RESEARCH BRIEF

Research Services

Vol. 1404
Revised February 2015

Dr. Aleksandr Shneyderman
Dr. Terry Froman

2013-2014 District Approach to Teacher Evaluation (Student Growth Component): Technical Report

This report describes the process of using students' achievement results to calculate the teacher-level outcome measures that were used as part of the teachers' summative performance evaluations in 2013-2014. The statistical procedures that were used for the 2013-2014 teacher evaluations can be split into three categories: (1) those based on the State's FCAT and Algebra VAM, (2) those based on the District-created covariance-adjusted models, and (3) those based on the students' achievement or learning gains without using any statistical adjustments.

Procedures for Use with the State FCAT and Algebra VAM

The State's contractor produced the teacher- and school-level outcomes connected to the student results in reading in grades 4-10, mathematics in grades 4-8, and Algebra in grades 8-9. Those outcomes include Teacher VAM Estimates, School Components and their corresponding Standard Errors (SE) produced separately for each grade level and subject area. The Teacher VAM Estimates in each grade level and subject area were converted to points using the following procedure:

Calculate the values of t, the ratio of the Teacher VAM estimate to its SE; that is $t = \frac{VAM\ Est.}{SE}$

Assign points in accordance with the values of *t:*

- If $t < -2$, assign 12.5 points,
- If $-2 \leq t < -1$, assign 25 points,
- If $-1 \leq t \leq 2$, assign 37.5 points,
- If $t > 2$, assign 50 points

Because the Statewide averages for the Teacher VAM Estimates were close to zero, this rule is approximately equivalent to assigning 50 points to those teachers whose Teacher VAM Estimates were high, more than 2 Standard Errors above the average. At the other end of the

Research Services
Office of Assessment, Research, and Data Analysis
1450 NE Second Avenue, Suite 208, Miami, Florida 33132
(305) 995 7512 Fax (305) 995-7522

range, 12.5 points were assigned to those teachers whose Teacher VAM Estimates were relatively low, more than 2 Standard Errors below the average.

Creating District Covariance-Adjustment Models

A one-level *Covariate-Adjusted Model* (multiple linear regression model) was created for each of the outcome measures. In this model, the students' test scores on certain assessments at the end of the 2013-2014 year were used as an outcome. The list of these assessments is presented later in this document. Because some of these assessments (EOC Assessments, SAT, and ACT tests), are administered several times during an academic year, the highest score achieved on a particular subject area test during an academic year was used as an outcome in a regression model.

Use of Covariates

The students' test scores at the end of the prior school year (or at the beginning of the current school year in a case of students in grade K) were used as covariates. For most outcome measures, direct academic pretests (such as FCAT 2.0 Reading in grade 3 serving as a predictor for FCAT 2.0 Reading in grade 4) were available. For others, related academic pretests were available (Algebra 1 EOC as a predictor for Geometry EOC Assessment results, for example). In some cases, only the pretests that can be described as "cognitive predictors" were available (such as the College Board's Preliminary SAT as a predictor for Postsecondary Education Readiness Test [PERT]). Even in these cases, it can be argued that the inclusion of such cognitive predictors in the covariance-adjusted model helps to adjust the outcomes for differences in initial levels of learning and cognitive abilities across classrooms compared.

Certain student demographic characteristics were used as covariates as well. All such covariates were dichotomized. A relative age indicator was used to adjust the outcomes for retained students. In addition, gifted status, status of a student as an English language learner (ELL), and status of a student having any identified primary exceptionality other than gifted (SPED status) were used as covariates. The use of covariates served as an attempt to "level the playing field" for teachers who had different students beginning a school year at different achievement levels. Because the State law known as the Student Success Act prohibits the use of student race or poverty indicators in teacher evaluations, they were not used in the model.

Model-Building Steps

A separate multiple linear regression model was constructed for each of the 2014 outcome measures and grade levels. Because some of the grade 3 students (those who have been previously retained) had Grade 3 FCAT results as a pretest while others had Grade 2 SAT results as a pretest, separate regression models were used for these two student groups.

When fitting a regression model, all covariates were entered into a model initially. Then, using a backward elimination process, the variables that did not contribute significantly to the statistical prediction of the outcome (if any) were removed. Some important covariates (such as ELL status) were retained in the model even if they were not statistically significant predictors of the outcome. The details on the regression models are provided in the Appendix 1.

Converting the Results of the District Covariance Adjustment Model

Once the regression model for a particular outcome was finalized, the outcome score predicted by the model was determined. Then, for each student a residual score was calculated as the difference between the actual 2014 score and the model-predicted score. These residuals were aggregated to the teacher-level. In addition, standard errors of these mean residuals were found. The teacher-level mean residuals and their standard errors were used to assign points to teachers in the same way as described above for the State VAM, except that the Teacher VAM Estimate was replaced by the mean residual.

Working with the FAA, AICE, AP, IB and IC Outcomes

Computing "Learning Gains" with FAA Results

For each student who participated in the two consecutive administrations of the *Florida Alternate Assessment* (FAA), a dichotomous variable indicating whether a student made a learning gain was determined using the State methodology. A learning gain could be made in one of three ways: (a) increasing an achievement level from the prior year, (b) maintaining a level 1-3 and increasing the exam score by at least five points, or (c) maintaining a level 4-9. For each teacher, the percentage of his/her students making learning gains was determined.

Computing Percentages of Students "Passing" AICE, AP, IB and IC Examinations

For each student who participated in the AP, IB, or AICE assessment a dichotomous variable indicating whether a student had "passed" the exam was computed. Passing meant achieving a score of 3 or higher on an AP exam, a score of 4 or higher on an IB exam, or scores of E through A (or A*) on an AICE exam. (AICE scores were reported as numeric values, where scores of at least 3 signified passing an exam.)

Then, for each teacher of an appropriate course, the percentage of students passing one of the three exams was computed. Subsequently, the weighted average of these percentages was found for each teacher within each of the following eight subject areas:

- Science
- Social Science
- English Language and Literature
- Visual and Performing Arts and Music
- Spanish Language and Literature
- Other Foreign Language and Literature
- Mathematics

- Probability and Statistics, Computer Science
- Other.

The list of courses classified in each of these areas is given in Appendix 3. In some cases, students taking AP courses participated in both the AP exams and EOC assessments. In these cases, only the student results on the AP exams were used in the District models.

The percentages of students passing any Industry Certification (IC) exam were considered separately.

Converting the Results of the FAA, AICE, AP, and IB Analyses

Once the percentages of students making learning gains or passing exams (as defined above) were found, the distributions of such percentages were constructed and the mean and standard deviations were determined. The teacher-level percentages of students making learning gains on the FAA, passing AP, IB, or AICE exams within each of the broad subject areas described previously, or passing an IC exam were converted to points. This was done in a way similar to the one described above for the State VAM, except that the Teacher VAM Estimate was replaced by the difference between teacher-level percentages and their Districtwide mean. Because these models did not take into account student demographic and academic characteristics, the following safeguards were used to prevent the assignment of the lowest number of points to teachers when a reasonable standard was met and to assign the maximum number of points when a high standard was met.

1. The teachers whose students met or exceeded the targets below were assigned 50 points even if the previous calculations resulted in a smaller number of points. The targets for the percentage of students passing the exams were 95% or above for Foreign Languages (including Spanish) and 75% or above for all other areas including Industry Certification.
2. At the lower end of the scale, the target of 5% or above passing rate for all areas was used. A teacher in any area could not be assigned 12.5 points if his/her students met or exceeded that target; instead, 25 points were assigned.

Summary of the State and District Models

State FCAT and Algebra Models

Grade	Outcome	State Model
4-10	Reading FCAT	State FCAT VAM
4-8	Mathematics FCAT	
8-9	Algebra EOC Assessment	State Algebra VAM

District Covariance-Adjustment Models

Grade	Outcome	Academic Covariates	Demographic Covariates
K	Stanford Early School Achievement Test (SESAT) Reading and Mathematics	Florida Assessments for Instruction in Reading (FAIR)	ELL Status Gifted Status SPED Status Relative Age
1-2	Stanford Achievement Test (SAT) Reading and Mathematics	SESAT/SAT Reading or Mathematics	
3	Florida Comprehensive Assessment Test (FCAT) 2.0 Reading and Mathematics	SAT Reading and Mathematics FCAT 2.0 Reading or Mathematics for students repeating Grade 3	
5, 8	FCAT 2.0 Science	FCAT 2.0 Reading	
7	End of Course (EOC) Civics	FCAT 2.0 Reading	
8-10	EOC Geometry	Algebra EOC	
8-11	EOC Biology	FCAT 2.0 Reading	
11	EOC US History	FCAT 2.0 Reading	
11-12	SAT, ACT, Florida Postsecondary Education Readiness Test (PERT): Reading Components	PSAT Reading	

District Learning Gains and Achievement Models

Grade	Outcome	Model Type
4-11	FAA	Learning Gain
10-12	AICE, AP, IB, IC	Achievement

Procedures for Creating Student Learning Outcomes for Teachers of Other Subjects

Teachers of certain subject areas, such as Art, Music, Physical Education, and other areas may not teach any courses associated with the outcomes of assessments described above. To enable the usage of the student-learning data for the purposes of teacher evaluation, the reading outcomes of the State and local models described above were used for students in these teachers' classrooms. In particular, the outcomes of the local SAT model for students in grades K-2, FCAT model for students in grade 3, Florida VAM for students in grades 4-10, and local PERT, SAT, ACT models for students in grades 11-12 were used. The model results were converted to points as described previously.

Calculating Points based on the Schoolwide Data

The School Components and their Standard Errors resulting from the 2013-2014 Florida VAM in reading and mathematics/algebra were used to calculate the points separately for each of the grade levels using the same procedure as described previously except that the School Components were used instead of the Teacher VAM Estimates. If a school had no grade levels included in the State VAM, the reading and mathematics results of the Distcrict models were used. Subsequently, these points were aggregated (via a weighted average with the numbers of teachers as weights) to a school level. The higher number of points in reading vs. mathematics/algebra was assigned to instructional personnel with schoolwide responsibilities.

Combining Points from Different Sources

For teachers of courses associated with the State and District models described previously, all available points were calculated for each grade level and subject area they taught in 2013-2014. As mentioned previously, for teachers who did not teach any courses associated with the Florida VAM or District Model outcomes, reading results of students in their classrooms were used. Finally, for instructional personnel with schoolwide responsibilities, the points based on the school-level reading outcomes from the State or District Model were used. Once the points from various sources were determined, they were aggregated for each teacher by computing the weighted average of all points using the numbers of students as weights.

Special Consideration

In some cases, the number of students whose assessment results were used for calculation of points was small, meaning that the amount of information about the teacher's effect on student learning was limited. If the total number of students whose assessment results were used for calculation of points was less than 10, then the aggregated number of points were compared with the schoolwide number of points and the larger of the two numbers was assigned.

In addition, some employees had Districtwide responsibilities with no student or schoolwide assessment results linked to them. In these cases, the Districtwide value of 37.5 points was assigned.

Linking Teachers with Students

The academic records of only those students who have been in the same school during both October and February FTE periods in the 2013-2014 academic year were used. The February 2014 course file was used to link student and teacher records.

Appendix 1

Details on the District Covariance Adjustment Models

A typical regression equation used in the District Covariance Adjustment Model had the following form:

$$Score_{2014,i} = \beta_0 + \beta_1 * Score_{2013,i} + \beta_2 * ELL_i + \beta_3 * Gifted_i + \beta_4 * SPED_i + \beta_5 * Rel.Age_i + e_i$$

Where $Score_{2014,i}$ is the score of student *i* on a particular 2014 assessment, β_0 is the overall intercept, coefficients β_1 through β_5 represent covariate slopes, and $e_i$represents the residual of student *i*.

This model was fitted to the 2013 and 2014 data for each grade level and subject area separately. Across all models used with core subject areas, the coefficient of multiple determination R^2 varied from a low of .327 for reading in grade K to a high of .760 for reading in grade K. The median value of R^2 was .518, indicating that about 52% of the variance in the outcome was accounted for by the covariates used in the typical model. The model with the median value of R^2 square was the model for geometry in grade 9.

To gauge the effectiveness of using the covariates to "level the playing field" for teachers who have different compositions of students in their classrooms, one can consider the correlation coefficients between the outcome resulting from that model (a residual at the student level or the mean residual at the teacher level) and various student characteristics presented below. These correlations are presented below for the student and teacher levels.

Student-Level Correlations

Covariate	Correlation Coefficient
ELL Status	-.001
Gifted Status	.002
SPED Status	.002
Relative Age	-.004
Prior Year Algebra Score	-.002

It can be seen that the values of the correlation coefficients at the student level are very close to zero, indicating that the inclusion of the covariates in the model achieved the goal of adjusting the outcome for the students' differing academic and demographic characteristics represented by the covariates.

Teache- Level Correlations

Covariate	Correlation Coefficient
ELL Status	-.130
Gifted Status	.210
SPED Status	-.078
Relative Age	-.181
Prior Year Algebra Score	.170

The values of the correlation coefficients at the teacher level, although larger in absolute value than those at the student level, are relatively small. That indicates that the effects of covariates on the outcome at the teacher level (different proportions of ELL students in teachers' classrooms, for example) were likely relatively small.

The issue of nonzero correlations at the teacher level could be addressed by the use of multilevel models in which the aggregates of the first-level covariates are used as predictors at the second level of the model. This would require the use of substantially more complicated statistical models similar to the ones used by the State's contractor. Their use, however, would come at the expense of the "explainability" of their results. In addition, we have some evidence that the teacher-level outcomes coming from the simpler District models are not very different from those coming from the Florida VAM in the vast majority of the cases. This evidence will be presented in Appendix 2.

The student poverty level was not taken into account in the District models because State law expressly prohibits it. For completeness, we present the correlation between the model outcome and the student status as a participant in the federal free or reduced price lunch program. At the student level for a model with the median value of R^2 (geometry in grade 9), this correlation was -.066, and at the teacher level it was -.173.

District Covariance Adjustment models used with different subjects or different grade levels had different values of correlation coefficients between the free/reduced price lunch status and the outcome. However, the general pattern of having almost no correlation at the student level and relatively small correlation at the teacher level was present with all such models.

Appendix 2

Concurrent Validity of District Covariance Adjustment Models

In an effort to ascertain the concurrent validity of the District Covariance-Adjustment Model, we applied it to the data that were used by the State contractor when implementing the Florida VAM. (For this investigation, we ignored the cases where a student was taught by more than one teacher in a particular course.) We constructed the measures analogous to what the Florida VAM calls the Teacher Effect and the School Component by calculating the student average residuals for teachers and schools and then finding the deviations of teacher-level mean residuals from the school means. That deviation served as a measure analogous to the Teacher Effect, while the school-level mean residual served as an analog to the School Component under the Florida VAM. We then added one-half of the "District School Component" to the "District Teacher Effect" to mimic the process used in the creation of the Teacher VAM Estimate under the Florida VAM.

We then examined the correlations between the student-, teacher-, and school-level outcomes from the District and State models. First, the correlations between the predicted scores resulting from the District and State models were very high. The values of the correlation coefficients varied between .98 and .99 for reading outcomes in grades 4-10 and between .97 and .99 for mathematics in grades 4-8. These high values demonstrate that the outcomes from the District and State models were very similar at the level of the student residual, which was the basic building-block level for the teacher and school outcomes.

At the teacher level, we calculated the correlations between the District and State Teacher VAM Estimates for teachers who had at least 10 students. In reading, the correlations varied from a low of .73 in grade 10 to a high of .91 in grade 4 with a median of .77. In mathematics, the teacher-level correlations varied from a low of .68 in grade 8 to a high of .95 in grade with a median of .90.

At the school level, the correlations between the District and State School Components varied from .70 to .93 in reading (with a median of .82) and between .75 and .95 in mathematics (with a median of .88). Although these teacher- and school-level correlations were not as high as those at the student level, they were well above the level of .50 generally considered as the minimum for demonstrating a concurrent validity of two instruments.

This demonstrates that the outcomes from the District Covariance Adjustment Model are sufficiently similar to those resulting from the Florida VAM, at least when the District Model is applied to the reading data for students in grades 4-10 or mathematics data for students in grades 4-8.

Appendix 3

Grouping of AP, IB, and AICE Courses

Area	Course Title
Science	Advanced Placement Biology
	Advanced Placement Biology, Gifted
	Advanced Placement Chemistry
	Advanced Placement Chemistry Gifted
	Advanced Placement Environmental Science
	Advanced Placement Physics B
	Advanced Placement Physics B Gifted
	Advanced Placement Physics C: Mechanics
	Advanced Placement Physics C: Mechanics, Gifted
	AICE Biology 2
	AICE Biology 1
	AICE Chemistry 2
	AICE Chemistry 1
	AICE Environmental Management
	AICE Marine Science 1
	AICE Marine Science 2
	AICE Physics 2
	AICE Physics
	Biology 2-IB
	Biology 3-IB
	Chemistry 2-IB
	Chemistry 3-IB
	Physics 2-IB
	Physics 3-IB
Social Science	Advanced Placement American History
	Advanced Placement Art-History of Art
	Advanced Placement Comparative Government and Politics
	Advanced Placement European History
	Advanced Placement European History Gifted
	Advanced Placement Human Geography
	Advanced Placement Macroeconomics
	Advanced Placement Macroeconomics Gifted
	Advanced Placement Microeconomics
	Advanced Placement Psychology
	Advanced Placement United States History
	Advanced Placement United States History Gifted
	Advanced Placement United States Government and Politics

Area	Course Title
	Advanced Placement World History
	Advanced Placement World History Gifted
	AICE Geography
	AICE Inter History Gifted
	AICE Inter History
	AICE Psychology
	AICE Sociology 1
	AICE Sociology 2
	AICE U.S. History
	American History-IB
	Contemp. History - IB
	Economics 2-IB
	Economics I-IB
	History of America -IB
	Psychology 1
	Psychology 1-IB
	Psychology 2-IB
	Social Anthropology 1-IB
English Language and Literature	Advanced Placement English Literature and Composition
	Advanced Placement English Language and Composition
	Advanced Placement English Language and Composition Gifted
	Advanced Placement English Literature and Composition Gifted
	AICE English Language
	AICE English Literature
	AICE English Literature 2
	AICE General Paper
	English 4-IB
Visual and Performing Art, Music	Advanced Placement Art-Drawing Portfolio
	Advanced Placement Music Theory
	Advanced Placement Studio Art Three-Dimensional Design Portfolio
	Advanced Placement Studio Art Two-Dimensional Design Portfolio
	AICE Art & Design Photo
	AICE Music 1
	Art Research 2-B-IB
	Art Studio 2-A-IB
	Dance 3-IB
	IB Film Studies 2
	Music 4-IB
	Theatre 2-IB

Area	Course Title
Foreign Language and Literature Spanish	Advanced Placement-Spanish Language
	Advanced Placement-Spanish Literature
	AICE Span Language
	AICE Span Literature
	Spanish 4-A-IB
	Spanish 4 -B-IB
	Spanish 5-A-IB
	Spanish 5-B-IB
	Spanish 6-IB
Foreign Language and Literature non-Spanish	Advanced Placement French Language
	Advanced Placement Italian Language and Culture
	Advanced Placement Chinese Language
	Advanced Placement German Language
	AICE French Language
	AICE French Language AS
	AICE Portuguese. Lang AS
	French 4-B-IB
	French 5-B-IB
	French 6 IB
	French-5-A-IB
	German 4-B-IB
	German 5-B-IB
	IB Italian 4
	IB Mandarin Chinese 4
	IB Portuguese 5 B
Mathematics	Advanced Placement Calculus AB
	Advanced Placement Calculus AB Gifted
	Advanced Placement Calculus BC
	Advanced Placement Calculus BC Gifted
	AICE Further Math
	AICE Math 1
	Analytic Geometry-IB
	Calculus-IB
	IB Calculus
	IB Math Analysis
	IB Math Higher Level
	IB Integral Differential Calculus
	Math Studies-IB

Area	Course Title
Computer Science, Probability and Statistics	Advanced Placement Computer Science A
	Advanced Placement Statistics
	AICE Math & Probability & Statistics 2
	AICE Math & Probability & Statistics 1
	Computer Studies – IB
	IB Design Tech 2
Other	AICE Business 1
	AICE Global Perspectives 1
	AICE Thinking Skills 2
	AICE Thinking Skills 1
	AICE Travel & Tourism 1
	IB Bus Management 1
	IB Bus Management 2
	Information Technology in Global Society IB 2
	International Finance & Law