI to select the most appropriate method for Wisconsin.

10. Proactive approach to ensure that best-practice school value-added results are not inappropriately used to evaluate principals

    Research suggests that it takes up to five years for the full impact of a newly hired principal to be realized. Hence, school performance measures, including school value-added, likely underestimate the performance of new principals in turnaround schools. VARC would examine what steps can be taken to reduce the likelihood that newly hired principals are not inaccurately and unfairly evaluated. One option that addresses this concern is the calculation of principal value-added estimates for those districts that want to use an appropriate student-based outcome measure as one of multiple measures to monitor principal performance.

CONCLUSION

This technical report describes the value-added model used to estimate the productivity of Wisconsin public schools and developed by the Value-Added Research Center of the Wisconsin Center for Education Research at the University of Wisconsin. For more information on the value-added research of the Value-Added Research Center of the Wisconsin Center for Education Research at the University of Wisconsin, visit VARC's website at http://varc.wceruw.org

APPENDIX

Options for continuous improvement of measures and models of school value-added

I. School Value-Added is Non-Neutral with Respect to School Size

Tasks:

- Develop value-added model with school and teacher components
- Derive average school value-added
- Show that variance of average school value-added is a function of school size
- Develop indirect method for estimating school and teacher variance components

Problem to be addressed: Differences in teacher performance are a source of variation in average school value-added. This variation tends to partially cancel out in large schools, thereby reducing the variance on school value-added. This problem is distinct from the fact that estimates for schools with larger student populations have lower noise due to student variation. Figures 1-3 at the end of this appendix provide evidence on the magnitude of this issue. The graphs in these figures report value-added estimates by school size for students in grades 4, 5, and 6, using estimates from the Wisconsin value-added model for 2015-2016.

Model of teacher-school value-added:

$$\hat{\alpha}_{jk} = \underset{\text{School}}{\eta_k} + \underset{\text{Teacher}}{v_{jk}} + \underset{\text{Noise}}{\varepsilon_{jk}} \tag{A1}$$

Average school value-added:

$$\hat{\beta}_k = \bar{\hat{\alpha}}_{\cdot k} = \eta_k + \bar{v}_{\cdot k} + \bar{\varepsilon}_{\cdot k} \tag{A2}$$

Variance of school value-added is not neutral with respect to school size:

$$\text{Var}(\hat{\beta}_k)\text{=E}(\hat{\beta}_k^2) = \omega_\eta^2 + \frac{\omega_v^2}{J_k} + \frac{\sigma_k^2}{J_k} \tag{A3}$$

Problem: Teacher value-added is unknown, so it is necessary to address the non-neutrality indirectly via a statistic that is observed and closely related to the number of teachers, namely, school size.

School size = # of teachers * average class size

$$N_k = J_k * n_k \Rightarrow J_k = N_k / n_k \tag{A4}$$

Substituting for the number of teacher J_k yields an equation for the variance:

$$\mathrm{Var}(\hat{\beta}_k) = \mathrm{E}(\hat{\beta}_k^2) = \omega_\eta^2 + n_k \omega_v^2 \left(\frac{1}{N_k} \right) + \frac{\sigma_k^2}{J_k} \tag{A5}$$

$$\begin{gathered} s_k^2 = \frac{\sigma_k^2}{J_k} \\ \hat{\beta}_k^2 - s_k^2 = \underbrace{\omega_\eta^2}_{\text{Intercept}} + \bar{n}\underbrace{\omega_v^2 \left(\frac{1}{N_k} \right)}_{\text{Slope}} + r_k \end{gathered} \tag{A6}$$

II. *Aggregating Measures over Time*

It is generally possible to improve the precision of measures of school (or teacher) performance using measures for multiple years. There are two important questions that need to be addressed when deciding how best to combine, or aggregate, growth measures from multiple years. First, what is the state, district, or school interesting in measuring – the measurement objective? Second, what is the correct (or approximately correct) model of school performance over time? Answers to these questions determine both how best to combine multiple-year indicators and how to accurately compute precision (standard errors). We address these questions below.

Measurement objective. Possible measurement objectives could include the following:

1. Current performance
2. Current performance and using past performance information to boost accuracy, but adjusting for systematic experience effects for early career teachers (with or without adjusting for the experience effect in the current year)
3. Persistent/stable component of performance (with or without adjustment for experience effect)
4. Future predicted performance (similar to #3, with or without adjustment for experience effect)
5. Average performance over current and past years (say, 3 years total, with or without adjustment for experience effect)

Note that all of the measures listed above benefit can be best estimated using information from multiple years. In other words, it is not the case that average performance as a measurement objective (#5) is the only measure that requires or benefits from multi-year data. One way to think of the list above is to ask: would the public want to know about performance if it could be estimated without error. Once the measurement objective is clarified, we then consider how best to estimate the preferred measure.

It is possible that different policy/management applications could imply the need for different measurement objectives, although having several different, but similar, measures might be confusing. For example, it might be best to measure current performance (#1 or #2) if the goal is to provide a summative rating of performance that reflects school productivity over the past school year. On the other hand, persistent/stable performance (#3) or future predicted performance (#4) might be a more appropriate indicator for the purpose of selecting which school to attend.

Choosing the Correct Model. A multi-year model of school performance will, in general, need to allow for the following components:

1. Noise, or estimation error, due largely to the fact that school performance estimates are based on a limited number of student data points.
2. A component of performance that is real, but transitory; that is, not stable or persistent from one year to the next.
3. For a static model: A persistent (long run) component of performance assumed to be fixed
4. For dynamic model: A persistent (long run) component of performance assumed to be subject to change: linear time trend model
5. For dynamic model: A persistent (long run) component of performance assumed to be subject to change: serial correlation model

Note that whether the model treats the persistent component as fixed or subject to change can have a large effect on how best to combine multiple year data. Recognizing that it is possible for school and teacher performance to change is arguably a desirable model feature in that it fits with how many, if not most, educators' view that careers. There are several ways to allow for change in the long run performance component. In the above list (#5 and #6) we include two dynamic processes that are widely used in statistics: a linear time trend and serial correlation.

Using the "model parts' defined above, we can construct several alternative models – essentially the menu of models that will inform the decision about how best to produce multi-year indicators and associated standard errors. We have identified the following menu of models:

1. Static model (components 1 – 3)
2. Dynamic model: linear time trend model (components 1 – 4)
3. Dynamic model: serial correlation model (components 1 – 3, 5)

The bottom line is that given the measurement objective and the model of performance, the best procedure for aggregating multi-year data can be developed.

Figure 1. Sample size vs. shrunk VA estimates, Grade 4, 2015-16.

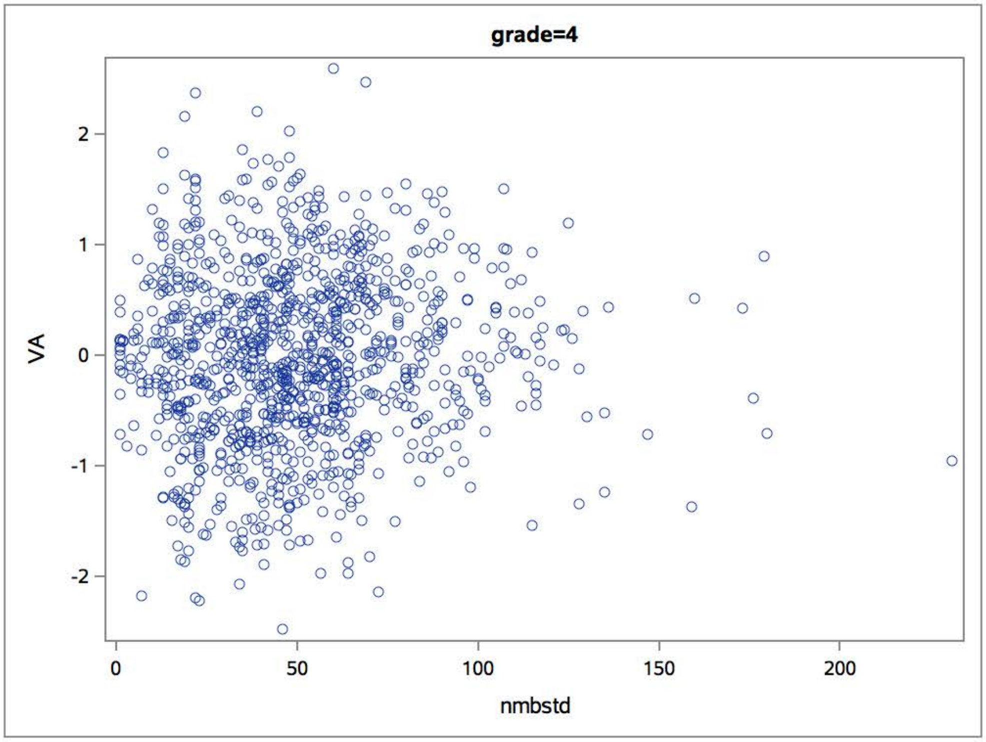

Figure 2. Sample size vs. shrunk VA estimates, Grade 5, 2015-16.

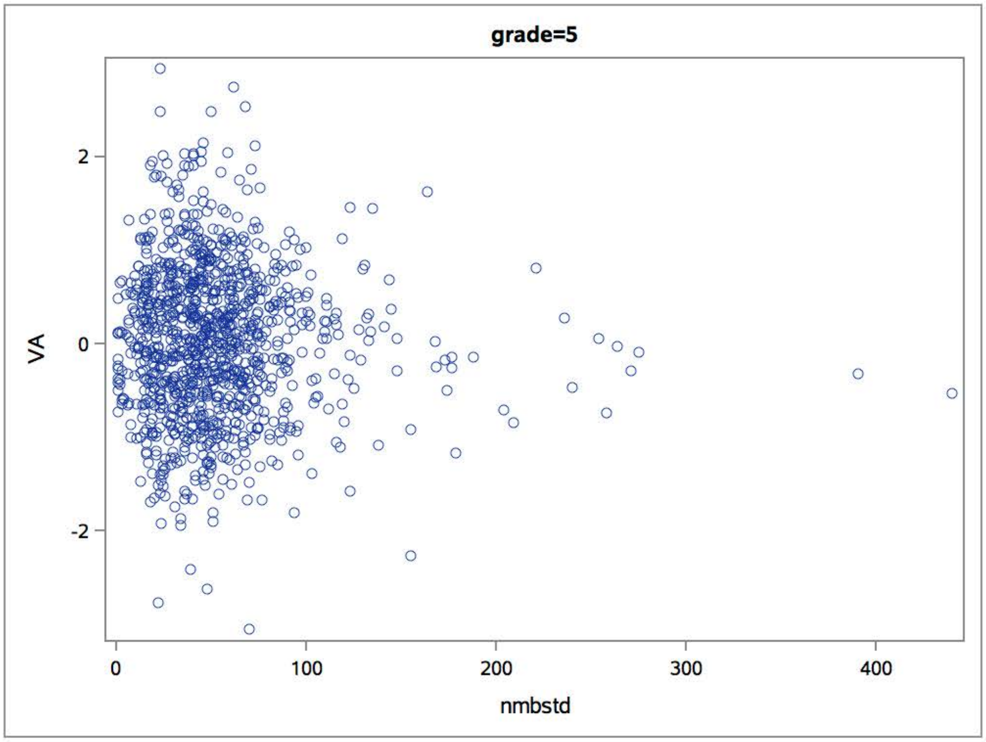

Figure 3. Sample size vs. shrunk VA estimates, Grade 6, 2015-16.