Florida's Student Growth Implementation Committee (SGIC)

University of Central Florida

Orlando, FL

February 7 and 8, 2012

Meeting Agenda

Tuesday, February 7, 2012

- 8:30 am   Coffee
- 9:00 am   Welcome, Agenda Overview, Updates
- 9:30 am   Business Rules Revisited
- 10:15 am   Results of the Algebra I End-of-Course Value-Added Model Evaluation
- 10:30 am   Break
- Noon   Lunch on Your Own
- 1:15 pm   Results of the Algebra I End-of-Course Value-Added Model Evaluation
- 3:00 pm   Break
- 4:30 pm   Begin Discussion on Final Recommendations on Algebra I End of Course Model and Business Rules
- 5:00 pm   Adjourn

Meeting Agenda

Wednesday, February 8, 2012

- 8:30 am   Coffee
- 9:00 am   Identification of Commonly Used Assessments to Develop Optional Value-Added Model Frameworks
- 10:30 am   Break
- 10:45 am   Alternate Assessment Growth Model
- Noon   Lunch on Your Own
- 1:15 pm   Alternate Assessment Growth Model
- 2:00 pm   Standard Setting and State Board of Education Rule
- 3:00 pm   Next Steps
- 3:30 pm   Adjourn

Business Rules Revisited for FCAT Model

Element	Current Business Rule
"Full Year" Status of Students	Students are included in the model if they are reported in either Survey 2 (October count) or Survey 3 (February count).
English Language Learners (ELL)	Students are identified as ELL if they are enrolled in ESOL services (reported as LY) for two years or less. Student who may be receiving ESOL services after two years, or are in follow-up services remain in the model; however, they are not coded as ELL.
Students with Disabilities (SWD)	The model currently uses only the primary exceptionality in identifying the disability of the student.

Business Rules Revisited for FCAT Model

Element	Possible Changes to Explore
"Full Year" Status of Students	• Require a Survey 2-3 match for inclusion in the model. • Challenge with semester-based courses; block scheduling.
English Language Learners (ELL)	• Remove the two-year limitation from the ELL definition, coding all students who are currently receiving ESOL services (LY) as ELL for purposes of the model.
Students with Disabilities (SWD)	• Include other exceptionalities in the identification of disability type for students in the model so that multiple disabilities of a particular student would be accounted for in the model.

AIR®

Value-Added Model Analysis for Algebra I End-of-Course (EOC)

Objectives

- Describe the four Algebra I EOC model variants that were analyzed

- Examine the results of each model variant by applying the same evaluation criteria used to examine the FCAT models
  - Statistical criteria for model
  - Impact data

- Review and confirm business rules

- Recommend an Algebra I EOC value-added model for Commissioner approval

AIR®

Algebra I EOC Data Description

- Scores on the Algebra I EOC test range from 325 to 475.

- The Algebra I EOC test does not have a vertical scale connecting scores with the FCAT or another test.

- Any student taking an Algebra I course is required to take the Algebra I EOC test.

Student Participation

The table below shows the number of students by grade taking the Algebra I EOC test in 2011.

Grade	Number of test takers	Percent
12	1,116	0.60
11	3,655	2.00
10	14,208	7.80
9	98,820	53.95
8	53,019	28.90
7	12,313	6.70
6	47	0.03

Algebra I EOC VAM Descriptions

- **Model 1:** One year prior math FCAT achievement scores and covariates

- **Model 2:** Two years prior math FCAT achievement scores and covariates

- **Model 3:** Two years prior math FCAT achievement scores, one year prior reading FCAT achievement scores and covariates

- **Model 4:** Two years prior math FCAT achievement scores, two years prior reading FCAT achievement scores and covariates

AIR®

Algebra I EOC Inclusion Rules

- **Model 1:** Each student must have an "immediate" prior math score available.

- **Model 2:** Each student must have one of the two prior math scores available.

- **Model 3:** Each student must have one of the two prior math scores available and an immediate prior reading score.

- **Model 4:** Each student must have one of the two prior math scores and one of the two prior reading scores.

AIR®

Defining Prior Scores

This table shows the grade levels from which prior scores for students are drawn.

Current Grade	Reading Prior Grade 1	Reading Prior Grade 2	Math Prior Grade 1	Math Prior Grade 2
12	10	9	10	8
11	10	9	10	8
10	9	8	8	7
9	8	7	8	7
8	7	6	7	6
7	6	5	6	5
6	5	4	5	4

AIR®

Grade 9 and 10 Math

- Notice that grade 9 reading is used as a prior score but not grade 9 math. This is because the grade 9 math test is no longer administered.

- While included in this analysis, grade 10 math will not be included in future analyses because FLDOE has discontinued its administration also.

AIR®

Covariates in Algebra I EOC Model

- The number of subject-relevant courses in which the student is enrolled
- Up to two prior years of achievement scores
- Students with Disabilities (SWD) status (primary disability)
- English Language Learner (ELL) status (only if <2 years)
- Gifted status
- Attendance
- Mobility (number of transitions)
- Difference from modal age in grade (as an indicator of retention)
- Class size
- Homogeneity of entering test scores in the class (based on prior math score)

AIR®

Algebra I EOC Model Characteristics

- One important difference between FCAT and EOC models is that the prior scores for EOC are taken from a different test altogether.

  - Prior scores may be related to the EOC but are designed to measure different things (e.g., math FCAT measures more than just Algebra).

Algebra I EOC Model Characteristics

- The Algebra I EOC model has a notable difference from the FCAT model in that the time between the current score and the prior scores varies by student.

- This example shows that student A will have a different predicted score than students B and C.

Student in Grade 9 2012	Grade 8: 2011	Grade 8: 2010	Grade 8: 2009
A	Most students		
B		Some retained students	
C			Some, but fewer, retained students

Algebra I EOC Model Characteristics

- Model 2 is most similar to the Reading and Math FCAT value-added model.
    - Uses two prior scores in the same subject
    - Uses same student-level characteristics (same covariates)
    - Includes both a teacher and a school component

Summary of a Teacher VAM Score

- Each model forms a statistical prediction for each student based on his or her prior scores and other characteristics.

- The difference between the student's actual score and his or her predicted score is called a *residual*.

- A *positive residual* means a student did better than was predicted and a *negative residual* means a student did not perform as well as predicted.

- The teacher VAM scores are (mainly) based on these residuals.

- Teachers with high VAM scores are those whose students did better than predicted.

- Teachers with low VAM scores are those whose students did not perform as well as predicted.

Model Evaluation Criteria

The following slides present data relevant to the following questions:

- Which model produces the most precise teacher VAM scores?

- Which model's prior scores are most helpful in forming predictions?

- Are any student or teacher groups advantaged or disadvantaged under the various models?

AIR®

Precision of the Teacher Component

- **Question:** What characteristics of value-added models lead to more **precise** estimates of the teacher component?

- **Statistic to examine:** Standard errors of the teacher component.

- **Evidence in favor of a desirable model:** A model with smaller standard errors, other things being equal, is more desirable than a model with larger standard errors.

- **Why:** Smaller standard errors for estimated teacher components indicate that they are estimated with greater precision.

Precision and Certainty: Standard Errors for Algebra I EOC Models

Model 2 produces the smallest median standard error.

Standard errors in models 3 and 4 are larger than in model 2.

Conditional Standard Errors of Teacher Component

Standard Errors of Teacher Component
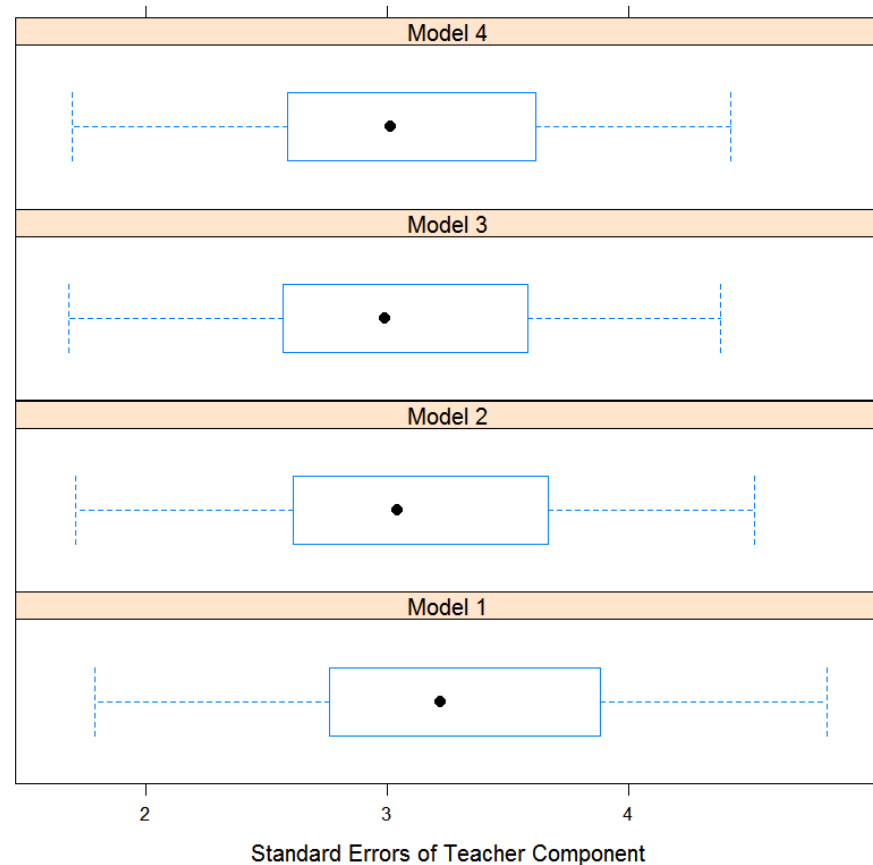
Magnitude of Teacher and School Component Standard Deviations

Standard deviations for teachers and schools are similar in magnitude across all models.

The school component is larger than the teacher component. This does not occur with FCAT.

This could suggest that there are large differences between middle and high schools that are pushed into the school component.

Teacher and School Standard Deviations

AIR®
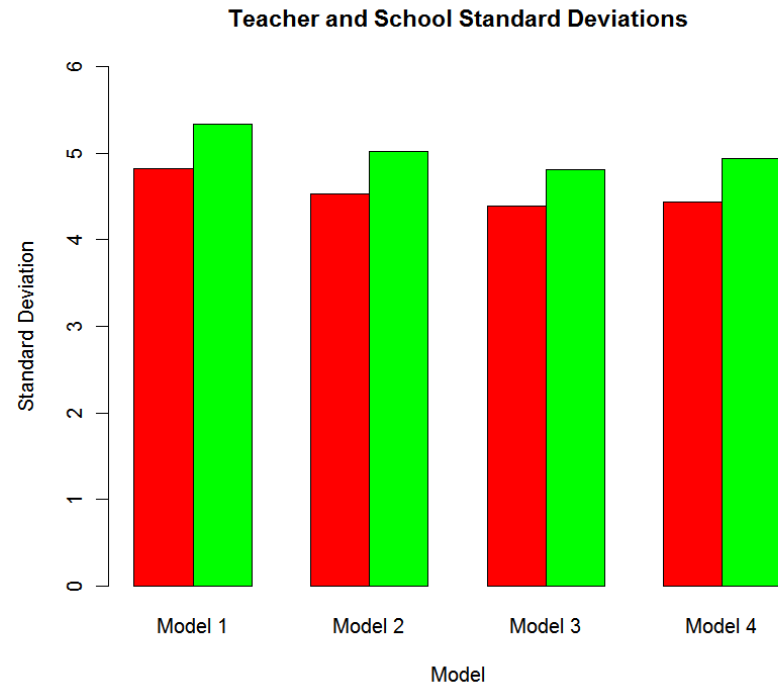
Algebra I EOC Model Parsimony

- **Question:** Does the model include control variables without being overly complicated?

- **Statistic to examine:** Percent of current year test score variance accounted for by control variables in models.

- **Evidence in favor of a desirable model:** High proportion of variance accounted for.

- **Why:** The model should not be needlessly complex.

Proportion of Variance in Current Year Test Score Explained by Control Variables

R-square is an overall measure of model fit.

Model 1 provides the smallest *R*-square.

The proportion of variance is approximately the same across all other models.

Model	*R*-Square
Model 1	0.41
Model 2	0.59
Model 3	0.59
Model 4	0.59

What Do the Data Suggest So Far?

- Model 2 produces the smallest mean standard error; adding the second math score increases R-square over one score.

- The R-square is largest (and equivalent) for models 2 through 4.

  - This indicates that the additional reading scores do not improve model fit beyond the two prior math scores.
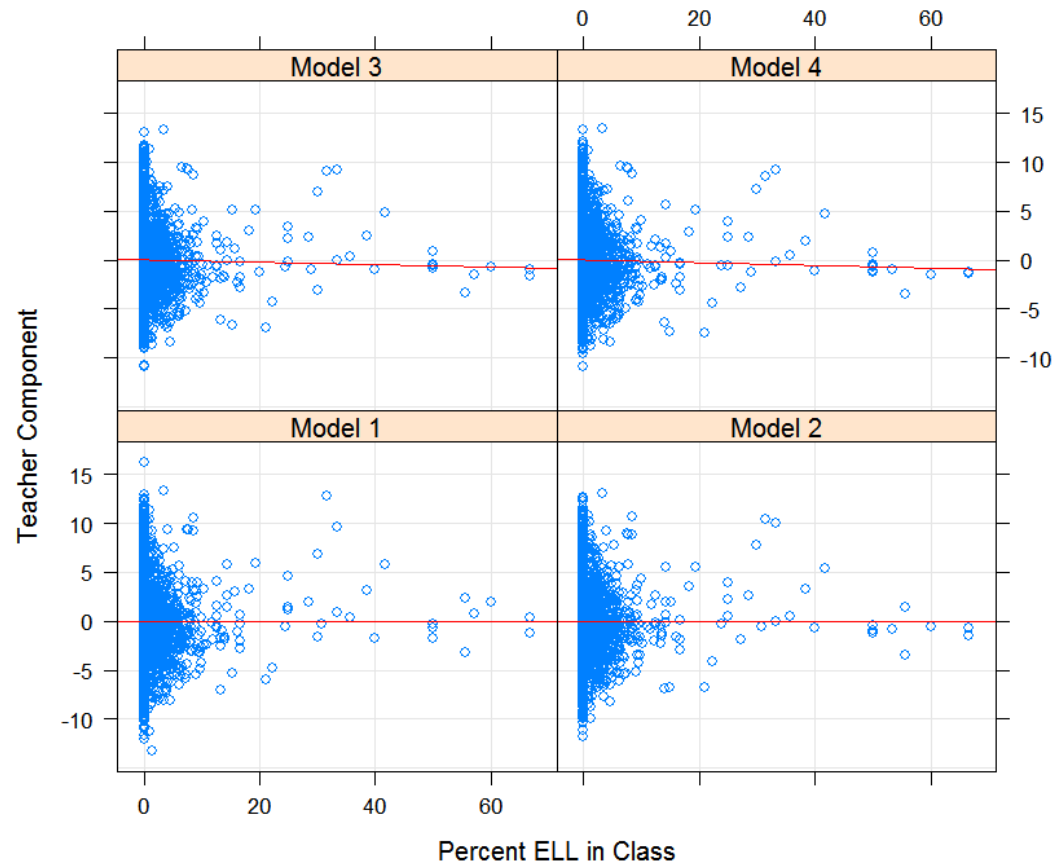
Correlation of Teacher Component and Percent Students with Disabilities

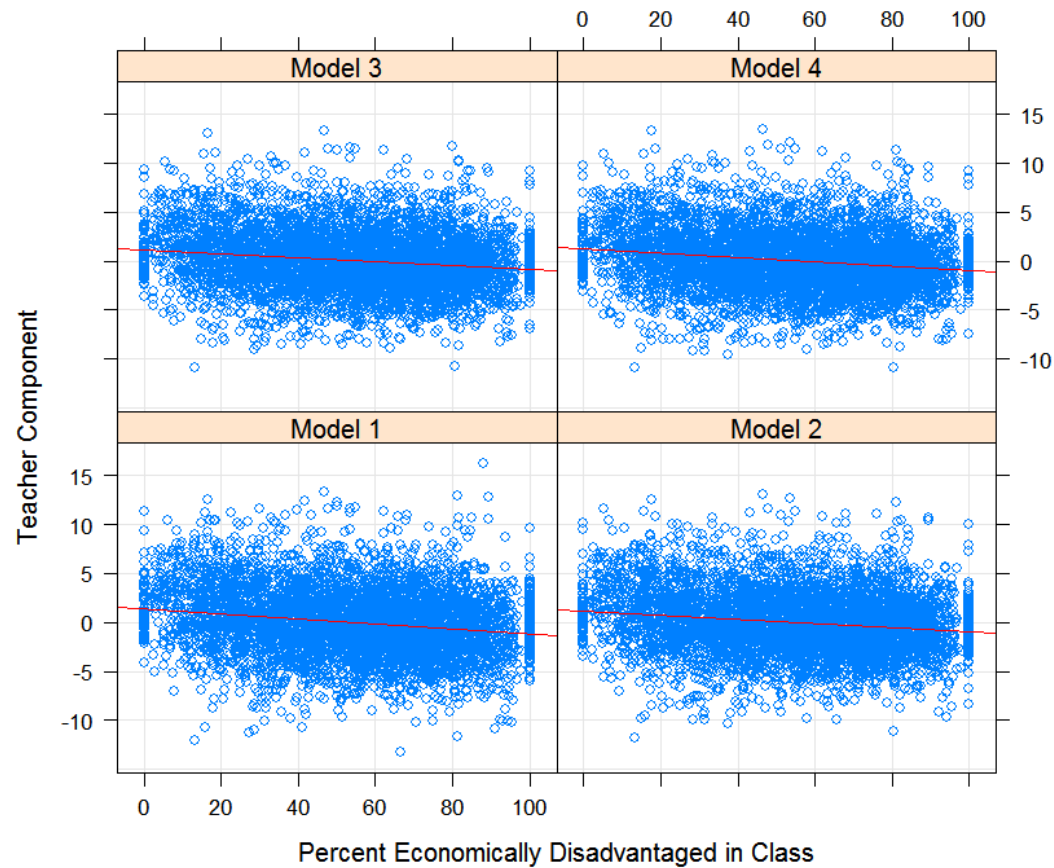**Correlation of Teacher Component with Percent SWD in Class**
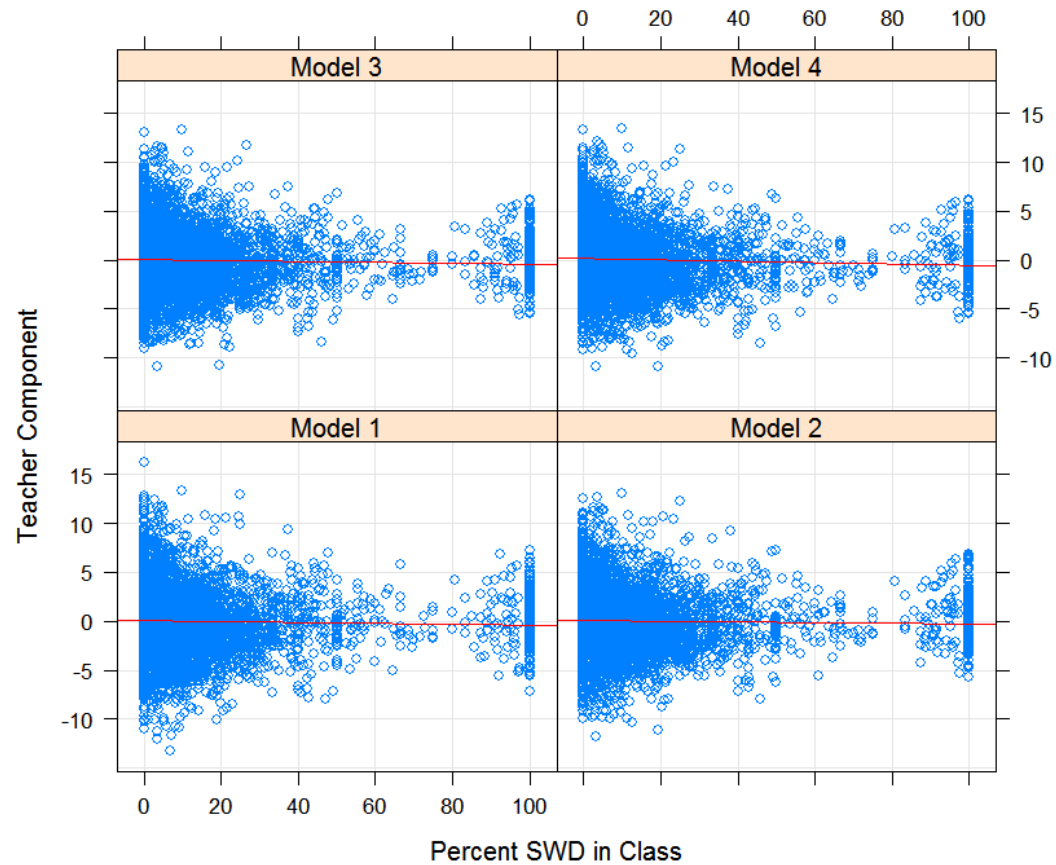
Correlation of Teacher Component and Percent English Language Learners

Correlation of Teacher Component with Percent ELL in Class
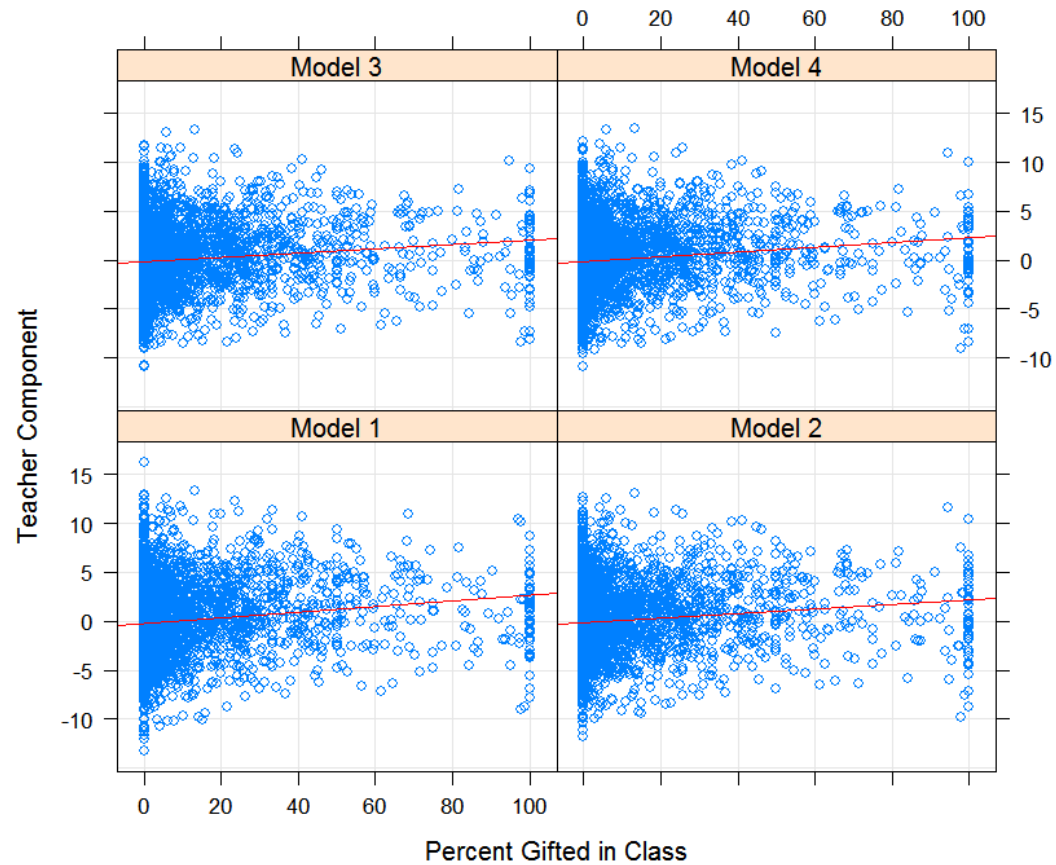
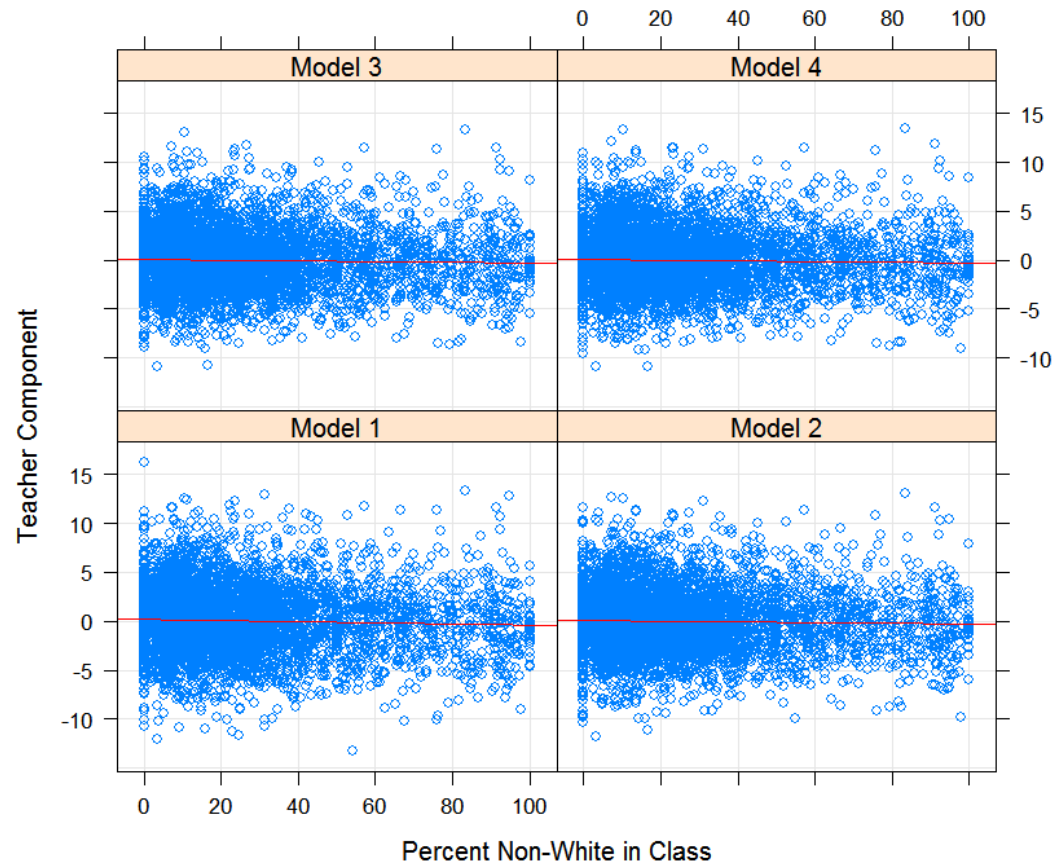# Correlation of Teacher Component and Percent Economically Disadvantaged

Correlation of Teacher Component and Percent Gifted

Correlation of Teacher Component with Percent Gifted in Class

Correlation of Teacher Component and Percent Non-White Students

Correlation of Teacher Component with Percent Non-White in Class

Distribution of Teacher Component by Modal Grade Taught

This plot shows that middle school teachers appear to have larger VAM scores than high school teachers.

It may be useful to consider running separate models for middle and high school.

Distribution of EOC Teacher Component by Grade
Model 2
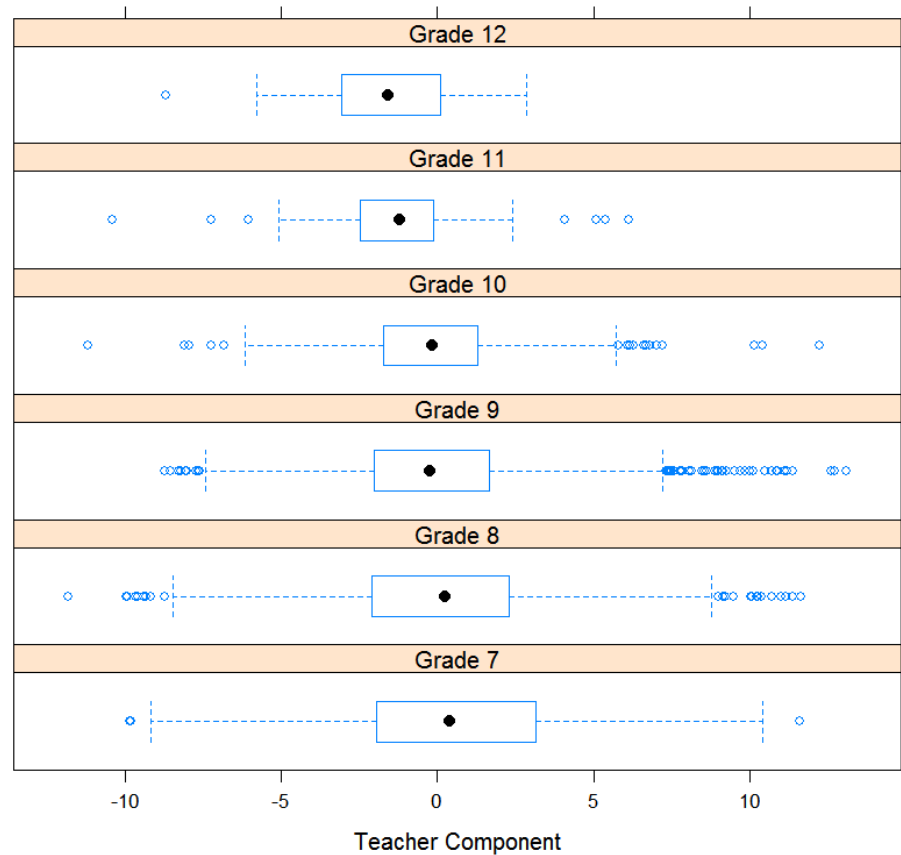
Evaluation Questions

We started with the following questions:

- Which model produces the most precise teacher VAM scores?

  - Answer: Model 2

- Which model's prior scores are most helpful in forming predictions?

  - Answer: Model 2

- Are any student or teacher groups advantaged or disadvantaged under the various models?

  - Answer: There are systematic relationships in teacher VAM scores and student characteristics that are not expected to exist in VAM models.

Summary of Impact Data

- The data suggest that there is a positive correlation with gifted status ($r = .13$).

  - The more gifted students in a class, the larger the teacher component.

- There is a negative correlation with economically disadvantaged (ED) students ($r = -.19$).

  - The more ED students in a class, the smaller the teacher component.

- Middle school teachers have larger teacher VAMs than high school teachers.

Summary of Impact Data

- Considering all data presented so far, the results suggest a selection effect remains.
    - This means that teachers may have higher value-added scores as a function of the students they taught.
- This selection effect seems to remain in part because the FCAT prior scores are not good control variables for how students may be sorted into Algebra I courses.

AIR®

Options to Control for Selection

- First, understand what methods districts use to determine when students take Algebra courses and the Algebra I EOC test.

- Are there any available measures that can be used as better pretests?

  - For example, are students ever required to take a prequalifying exam?

- Why would a better pretest help?

  - It would help to adjust for selection and sorting of students into EOC classes.

AIR®

Options to Control for Selection, (Cont)

- Can we consider any new implementation rules that may mitigate this effect?
  - For example, separate middle and high school models. Students in upper grades (i.e., grades 10 to 12) may be sorted into homogenous groups more intentionally than students in lower grades.

Floor and Ceiling Effects

- The Algebra I EOC test appears to have a "floor" and a "ceiling" effect.

- A floor effect is when there is a lower bound cutoff score that students cannot score below.

  - This is a problem because some students may have abilities that are lower than the test allows them to demonstrate.

- A ceiling effect is when there is a cutoff at the high end that cannot be exceeded.

  - This is a problem because some students may have abilities that are higher than what the test allows them to demonstrate.

How Do Floor and Ceiling Scores Affect VAM?

- Regression slope is altered, thus potentially affecting all teachers.

- Teacher component could have upward bias for low-performing teachers and downward bias for high-performing teachers.

- Low-performing teachers may have students receiving scores higher than what they may truly deserve.

- High-performing teachers may have students receiving scores lower than what they may truly deserve.

Discussion

- As a result of the floor and ceiling effects, some students have predictions outside the range of 325 to 475.

    - N = 6,254 at minimum score 325 (3.4%)

    - N = 332 at maximum score 475 (0.18%)

- The value-added literature has not yet proposed methods for censored data.

- AIR is currently investigating ways to best work with censored data with a VAM.

AIR®

Optional Value-Added Models

- AIR will develop VAMs that districts have the option to use with selected standardized assessments prevalently used in Florida.

- White paper explaining the optional VAMs.

- SAS code to compute optional VAMs.

- Data files created running the optional VAMs.

- Training and training materials for districts on the optional VAMs.

AIR

Possible Data Available for Optional Value-Added Models

- Advanced Placement and International Baccalaureate exam scores
  - Five-point scale may not differentiate enough for VAM
- Florida Assessment of Instruction in Reading (FAIR)
- ACT and SAT exam scores
- Postsecondary Education Readiness Test for 1th graders if scoring a certain achievement level on FCAT from FLDOE
- SAT 9/10 **from districts**

AIR®

Alternate Assessment Growth Model

- Description of current Florida Alternate Assessment (FAA) growth model developed by Measured Progress for school accountability purposes and its application for use in teacher evaluations in 2011–12.

- Future considerations/options regarding uses for teacher evaluations.

Standard Setting and State Board of Education Rule

Calculations of Student Learning Growth Using Statewide Assessment Data for Use in School Personnel Evaluations

- New Rule 6A-5.0411.

- Rule includes the growth formulas selected by the Commissioner.

- "Specific, discrete" student learning growth standards for each performance level:

  - **Highly Effective and Effective:** A standard that must be met in order for an employee to receive each rating, respectively.

  - **Unsatisfactory:** A standard that if not met will result in the employee receiving an unsatisfactory performance evaluation rating.

  - Commissioner must consult with experts, instructional personnel, school administrators, and education stakeholders in developing the criteria for the performance levels.

Breakdown of Timeline for Rule 6A-5.0411

Date	Event/Process
August 1–2, 2011	School district technical assistance meetings on value-added model and delivery of three years of historical data to each district.
September 16, 2011	Rule development noticed in Florida Administrative Weekly (FAW).
September 20, 2011	Presentation of timeline to SBOE for discussion and approval.
September 30, 2011	RTTT participating LEA final 2011–12 evaluation systems documents submitted.
October 2011	Analysis of district plans for student growth standards. Provide state assistance as needed for districts that are not successful in completing acceptable plans by the September 30 deadline.
December 2011	Proposed rule text for use in evaluations for 2011–12 school year noticed in FAW.
February–May 2012	Rule with statewide formula for use in 2011–12 presented to SBOE for adoption in February. Regional workshops on standards with school district assessment and personnel evaluation teams to receive input. ▪ Education on the standard-setting process. ▪ Input from educators and the public on content of proposed standards rule.
July 2012	Value-added results calculated for all teachers and principals provided to districts to complete their 2011–12 personnel evaluations. Final values for standards completed based on second year of FCAT 2.0 and Alg I EOC data; rule development workshop to be held via statewide conference call.
August 2012	Final rule with statewide standards for all performance levels presented to the SBOE for adoption. Standards applied to personnel evaluations for 2012–13 school year.

Questions and Next Steps

Information about the activities, membership, meeting schedule and materials, and recordings of conference calls and webinars of the SGIC are posted at www.fldoe.org/arra/racetothetop.asp.

AIR

Contact Information

Juan Copa

Director of Research and Analysis in Educator Performance

Florida Department of Education

Phone: 850-245-0744 (office)

Email: Juan.Copa@fldoe.org