Florida Value-Added Models

Student Growth Implementation Committee (SGIC)
February 27, 2013

AIR®
American Institutes for Research®

Presentation Outline

How models are designed and evaluated

- End-of-course exams:
  - Algebra I
  - Biology
  - Geometry
- Optional VAM
  - SAT-10
  - Advanced Placement (AP) Calculus AB
  - Advanced Placement (AP) English Language and Literature
- Next Steps

Structured Review Process

- Are the input data accurate and sensible?
  - Examine the descriptive statistics
  - Are there any red flags?
- Do the models behave as expected?
  - Examine the variance components
  - Examine R-squared to determine model fit
  - Precision of the value-added scores
- Do the results suggest advantages to certain groups?
  - Impact data based on correlations between value-added scores and class characteristics

Thoughts on Covariates

- Ideally, the predictor variables should have the following properties:
  - A high statistical correlation with the outcome
  - A high curricular relationship with the outcome
  - A correlation with factors that contribute to student learning but are not in the control of teachers and schools
  - A high correlation with the unobservable processes by which students are sorted into schools and classes
- If predictors do not fully capture selection effects, teacher and school value-added estimates may be biased.

Covariates Included in Most Models

- Prior test scores
- Students with Disabilities (SWD) status
- Gifted status
- English Language Learner (ELL) status (time as ELL)
- Attendance
- Mobility (number of transitions)
- Difference from modal age in grade
- Class size
- Homogeneity of entering test scores in the class
- Percentage in each grade, when appropriate
- Percent gifted in class
- Number of subject-relevant courses

End of Course Value-Added Model: Algebra I

AIR®
American Institutes for Research®

Algebra I Background Information

- Students are included only if they have a 2010–11 FCAT 2.0 math score available as a predictor variable.
- The model was run three times, each with a different subset of students:
  - Model 1a: Includes all students
  - Model 1b: Includes students in grades 6–8
  - Model 1c: Includes only students in grade 9

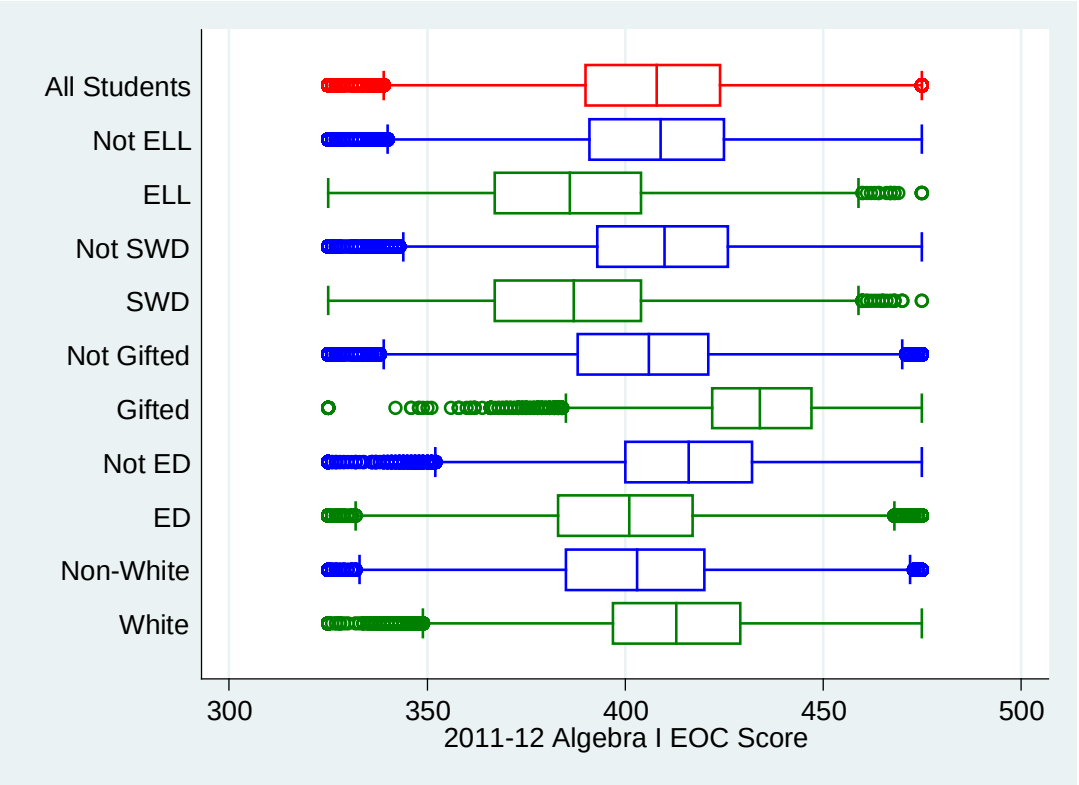
Number of Students per Model

Model	N
Model 1a (All Students)	155,581
Model 1b (Grades 6–8)	57,988
Model 1c (Grade 9)	97,593

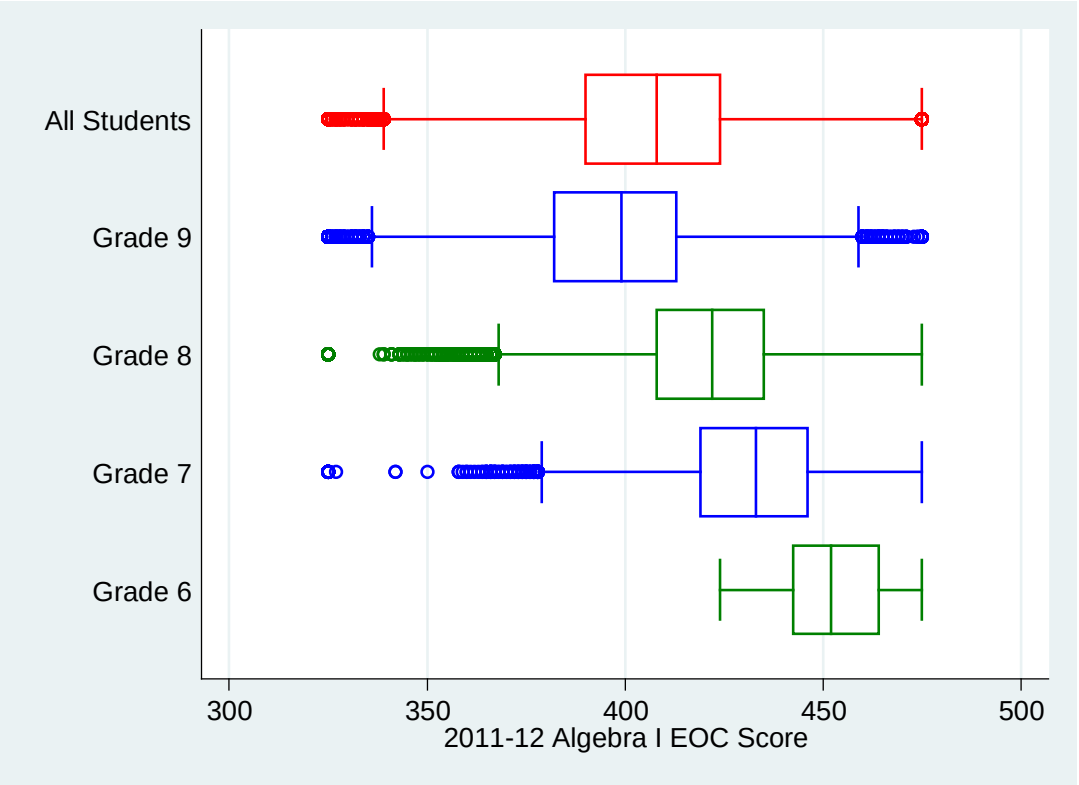
Descriptive Statistics

The following descriptive statistics are presented to show that the data seem reasonable and that observed patterns in the level scores are also observed in the value-added scores.

2011–12 Algebra I EOC Scores, Overall and by Subgroup

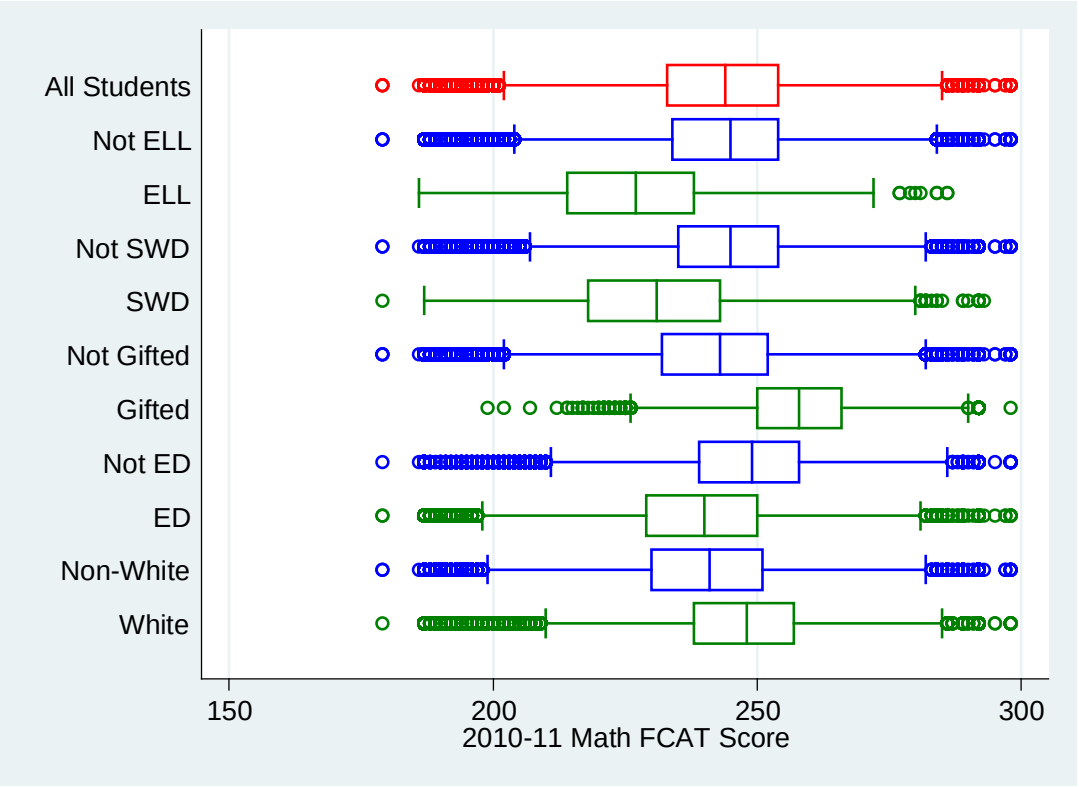


2011–12 Algebra I EOC Scores, Overall and by Grade

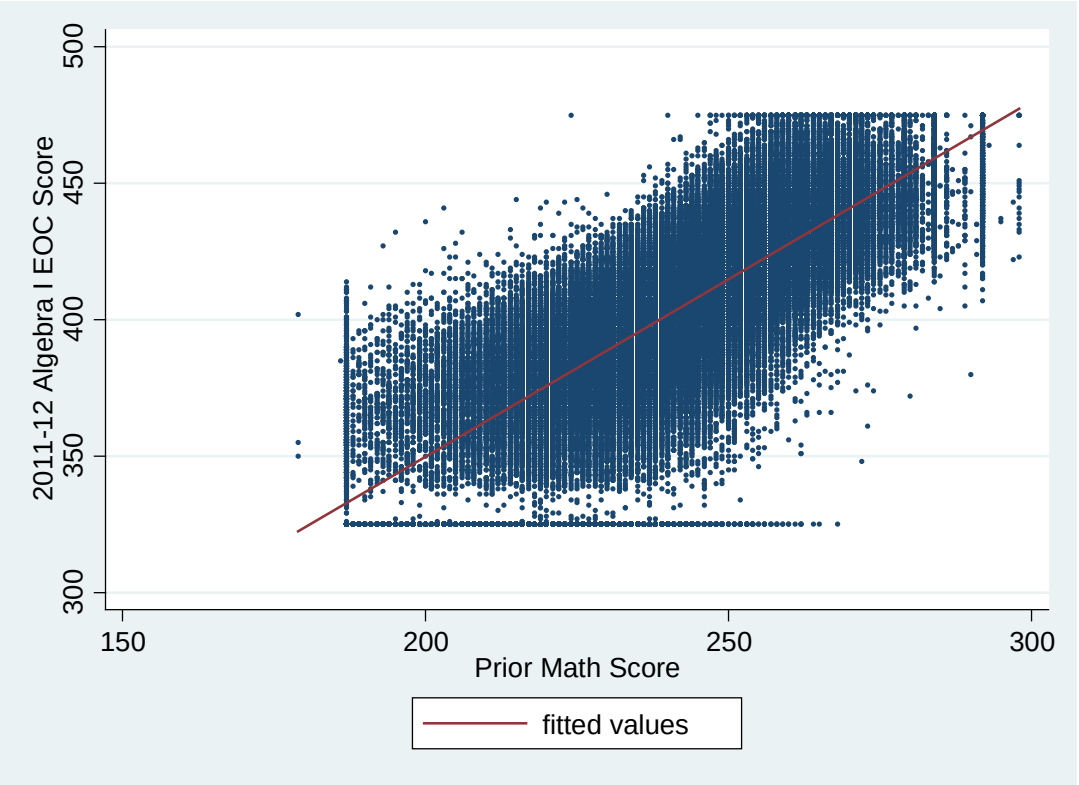


Algebra I

2010–11 Math FCAT Scores, Overall and by Subgroup



Algebra I EOC and Math 8 FCAT Scores (Correlation = 0.70)



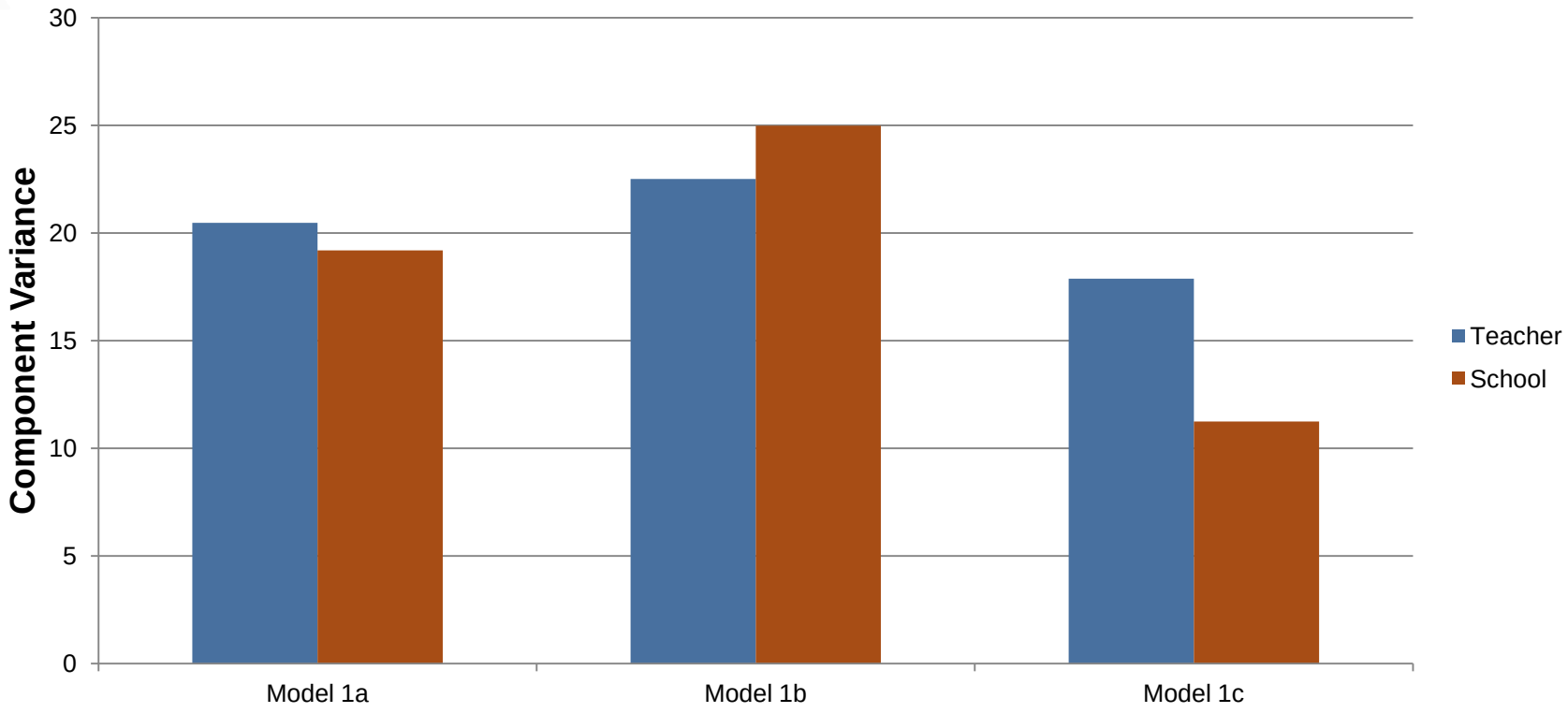
Summary of Descriptive Statistics

- The data show that students in lower grades score higher on the Algebra I EOC than students in the higher grades.
- There are large systematic differences between student groups.
- The correlation between the Algebra EOC and the Math 8 FCAT is 0.70.

Standard Deviations of Teachers and Schools

- The next slide shows the teacher and school standard deviations.
- The teacher component is typically expected to have more variability than the school component.
- The school component is larger than expected in two of the three Algebra I EOC models.

School-Level Variation Is Larger than Expected Relative to Teacher-Level Variation



The R-Squared Is One Indicator of Model Fit

The closer the value is to 1, the better the model predicts the outcome scores. Model 1a, which includes the most observations, provides the best fit of the data.

Model	R-Squared
Model 1a (All Students)	0.63
Model 1b (Grades 6–8)	0.53
Model 1c (Grade 9)	0.51

Both Models Are Able to Identify More and Less Effective Teachers

- Reliability Ratio numerator: How precise are the teacher estimates on average?
- Reliability Ratio denominator: What is the overall distribution of teacher estimates?
- Low ratio → Better able to distinguish among teachers on the basis of effectiveness

Teacher Reliability Ratios

Model	Ratio
1a (All Students)	0.90
1b (Grades 6–8)	0.89
1c (Grade 9)	0.95

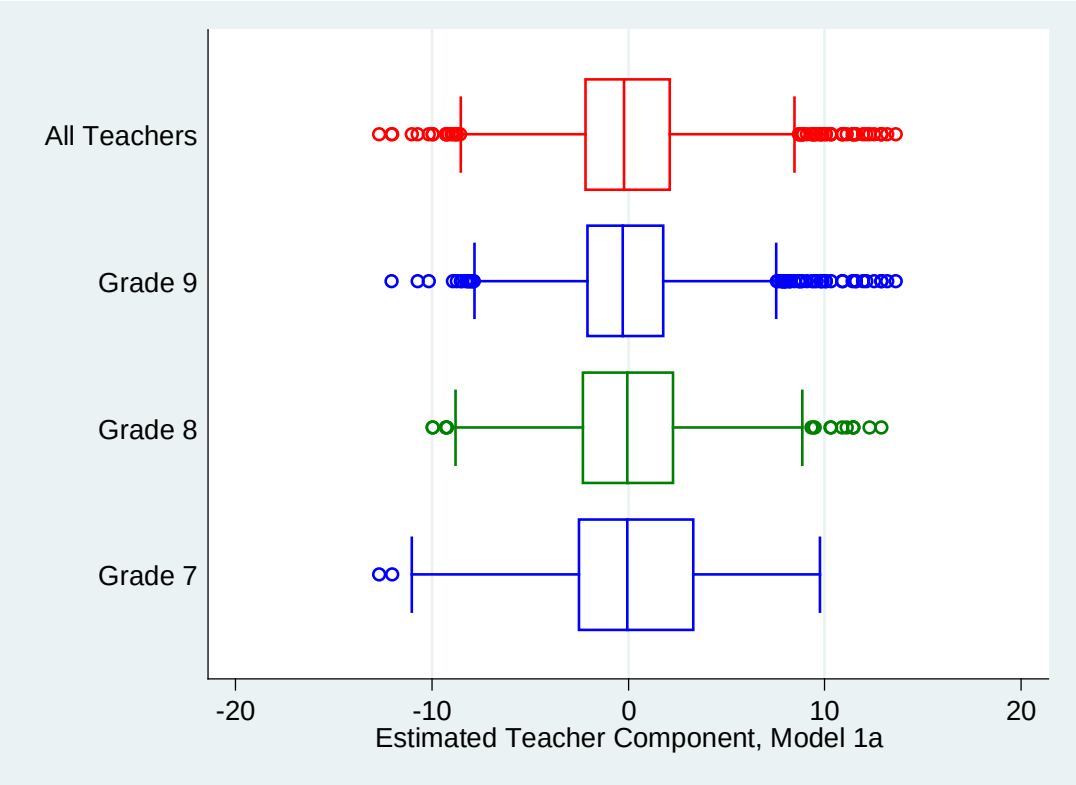
Algebra I

Percent of Teachers and Schools Significantly Different from Average

Model	Teachers (above and below)	Schools (above and below)
1a (All Students)	12%	14%
1b (Grades 6–8)	11%	14%
1c (Grade 9)	12%	11%

Algebra I

Teacher Component Estimates by Modal Grade in Class

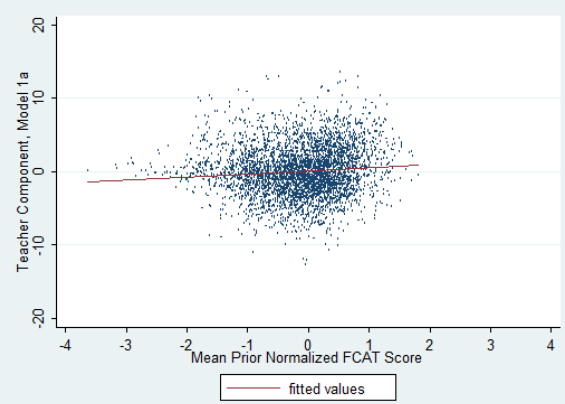


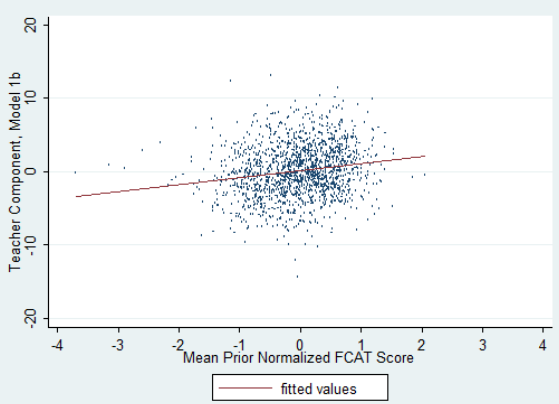
Impact Data Results

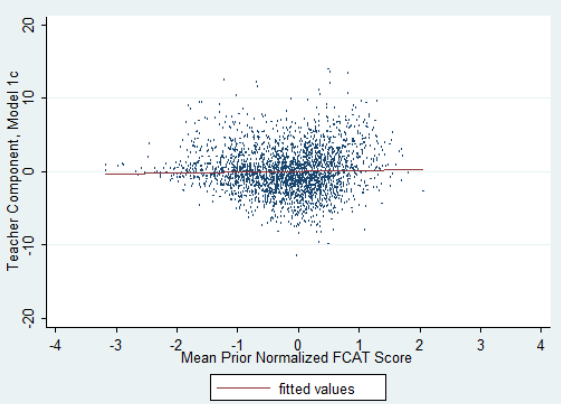
- Impact data slides show the relationship of the teacher score to various classroom characteristics.
- There are two ways to interpret a non-zero relationship:
  - Teachers are not distributed randomly across students.
  - Classroom characteristics affect the rate of student learning and lead to biased value-added estimates.

Algebra I

Teacher Component and Mean Normalized Prior Score

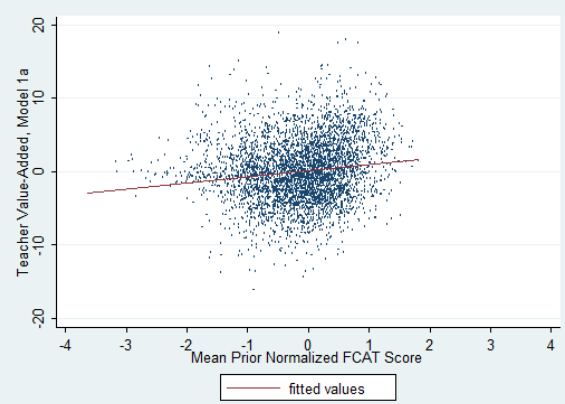


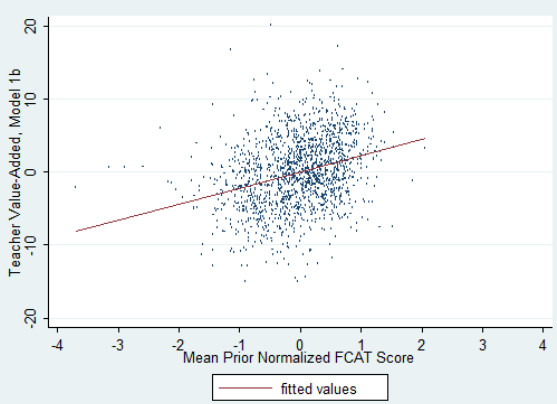


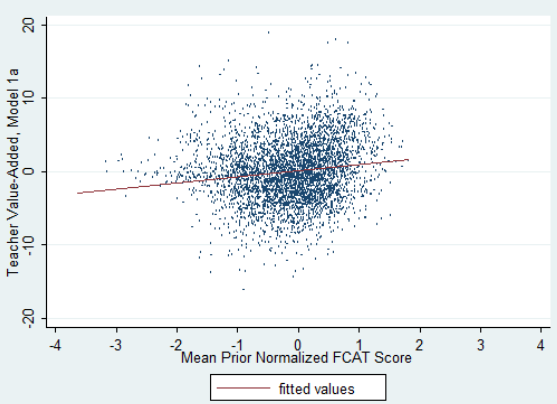


Algebra I

Teacher Value-Added and Mean Normalized Prior Score

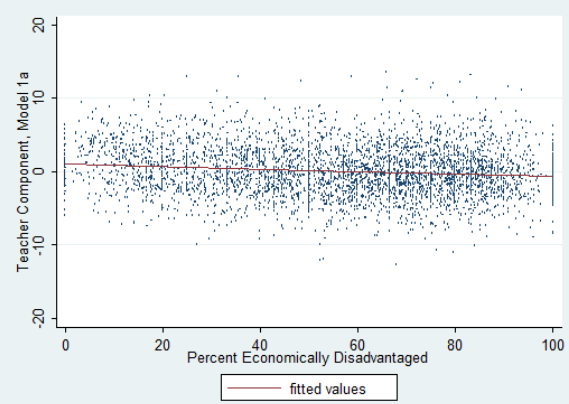


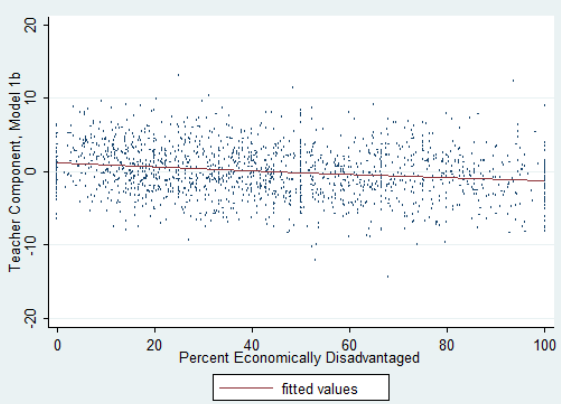


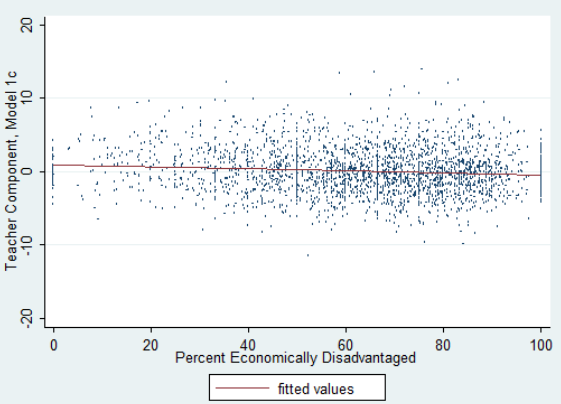


Algebra I

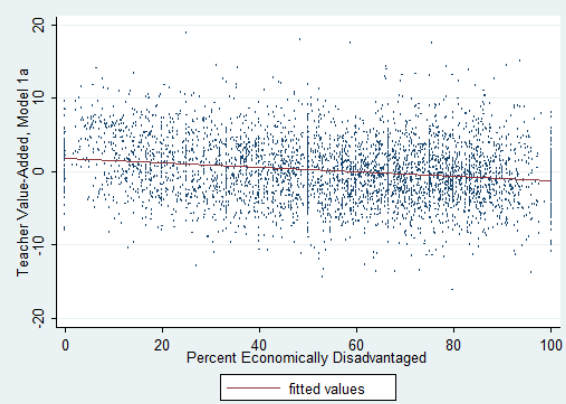
Teacher Component and Percent Economically Disadvantaged

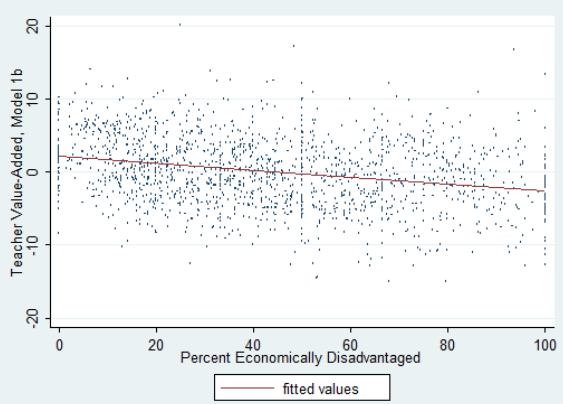


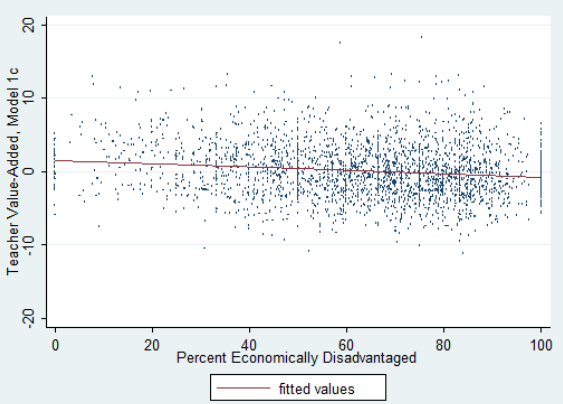




Teacher Value-Added and Percent Economically Disadvantaged

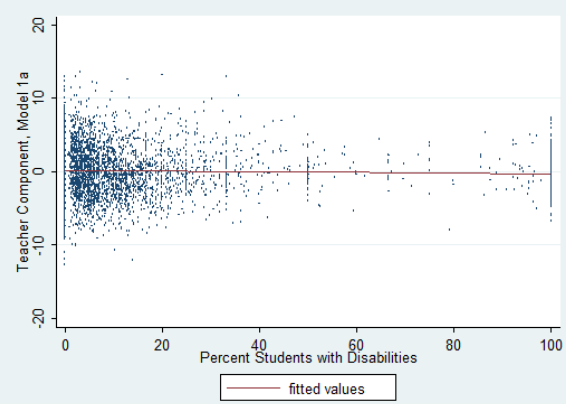


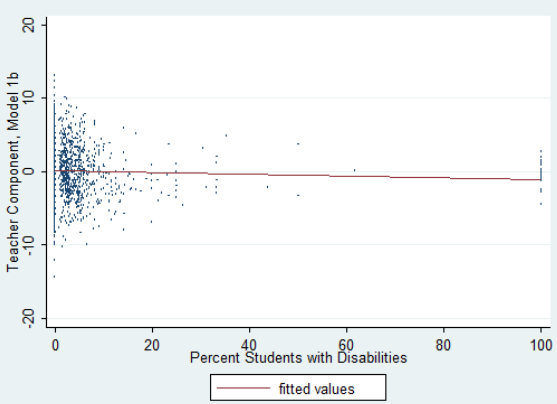


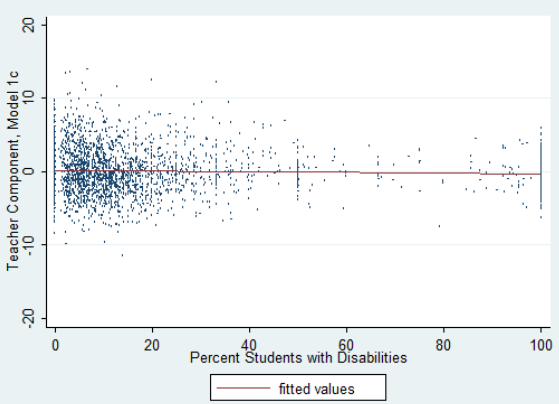


Algebra I

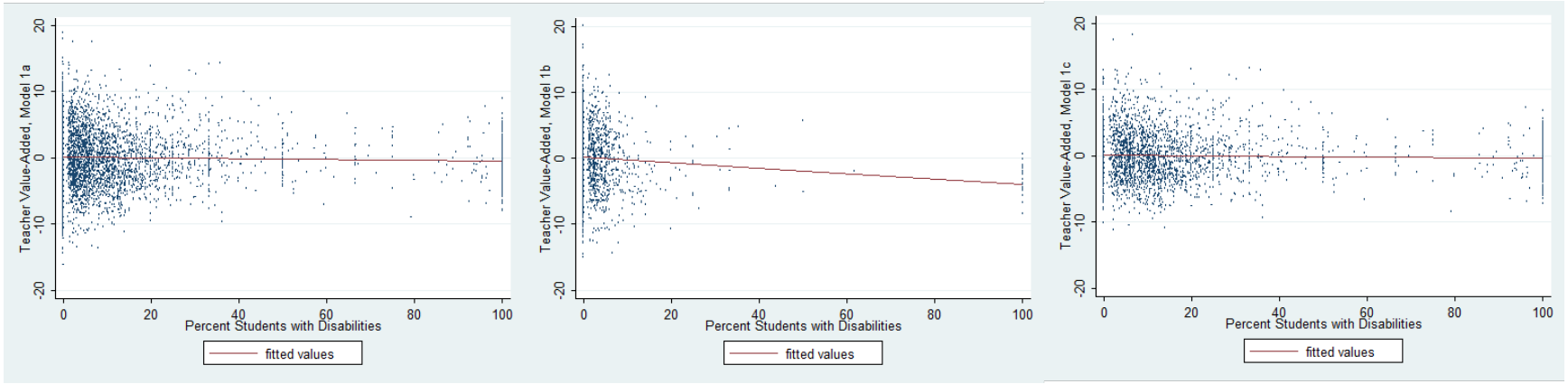
Teacher Component and Percent Students with Disabilities





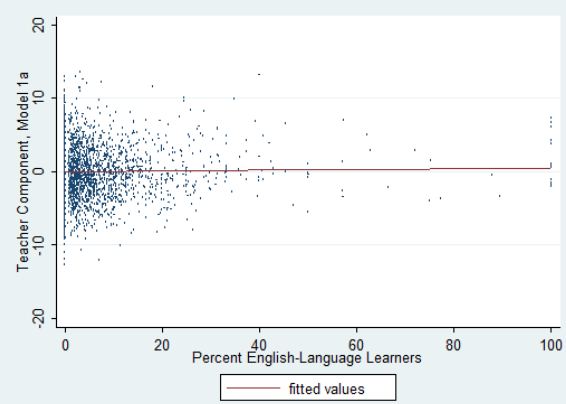


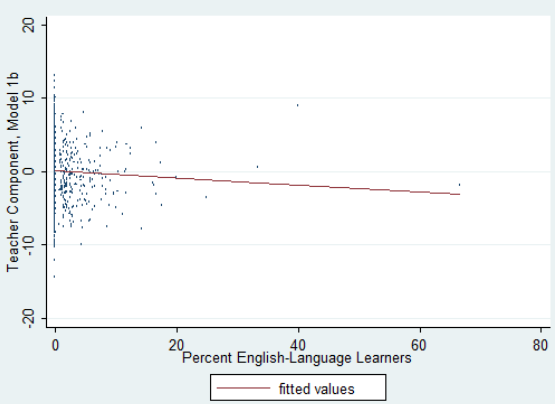
Teacher Value-Added and Percent Students with Disabilities

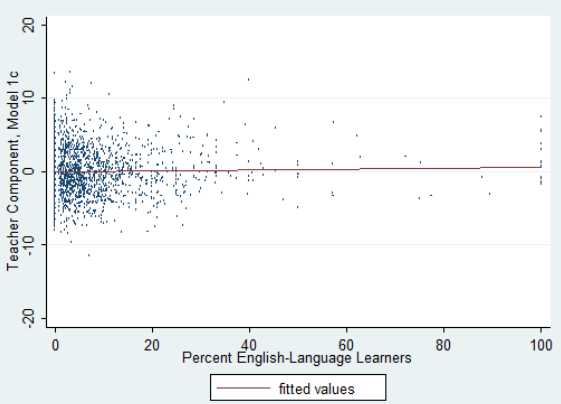


Algebra I

Teacher Component and Percent English Language Learners

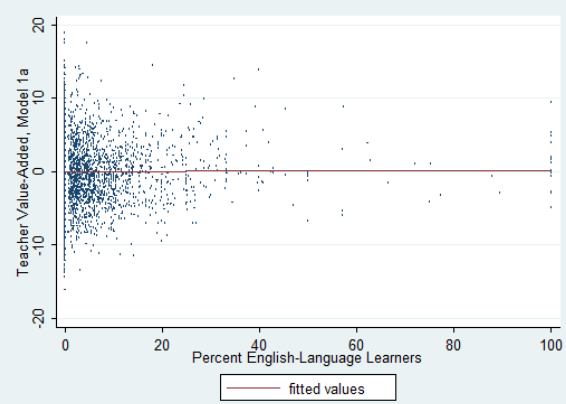


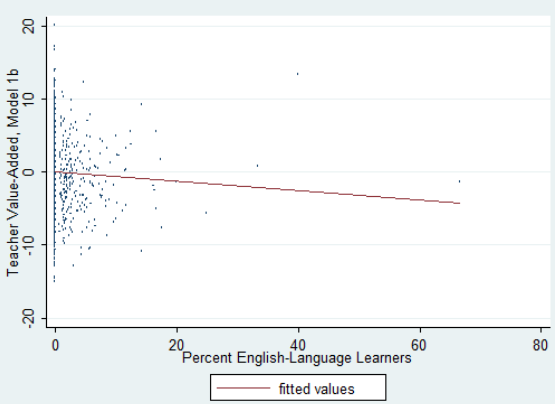


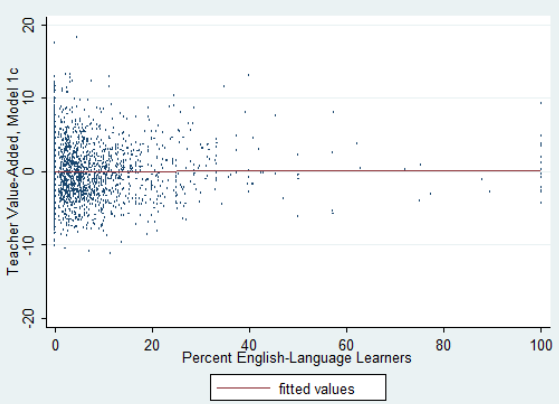


Algebra I

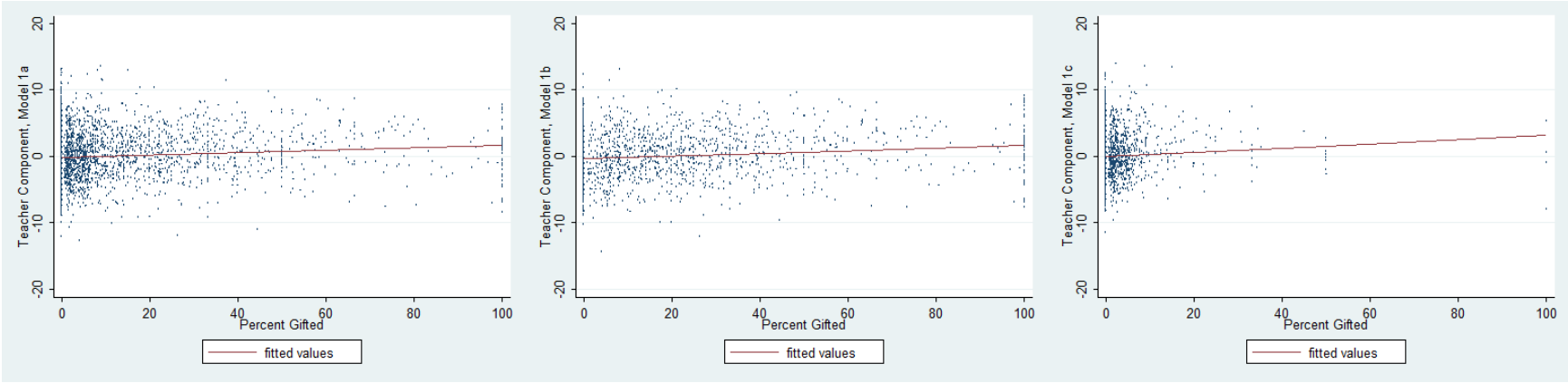
Teacher Value-Added and Percent English Language Learners





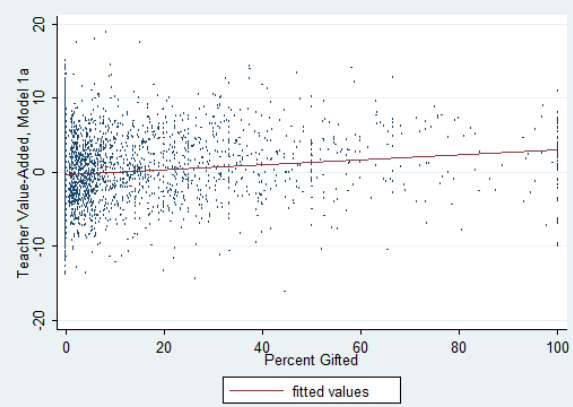


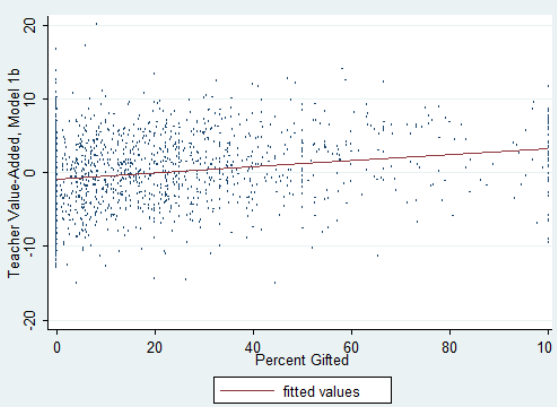
Teacher Component and Percent Gifted

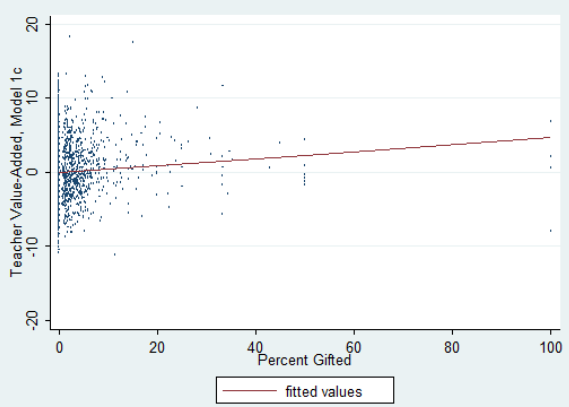


Algebra I

Teacher Value-Added and Percent Gifted

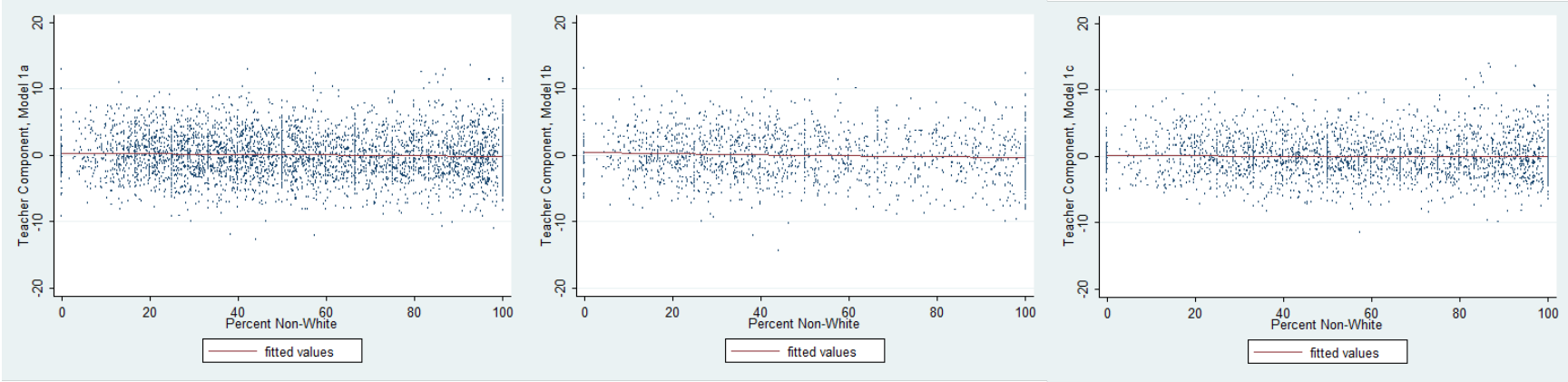






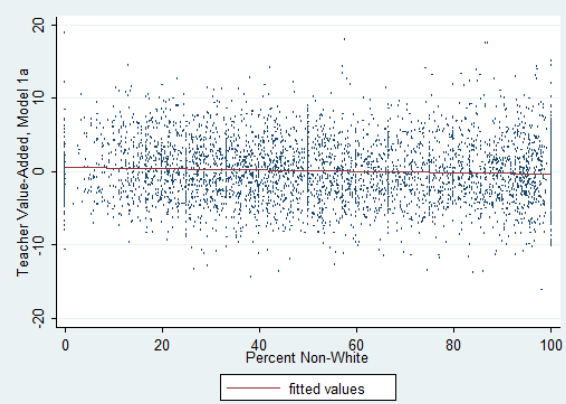
Algebra I

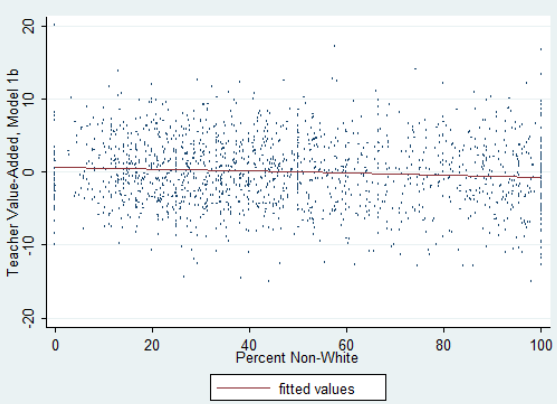
Teacher Component and Percent Non-White

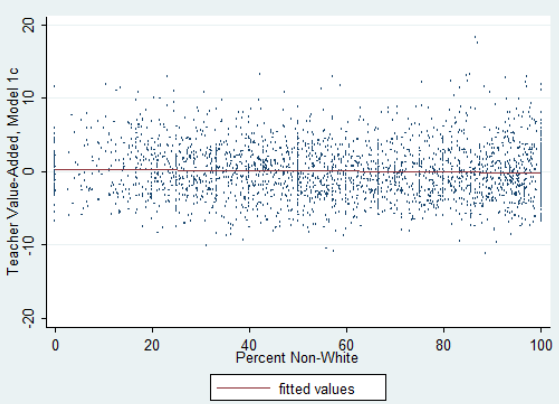


Algebra I

Teacher Value-Added and Percent Non-White







Algebra I

Observed Correlations with Teacher Value-Added Scores

Model	Model 1a		Model 1b		Model 1c	
	No School	School	No School	School	No School	School
Mean Prior	0.08	0.13	0.17	0.29	0.02	0.03
%ED	-0.13	-0.18	-0.19	-0.27	-0.10	-0.14
%SWD	-0.03	-0.04	-0.04	-0.11	-0.04	-0.05
%ELL	0.01	0.00	-0.04	-0.04	0.02	0.00
%Gifted	0.09	0.14	0.14	0.21	0.06	0.08
%Non-White	-0.03	-0.06	-0.06	-0.08	-0.02	-0.03

Impact Data Results

- Note that the relationship between student characteristics and teacher estimates increases when the school component is added.
- The change is much larger in models 1a and 1b than in 1c.
- This is as we'd expect, given the sizes of the teacher and school variances in each model.

Impact Data Results

- Not only are there average differences in level scores between groups of students, but there are also average differences in value-added scores across classrooms and schools with different student demographic characteristics.
- It is not possible to determine the source of the differences across classrooms and schools.

End-of-Course Value-Added Model: Biology

AIR®
American Institutes for Research®

Three Different Model Specifications Were Estimated

- The three models are identical except for the different prior achievement scores included:
  - Model 2a: Science FCAT score
  - Model 2b: Science FCAT score and up to two prior Math FCAT scores
  - Model 2c: Science FCAT score and up to two prior Reading FCAT scores

Biology

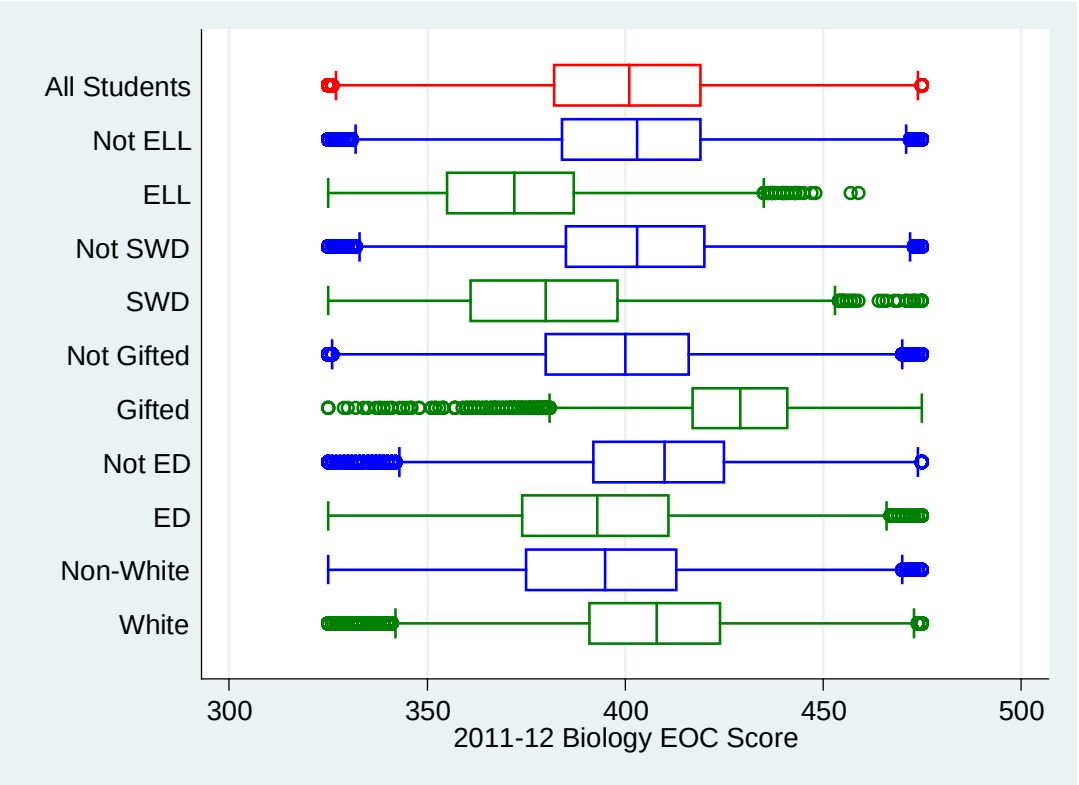
Prior FCAT Score Depends on Student's Grade

Current Grade	Science FCAT	First Math FCAT	Second Math FCAT	First Reading FCAT	Second Reading FCAT
12	8	8	7	10	9
11	8	8	7	10	9
10	8	8	7	9	8
9	8	8	7	8	7
8	5	7	6	7	6

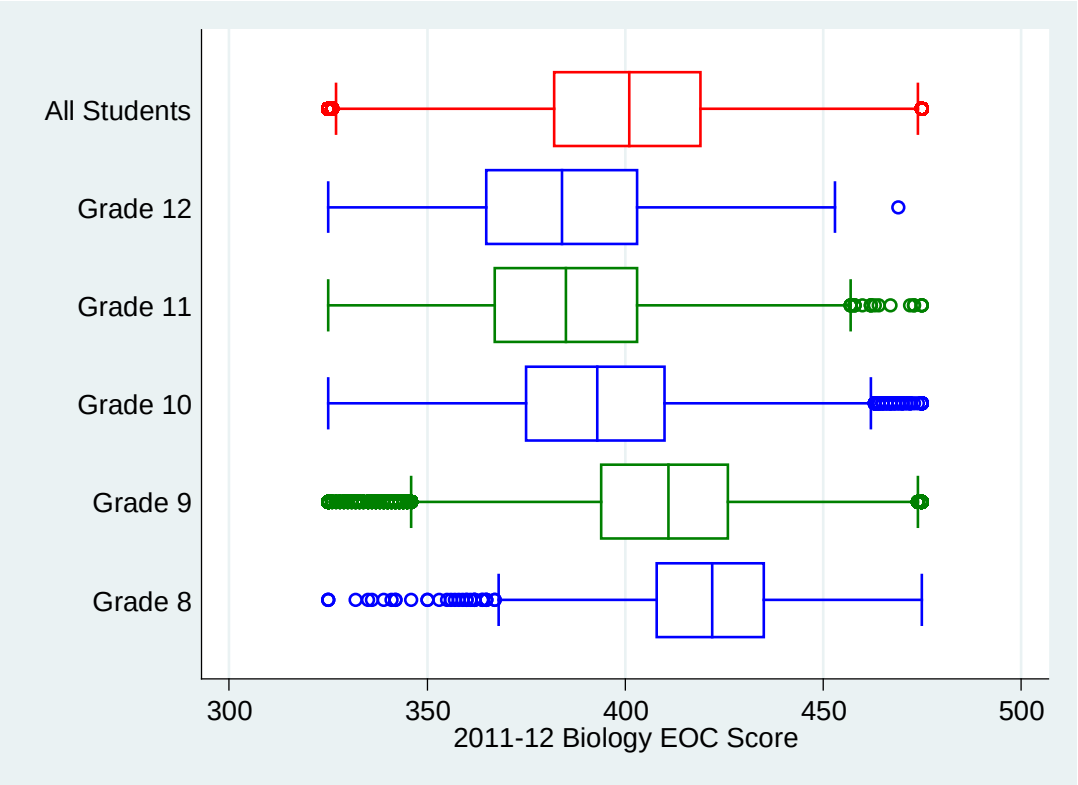
Number of Students per Model

Model	N
Model 2a (Science FCAT)	**147,869**
Model 2b (Science and Math FCATs)	**160,376**
Model 2c (Science and Reading FCATs)	**168,713**

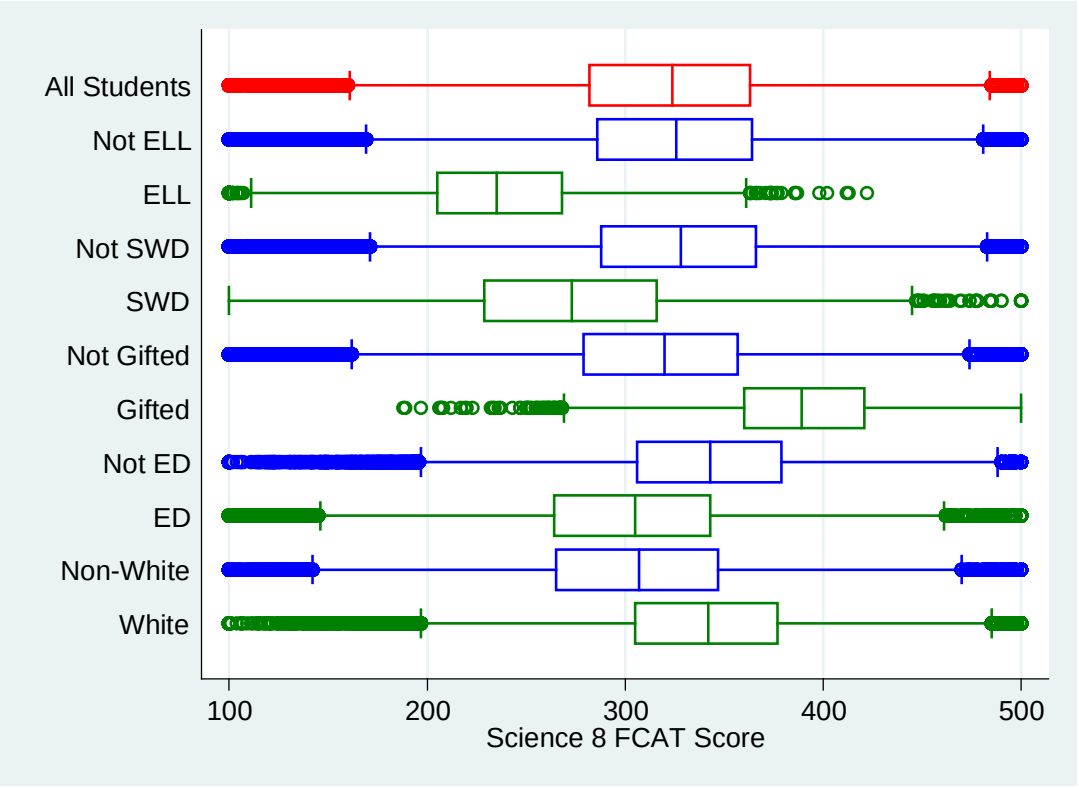
2011–12 Biology EOC Scores: Overall and by Subgroup



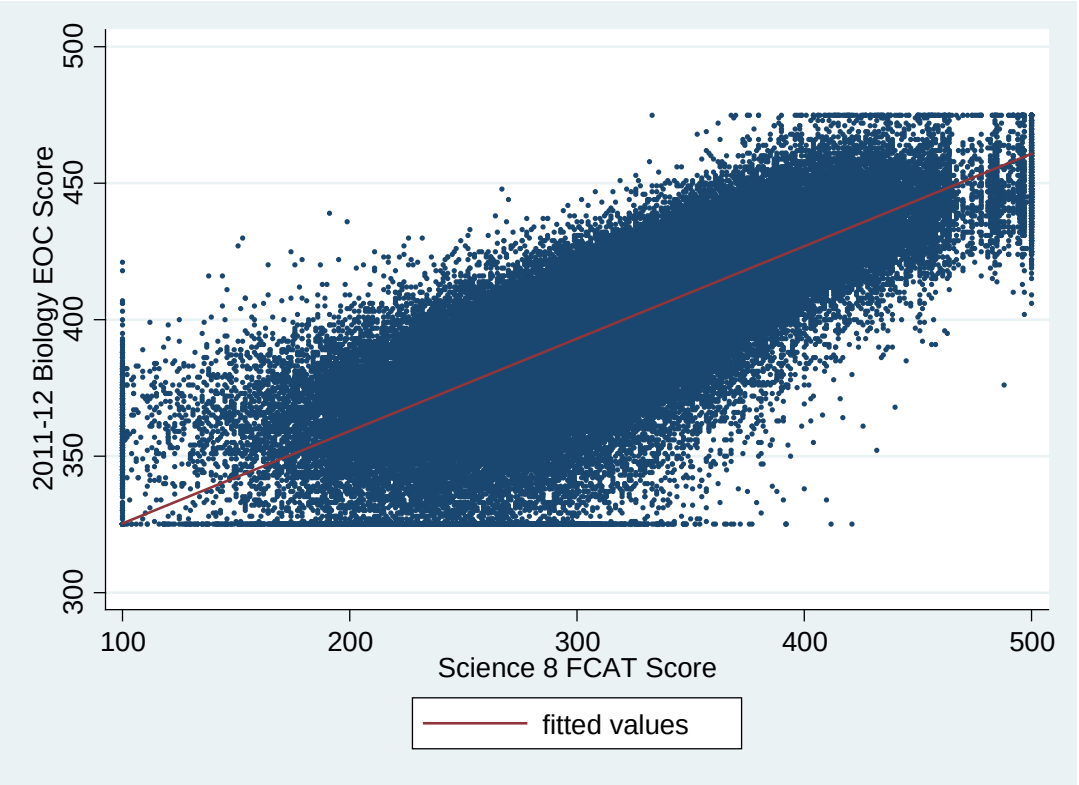
2011–12 Biology EOC Scores: Overall and by Grade



Science 8 FCAT Scores: Overall and by Subgroup



Biology EOC and Science 8 FCAT Scores (Correlation = 0.78)



Summary of Descriptive Statistics

- The data show that students in lower grades score higher on the Biology EOC than students in the higher grades.
- There are large systematic differences between student groups.
- The correlation between the Biology EOC and Science 8 FCAT is within the expected range.

AIR®
American Institutes for Research®

Month 20XX

R-Squared Is Similar Across Models

Model	R-Squared
Model 2a (Science 8)	.62
Model 2b (Science 8 and Math)	.61
Model 2c (Science 8 and Reading)	.63

Biology

Percent of Teachers and Schools Significantly Different from Average

Model	Teachers (above and below)	Schools (above and below)
2a (Science)	12%	10%
2b (Science and Math)	12%	10%
2c (Science and Reading)	12%	9%

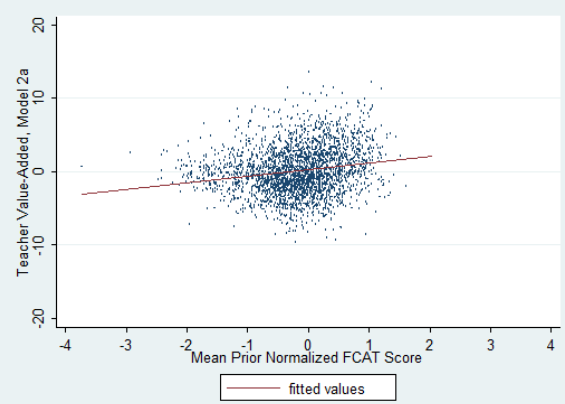
Reliability Ratio Is Not Atypical

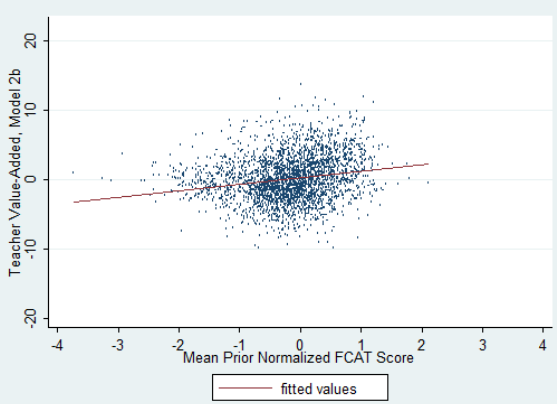
Model	Teachers
2a (Science)	0.96
2b (Science and Math)	0.98
2c (Science and Reading)	0.97

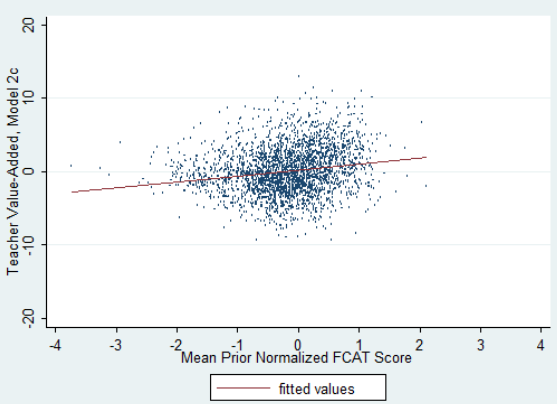
Teacher Component and Mean Normalized Prior Score

Biology

Teacher Value-Added and Mean Normalized Prior Score

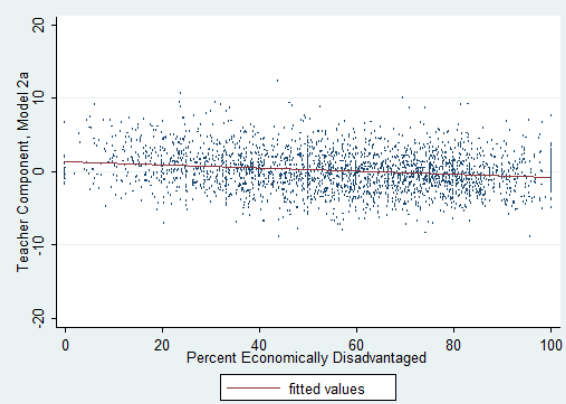


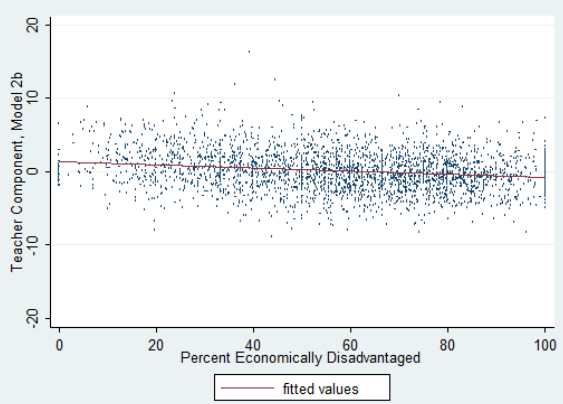


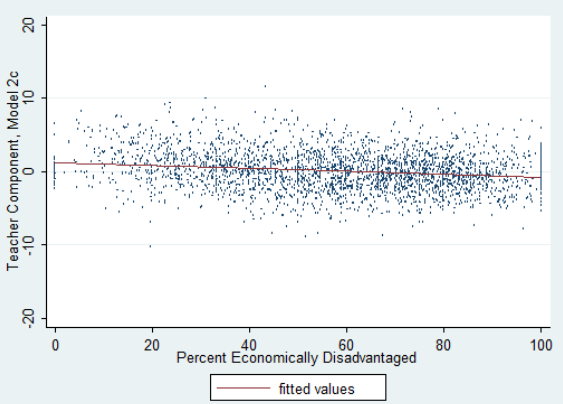


Biology

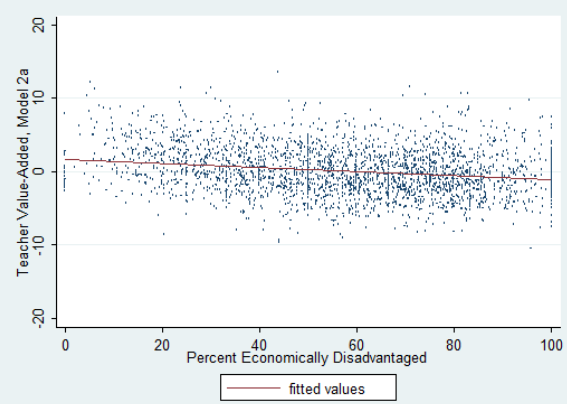
Teacher Component and Percent Economically Disadvantaged

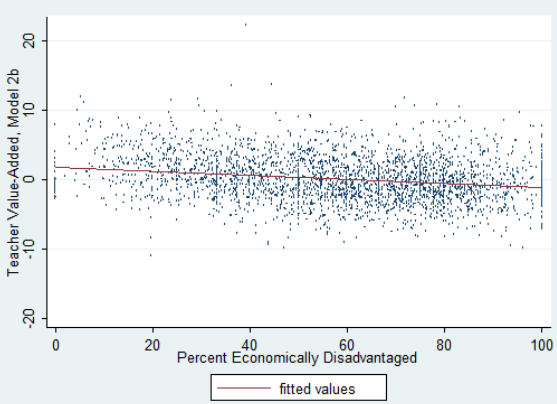


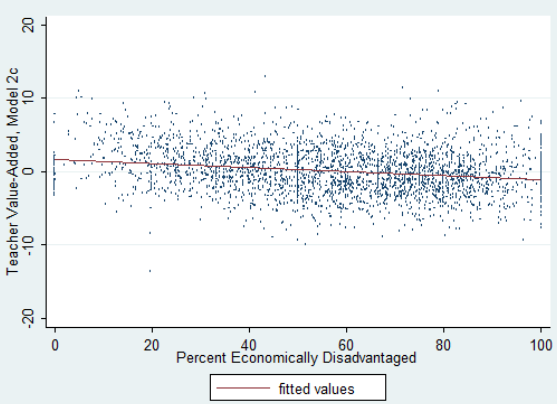




Teacher Value-Added and Percent Economically Disadvantaged

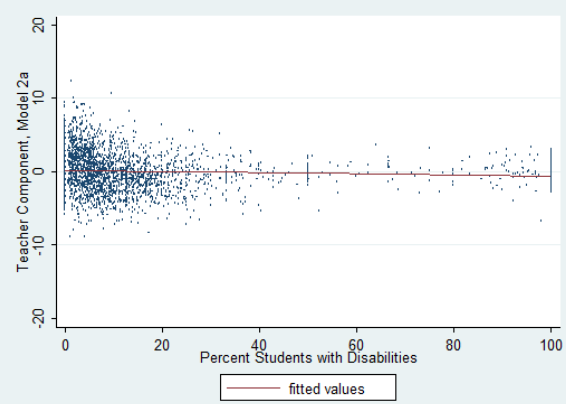


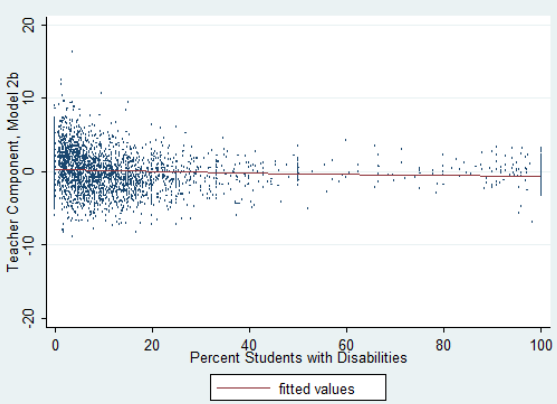


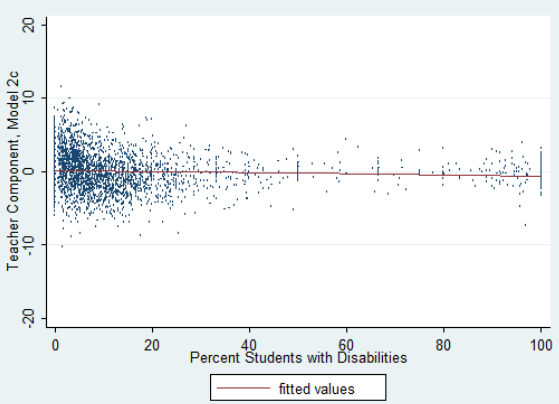


Biology

Teacher Component and Percent Students with Disabilities

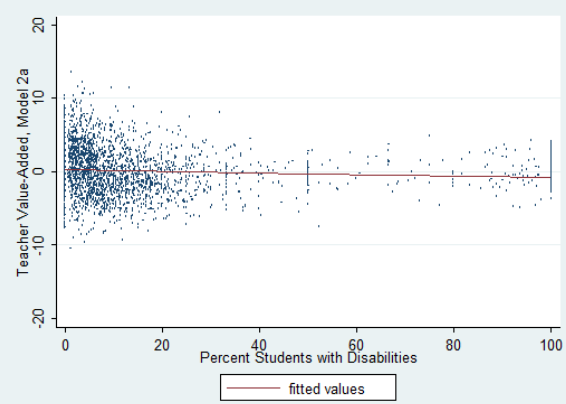


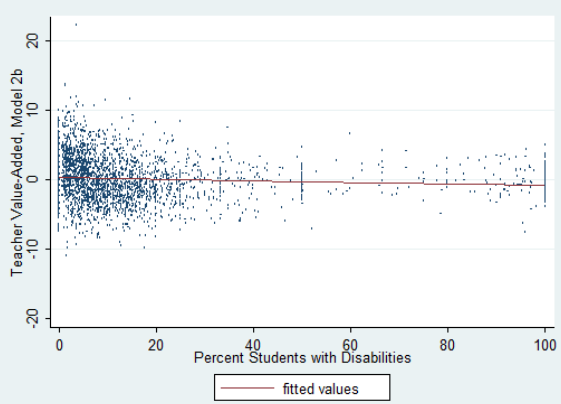


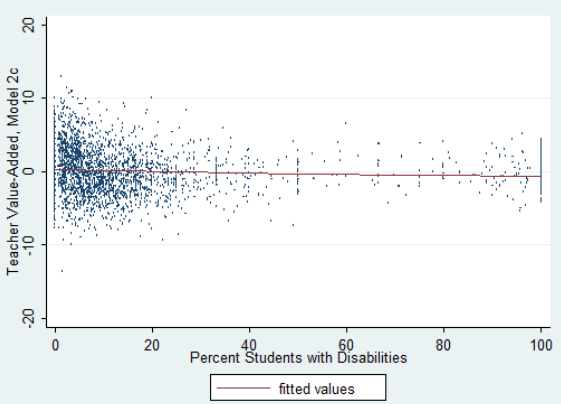


Biology

Teacher Value-Added and Percent Students with Disabilities

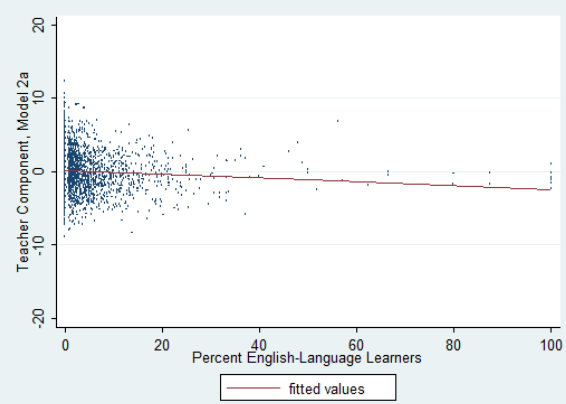


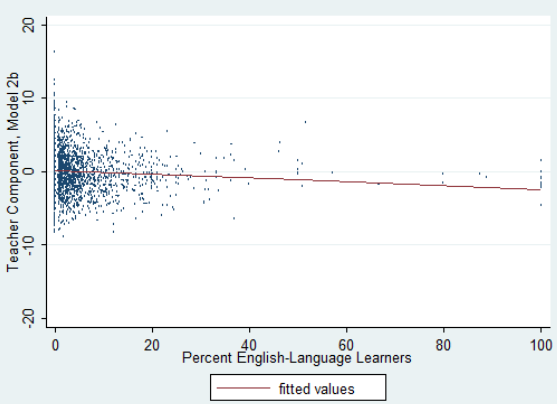


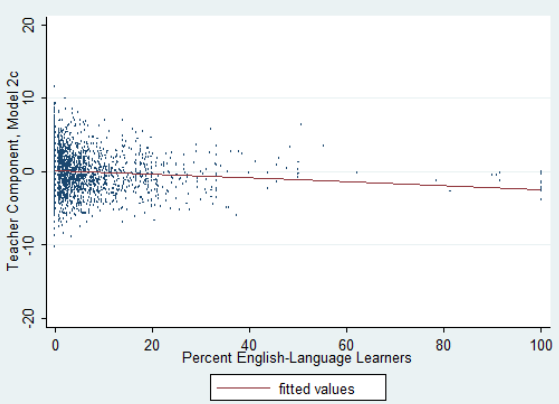


Biology

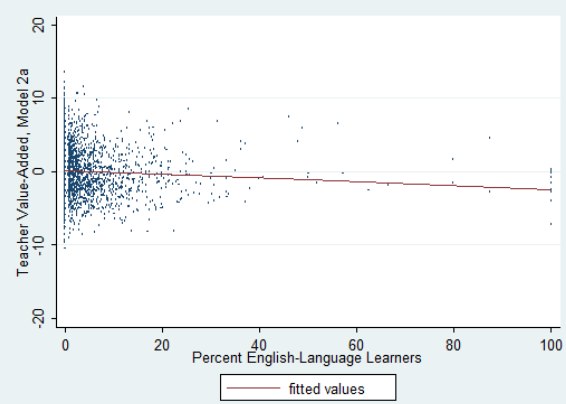
Teacher Component and Percent English Language Learners

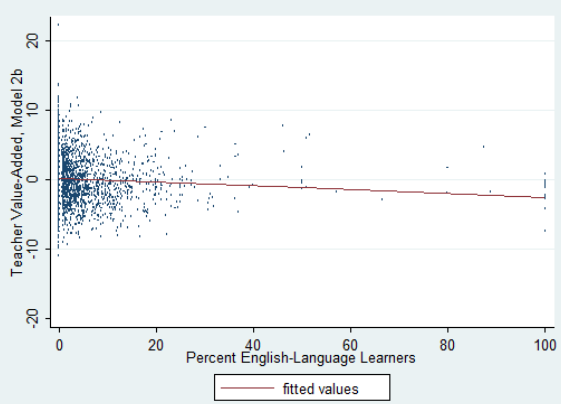


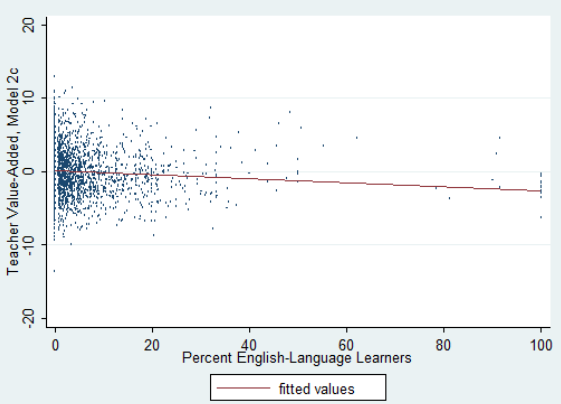




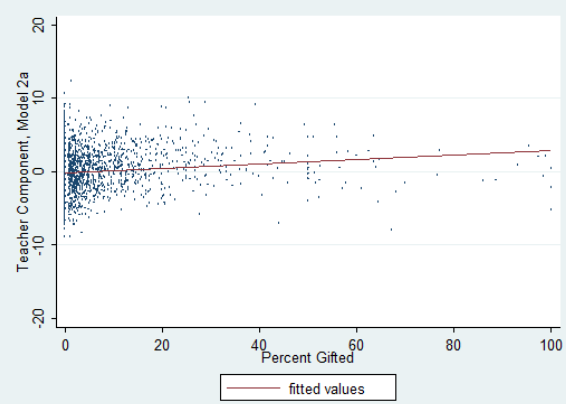
Teacher Value-Added and Percent English Language Learners

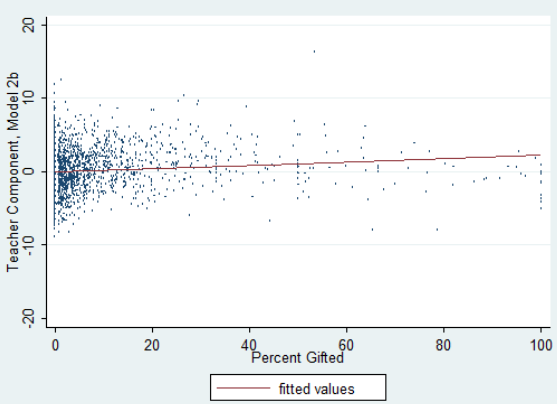


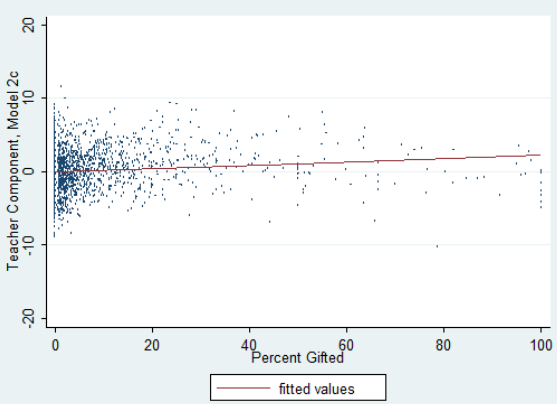




Teacher Component and Percent Gifted

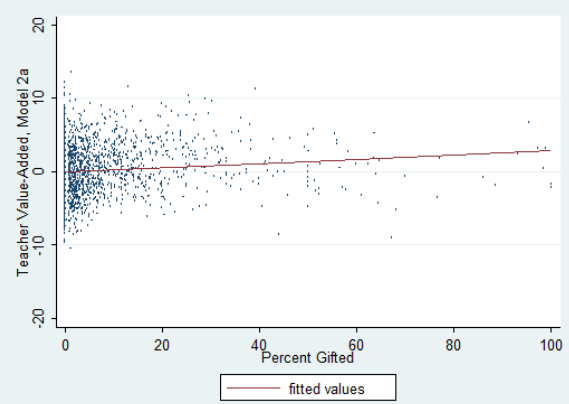


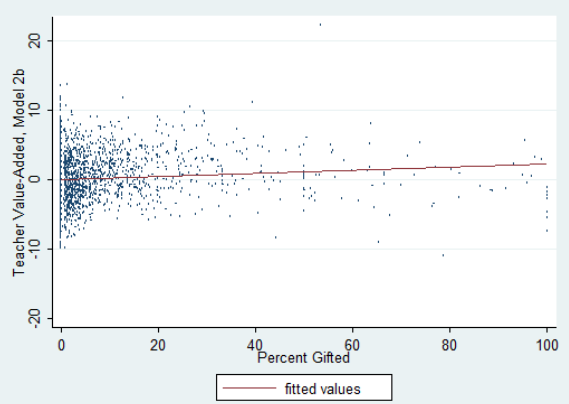


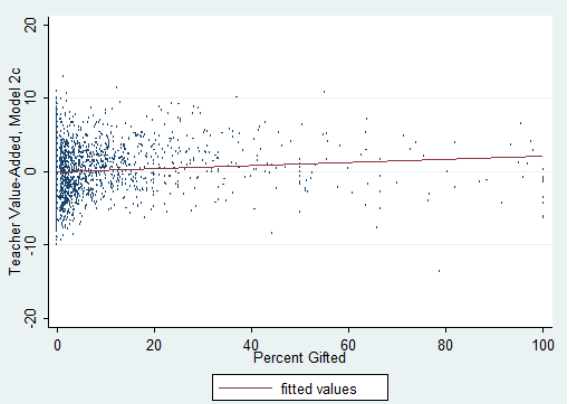


Biology

Teacher Value-Added and Percent Gifted

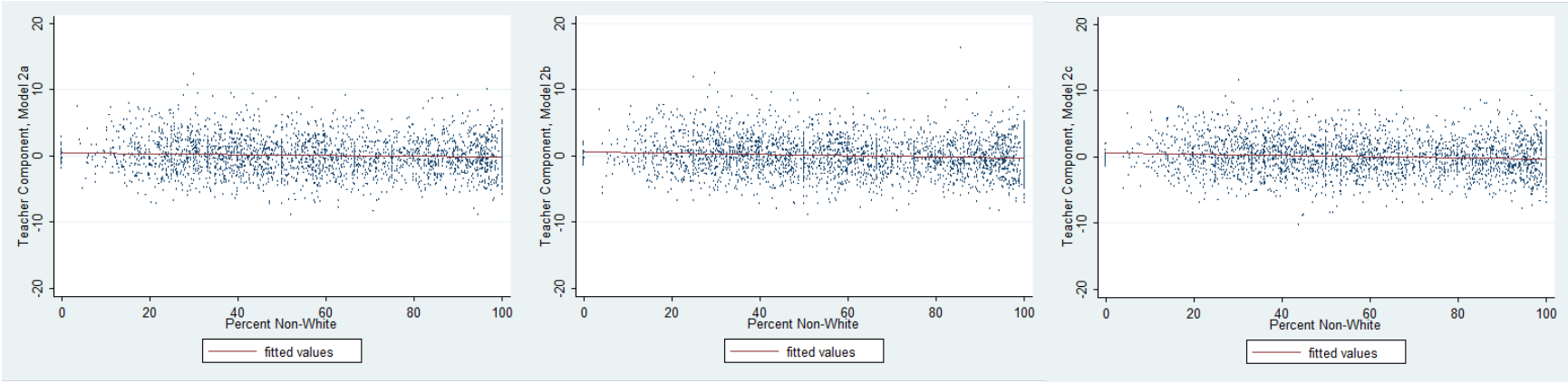




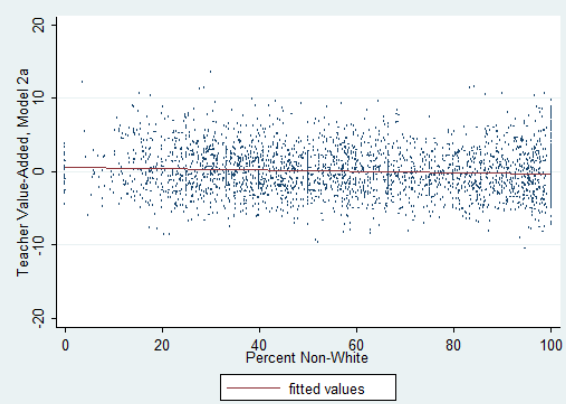


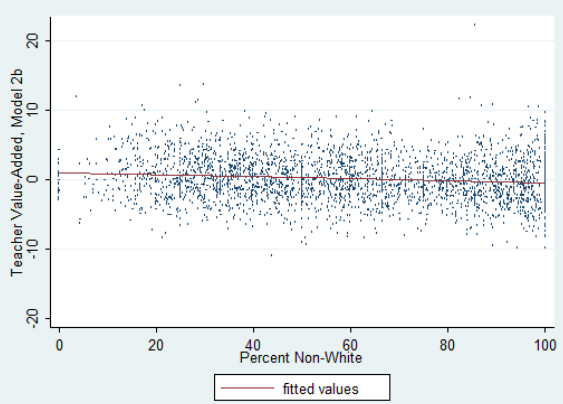
Biology

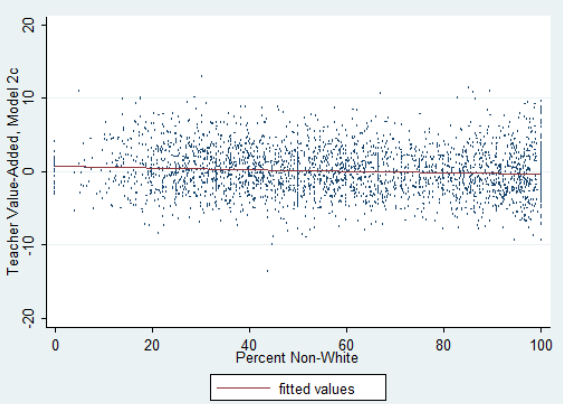
Teacher Component and Percent Non-White



Teacher Value-Added and Percent Non-White







Biology

Observed Correlations with Teacher Value-Added Scores

Model	Model 2a		Model 2b		Model 2c	
	No School	School	No School	School	No School	School
Mean Prior	0.21	0.19	0.21	0.20	0.21	0.18
%ED	-0.19	-0.21	-0.19	-0.22	-0.19	-0.21
%SWD	-0.08	-0.08	-0.08	-0.09	-0.08	-0.08
%ELL	-0.09	-0.08	-0.09	-0.08	-0.11	-0.09
%Gifted	0.12	0.10	0.13	0.10	0.13	0.10
%Non-White	-0.07	-0.07	-0.09	-0.11	-0.08	-0.09

Impact Data Results

- Unlike the Algebra EOC models, the relationship between student characteristics and teacher estimates increases when the school component is added.
- This is as we might expect, given that the variation in teacher quality is greater across teachers than across schools.

End-of-Course Value-Added Model: Geometry

AIR®
American Institutes for Research®

Three Different Geometry EOC Model Specifications Were Estimated

- The three models are identical except for the different prior achievement scores that were included:
  - Model 2a: Algebra I EOC scores
  - Model 2b: Up to two prior Math FCAT scores
  - Model 2c: Algebra I EOC scores and up to two prior Math FCAT scores

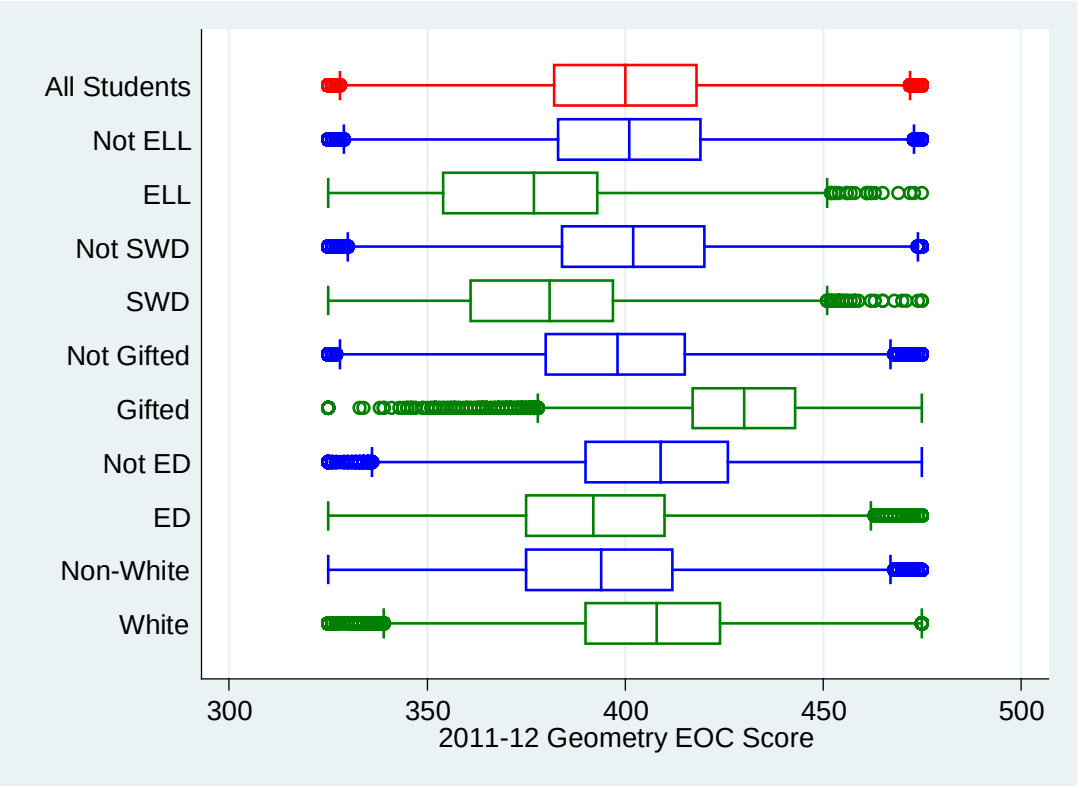
Prior Scores Included Depend on the Student's Current Grade

Current Grade	Algebra I EOC	First Prior Math FCAT	Second Prior Math FCAT
12	Algebra I	8	7
11	Algebra I	8	7
10	Algebra I	8	7
9	Algebra I	8	7
8	Algebra I	7	6

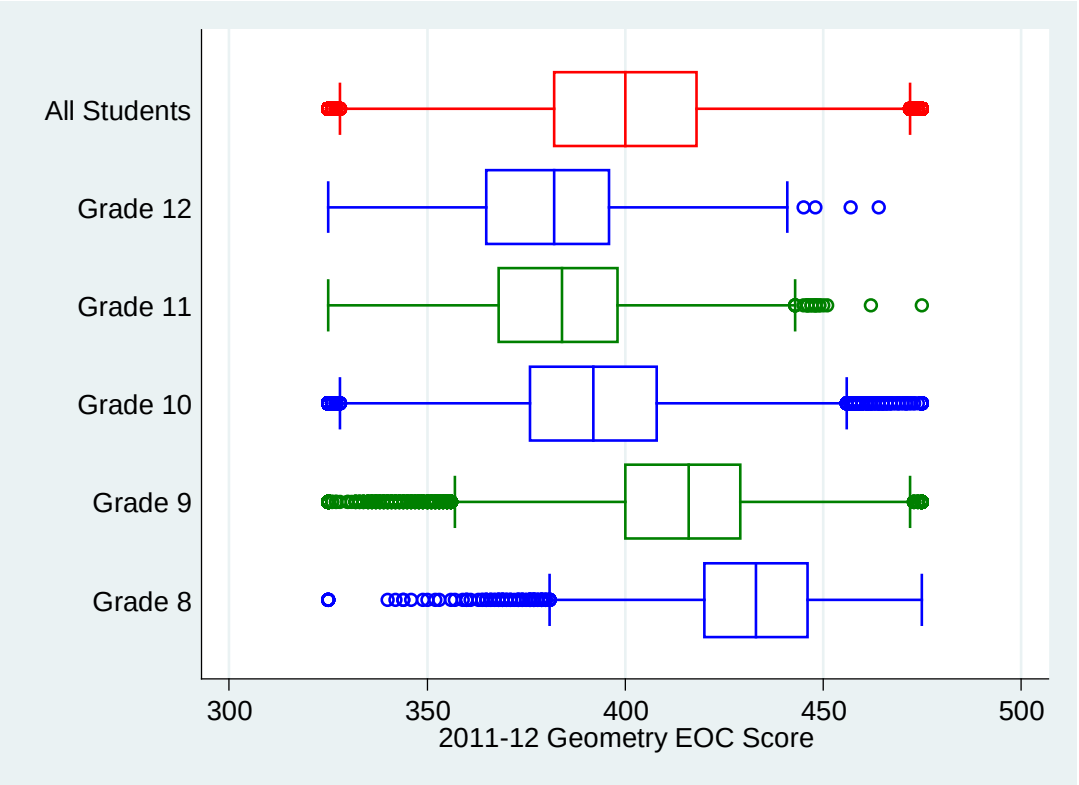
Number of Students per Model

Model	N
Model 2a (Algebra EOC)	**142,956**
Model 2b (Math FCAT)	**155,859**
Model 2c (Algebra EOC and Math FCAT)	**165,843**

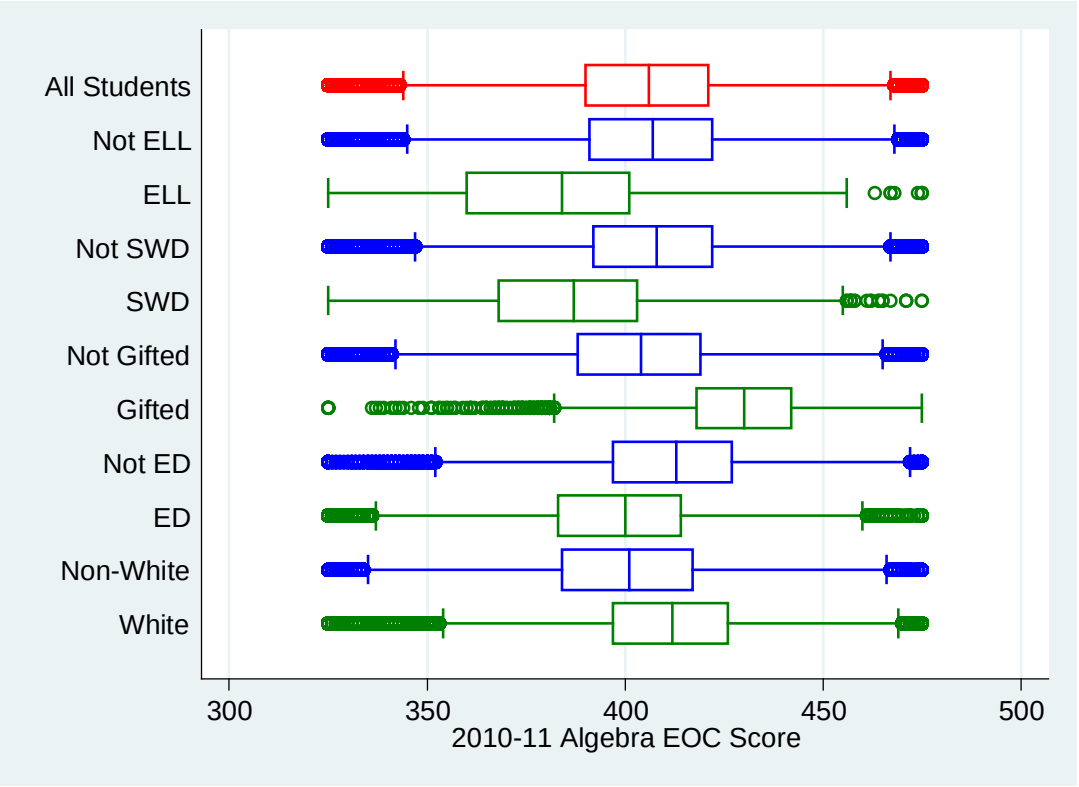
2011–12 Geometry EOC Scores, Overall and by Subgroup



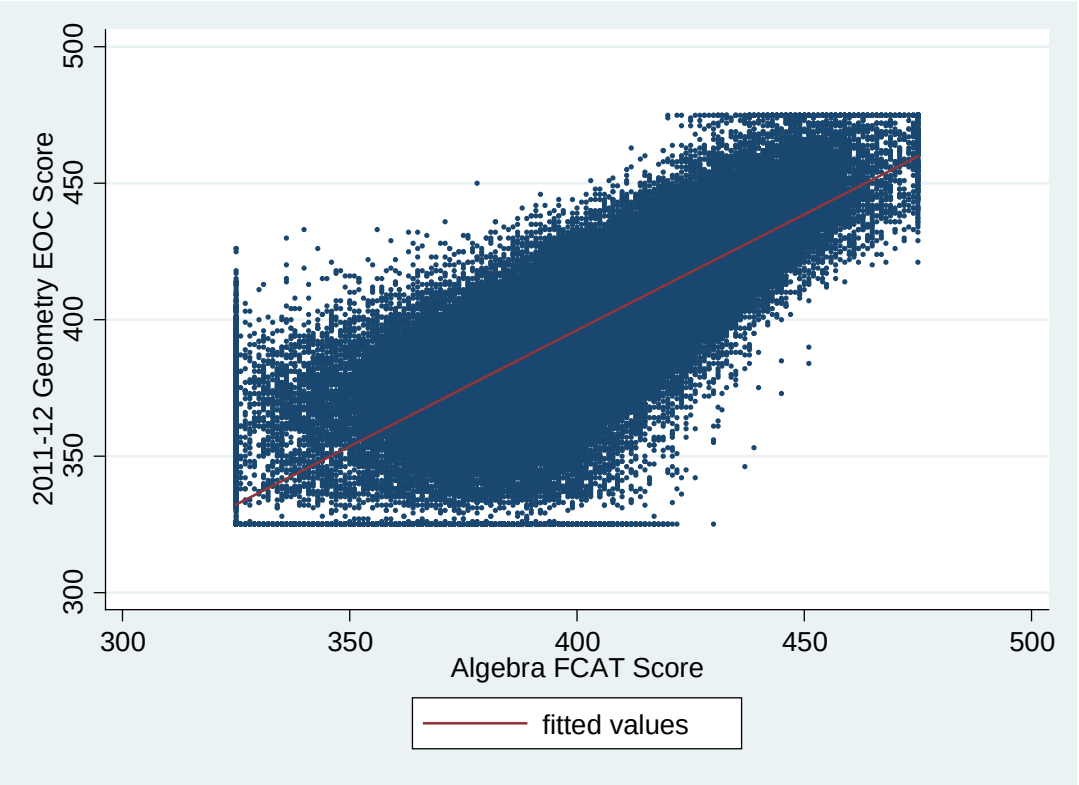
2011–12 Geometry EOC Scores, Overall and by Grade



Prior Algebra EOC Scores, Overall and by Subgroup



Geometry EOC and Algebra EOC Scores (Correlation = 0.76)



Summary of Descriptive Statistics

- The data show that students in lower grades score higher on the Geometry EOC than students in the higher grades.
- There are large systematic differences between student groups.
- Correlation between Geometry EOC and Algebra EOC scores is within the expected range.

AIR®
American Institutes for Research®

Month 20XX

R-Squared Is Similar Across Models

Model	R-Squared
Model 2a (Algebra EOC)	.62
Model 2b (Math FCAT)	.62
Model 2c (Algebra & Math FCAT)	.65

Geometry

Percent of Teachers and Schools Significantly Different from Average

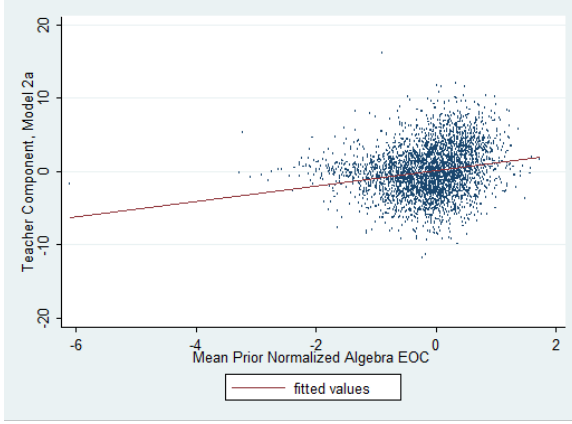
Model	Teachers (above and below)	Schools (above and below)
2a (Algebra EOC)	18%	6%
2b (Math FCAT)	17%	11%
2c (Algebra EOC and Math FCAT)	18%	9%

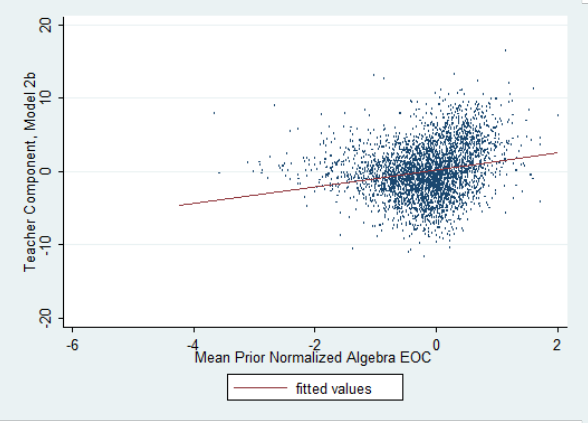
Reliability Ratio

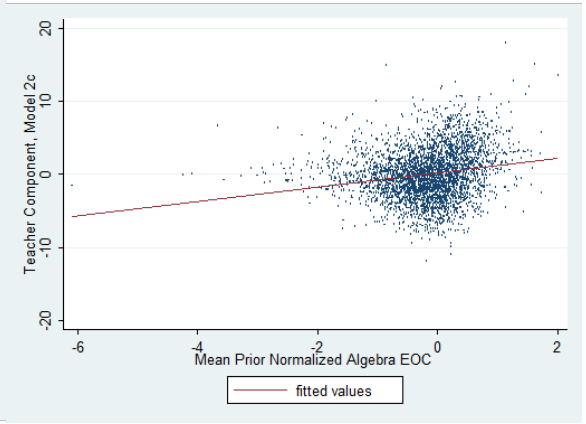
Model	Teachers
2a (Algebra EOC)	0.81
2b (Math FCAT)	0.84
2c (Algebra EOC and Math FCAT)	0.82

Geometry

Teacher Component and Mean Normalized Prior Score

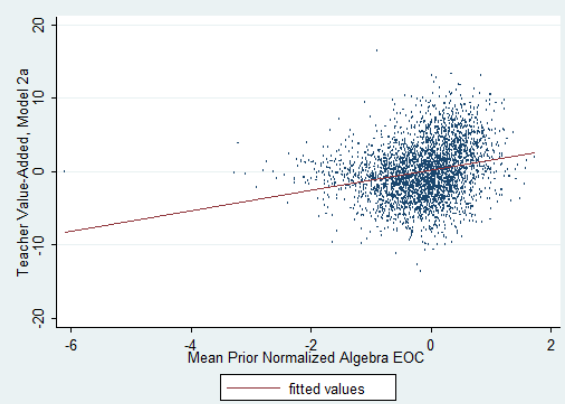


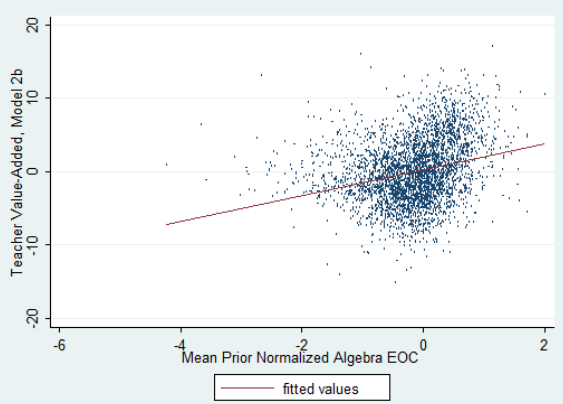


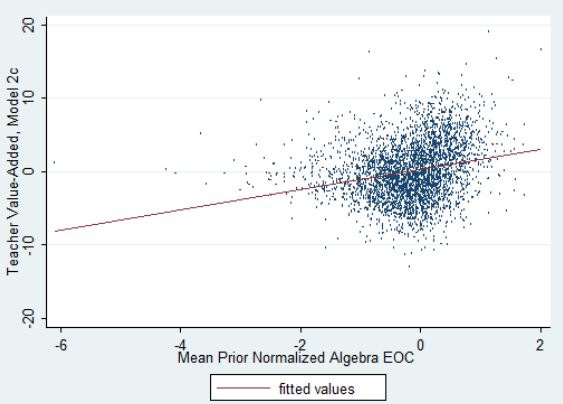


Geometry

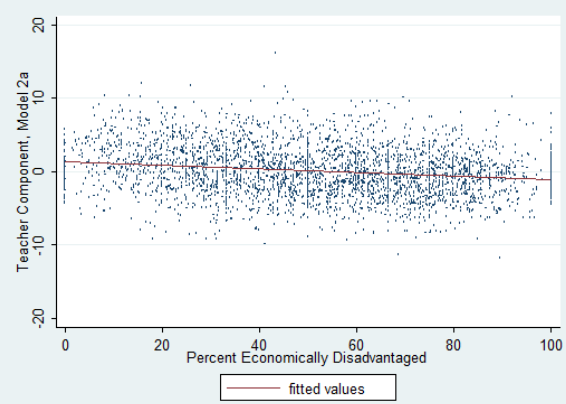
Teacher Value-Added and Mean Normalized Prior Score

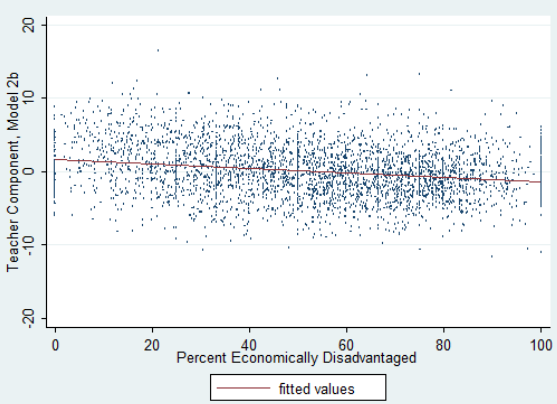


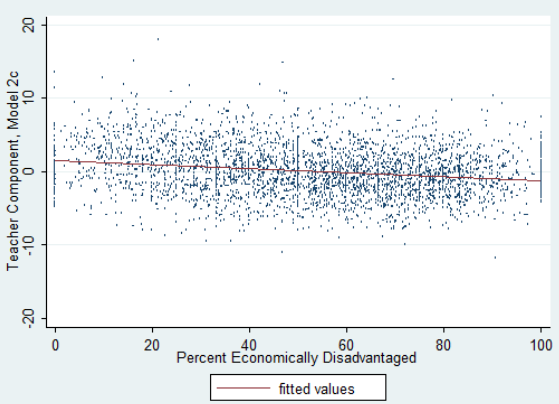




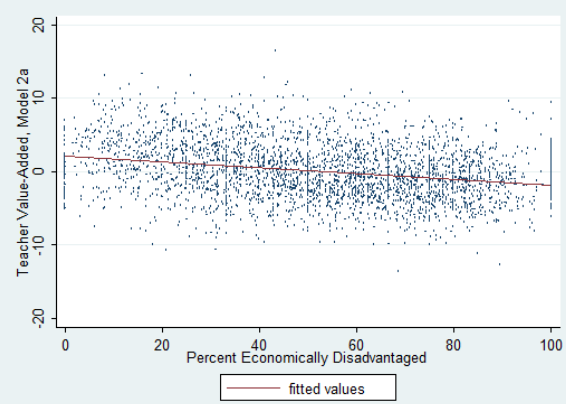
Teacher Component and Percent Economically Disadvantaged

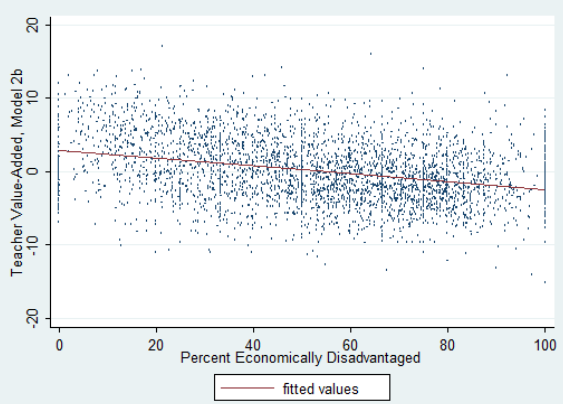


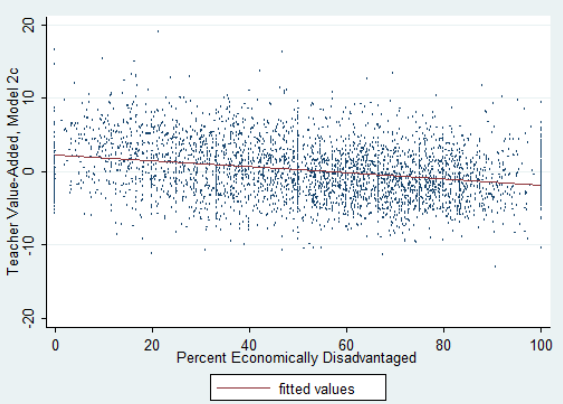




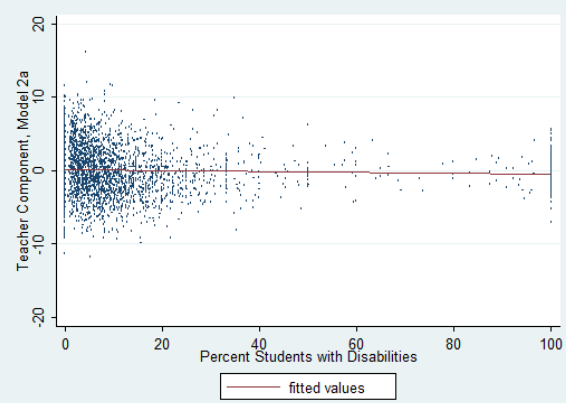
Teacher Value-Added and Percent Economically Disadvantaged

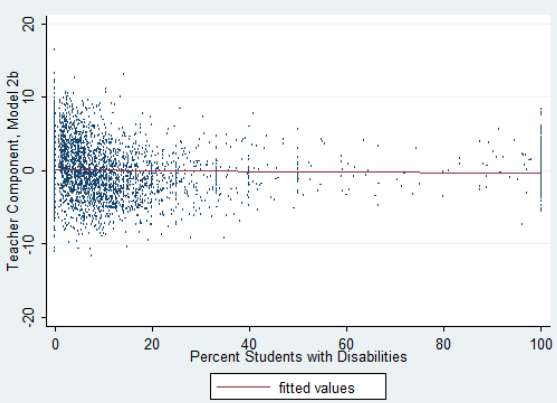


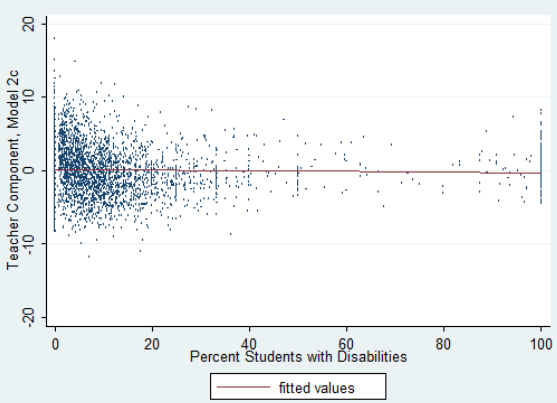




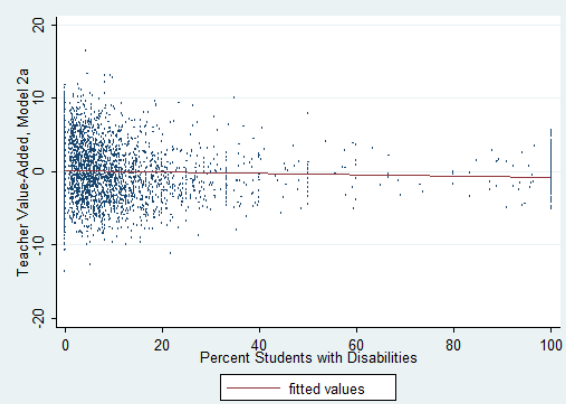
Teacher Component and Percent Students with Disabilities

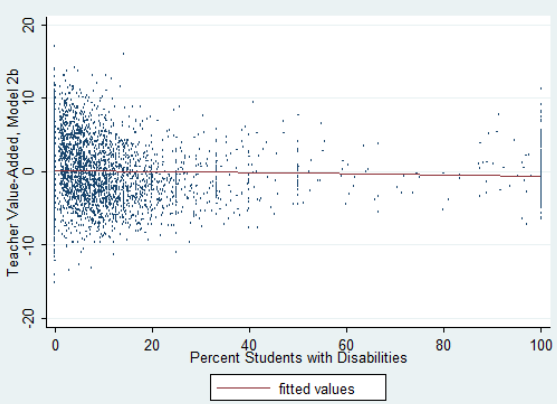


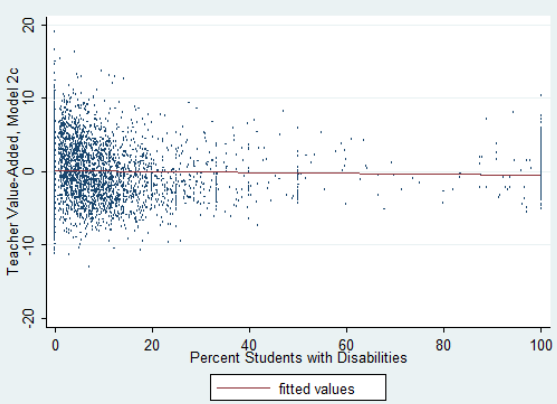




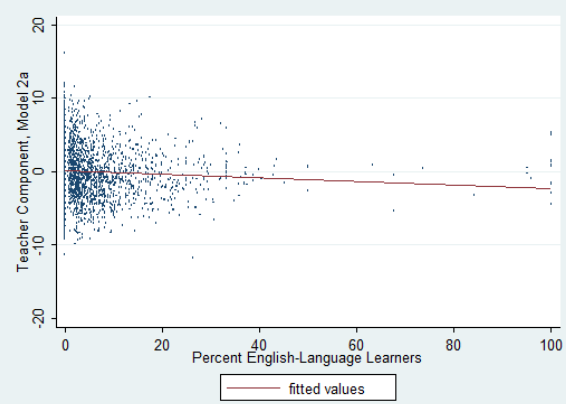
Teacher Value-Added and Percent Students with Disabilities

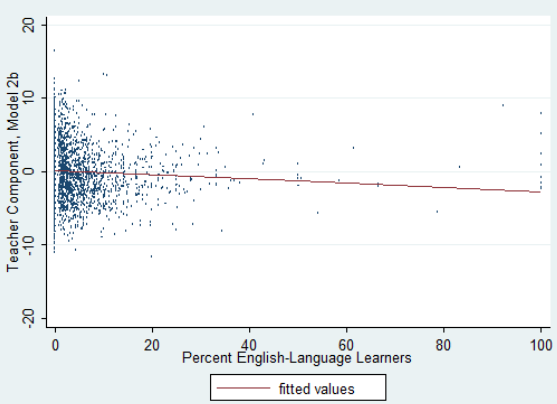


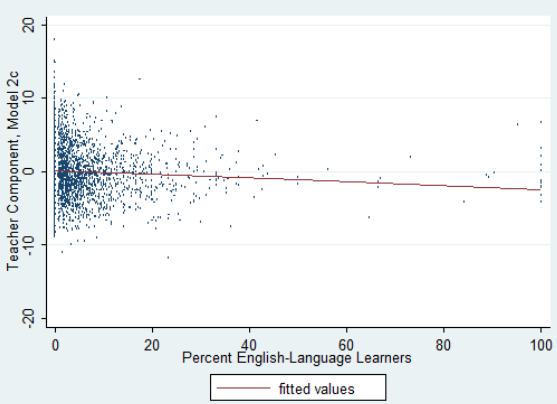




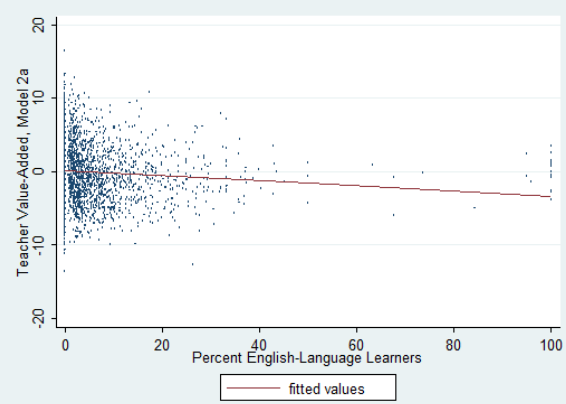
Teacher Component and Percent English Language Learners

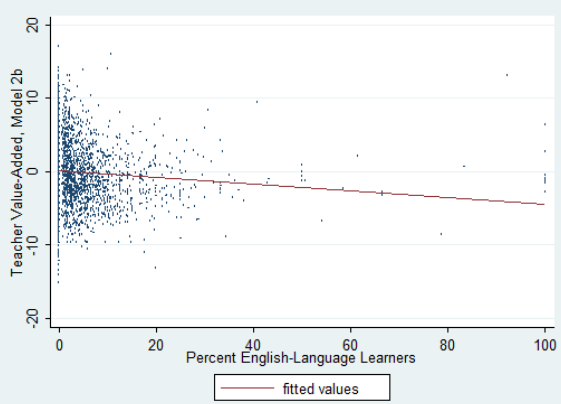


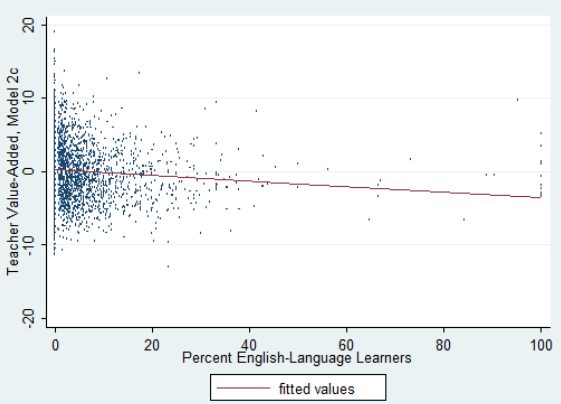




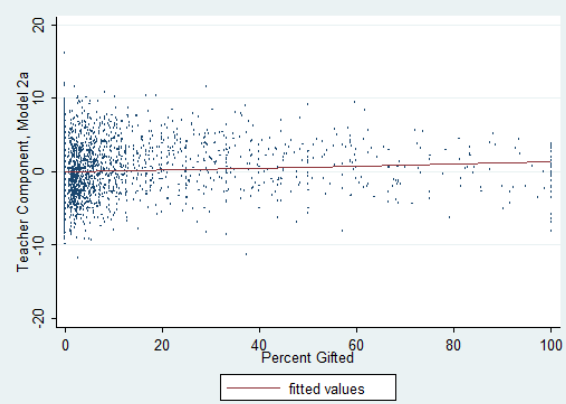
Teacher Value-Added and Percent English Language Learners

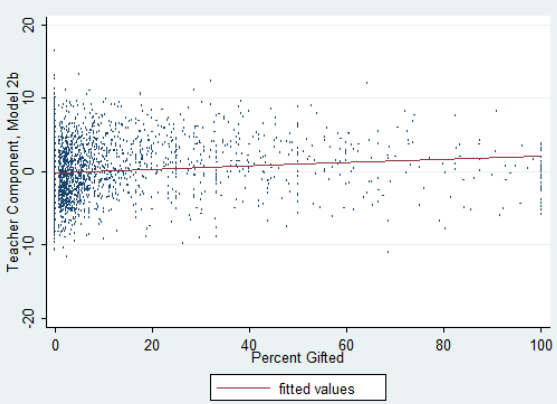


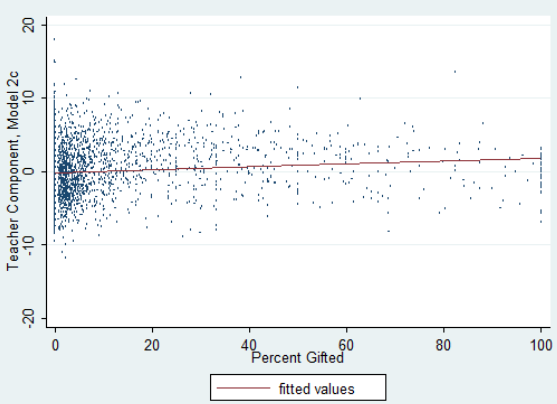




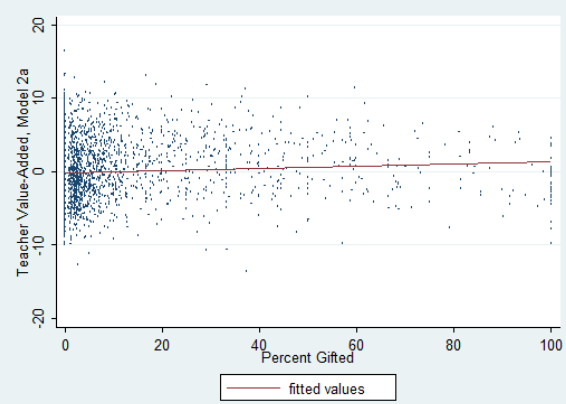
Teacher Component and Percent Gifted

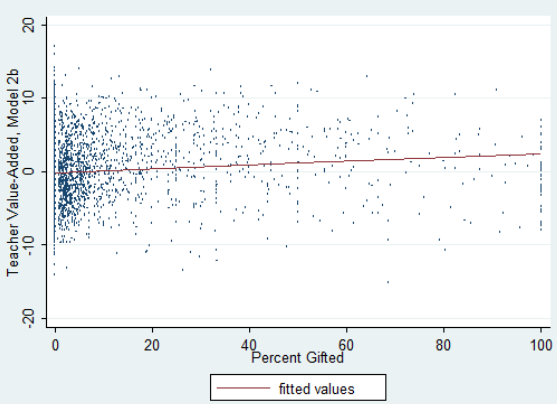


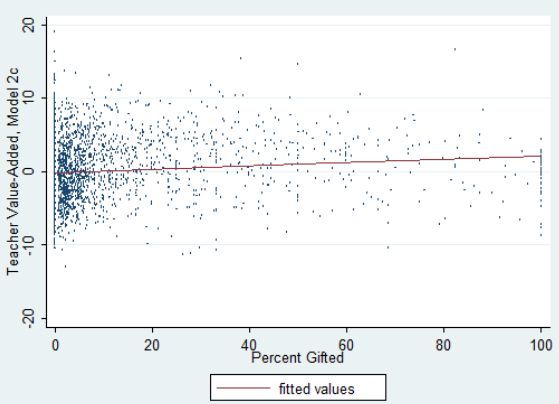




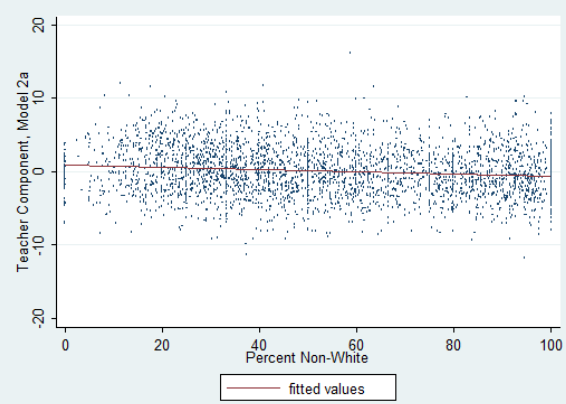
Teacher Value-Added and Percent Gifted

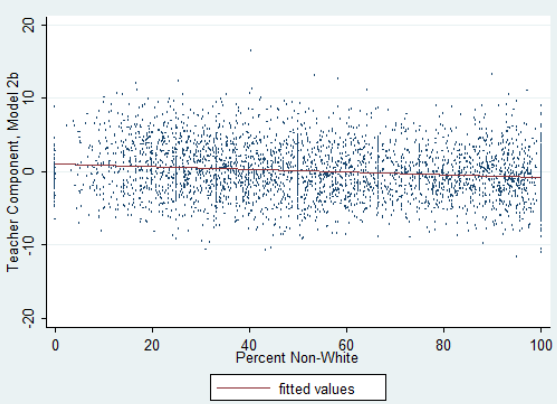


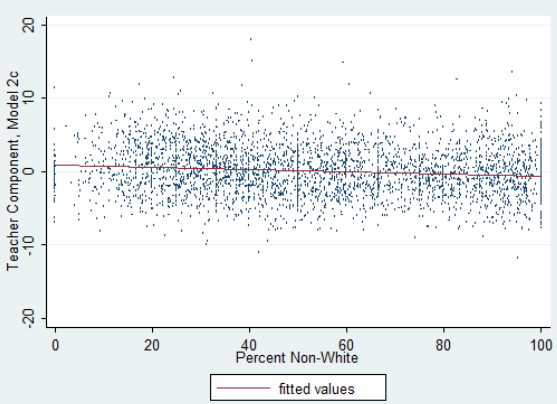




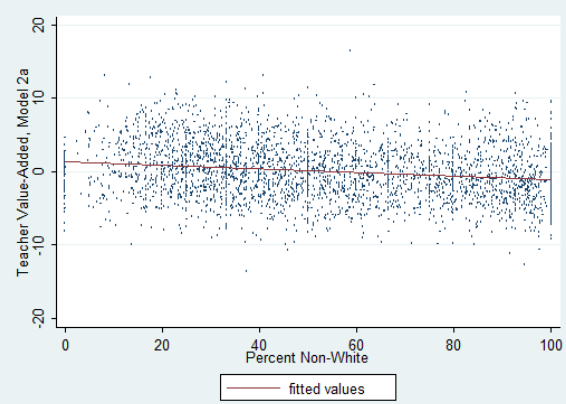
Teacher Component and Percent Non-White

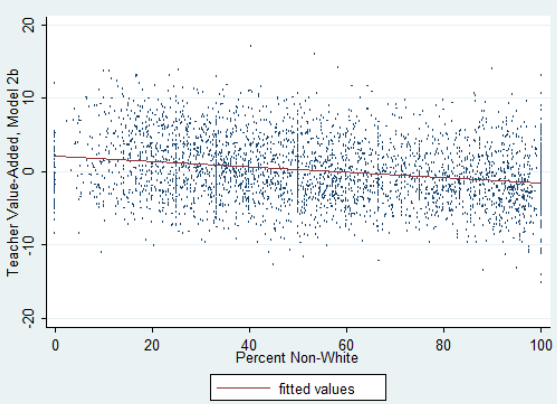


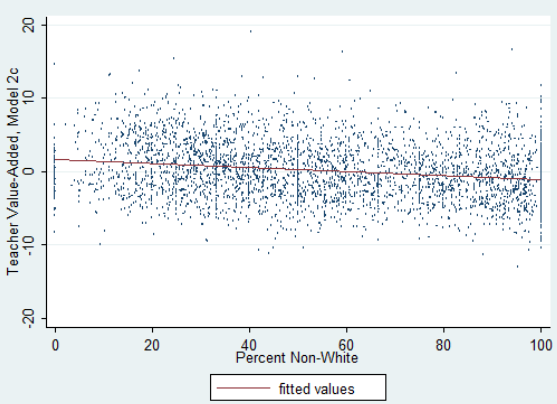




Teacher Value-Added and Percent Non-White







Geometry

Observed Correlations with Teacher Value-Added Scores

Model	Model 2a		Model 2b		Model 2c	
	No School	School	No School	School	No School	School
Mean Prior	0.20	0.23	0.21	0.26	0.19	0.23
%ED	-0.20	-0.26	-0.22	-0.31	-0.20	-0.27
%SWD	-0.05	-0.06	-0.03	-0.05	-0.03	-0.04
%ELL	-0.07	-0.09	-0.07	-0.09	-0.07	-0.09
%Gifted	0.07	0.07	0.11	0.11	0.10	0.10
%Non-White	-0.13	-0.19	-0.14	-0.24	-0.12	-0.20

Impact Data Results

- The impact of the mean prior score, the percent ED, and the percent non-white is larger than the impact of other characteristics.
- Adding the school component increases the impact of percent ED and percent non-white more than it affects the impact of other school characteristics.

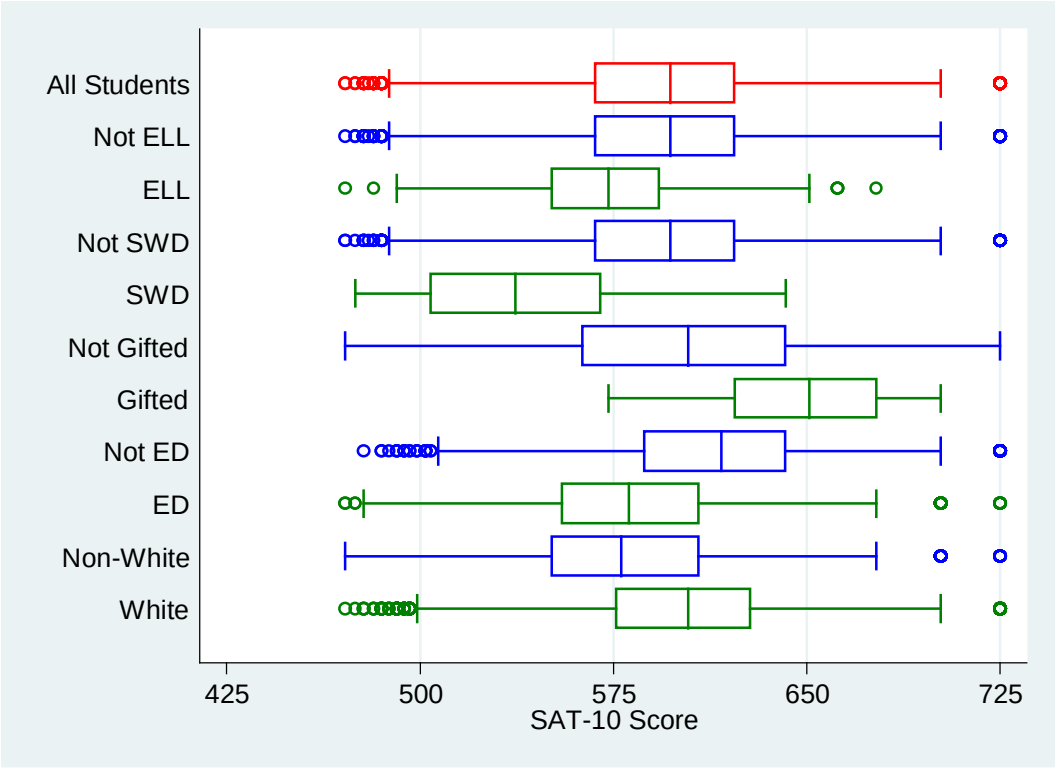
Optional Value-Added Model: SAT-10

AIR®
American Institutes for Research®

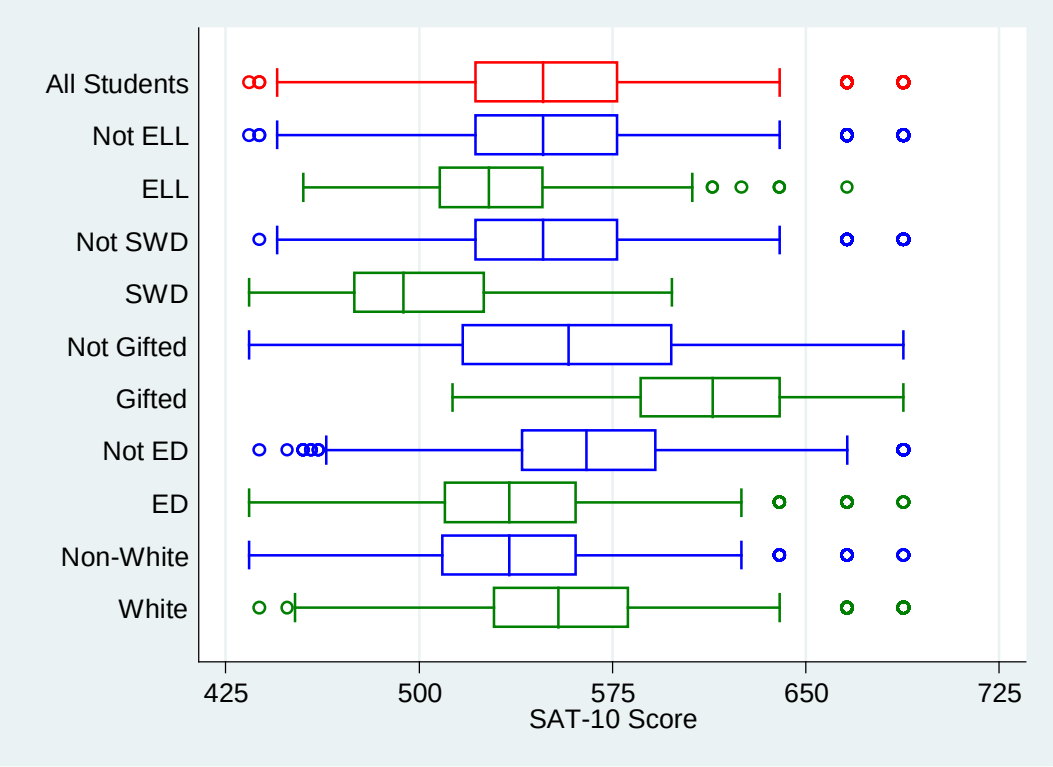
SAT-10 Background Information

- SAT-10 scores are used to create value-added scores for grade 2 teachers.
- Grade 1 scores are used as predictors for the grade 2 outcome variable.
- SEMs were not provided; as a result, measurement error is not accounted for.
  - If SEMs are available, they should be used to account for measurement error.
- The VAM implemented for SAT-10 is the same statistical model used for the FCAT VAMs.

2010–11 SAT-10 Scores: All Students and by Subgroup



Prior Year SAT-10 Scores: All Students and by Subgroup



Summary of Descriptive Statistics

- The differences between groups are typical for in-level score analyses.
- All discrepancies appear normal.
- Correlation between current and prior score (0.77) is typical.

AIR®
American Institutes for Research®

Month 20XX

R-Squared Is One Indicator of Model Fit

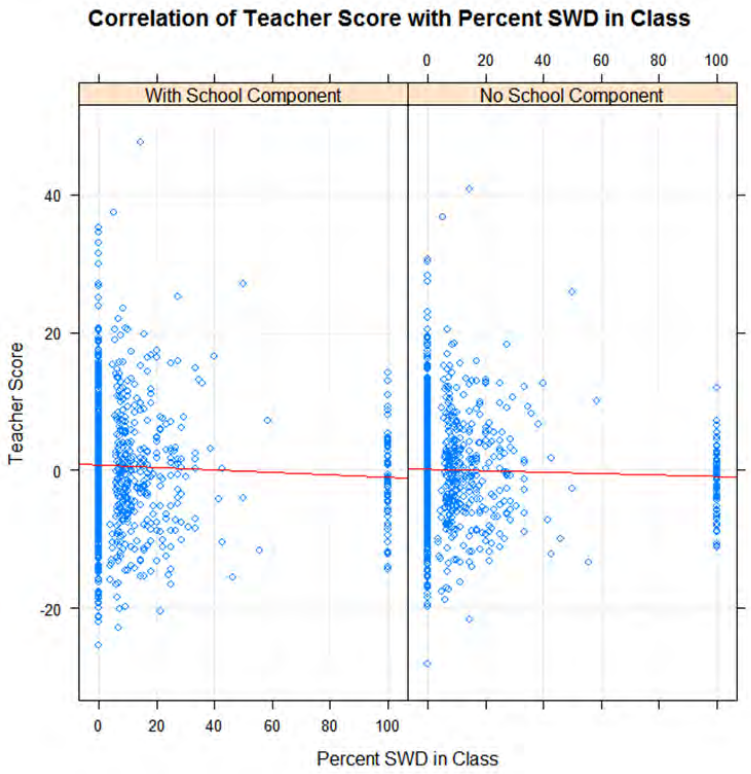
- For the SAT-10, the R-squared is 0.62.
- This is on par with the FCAT R-squared.

Reliability Ratio

- For SAT-10, the teacher reliability ratio is 0.95.
- Percent significantly above or below average:
  - Teachers: 8.9%
  - Schools: 16.8%

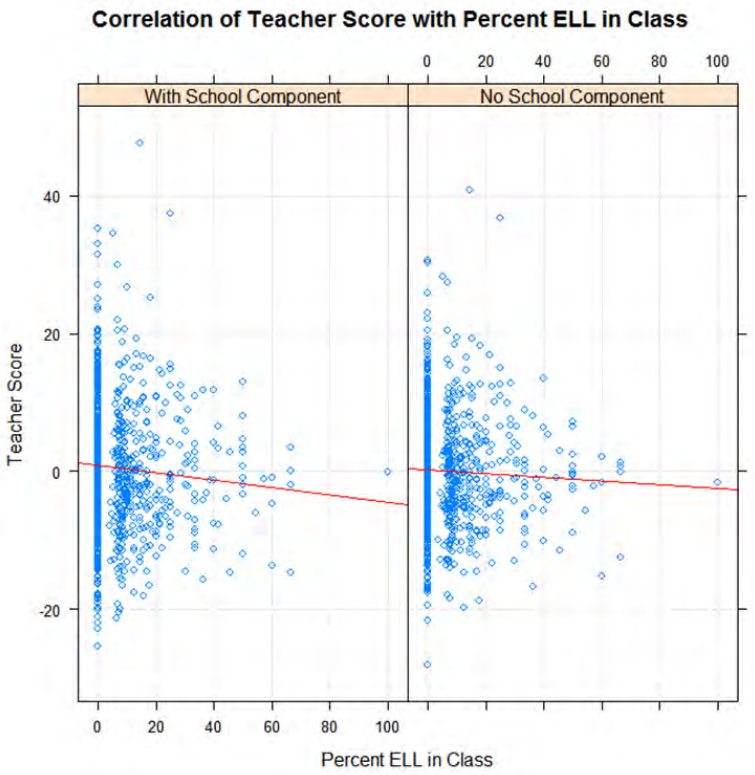
SAT-10

Teacher Value-Added and Percent Students with Disabilities



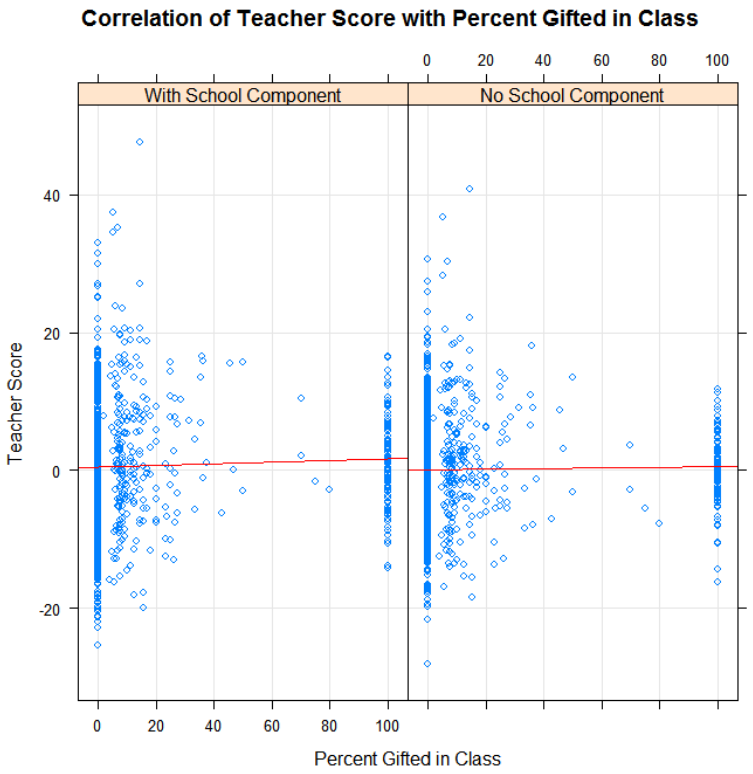
SAT-10

Teacher Value-Added and Percent English Language Learners



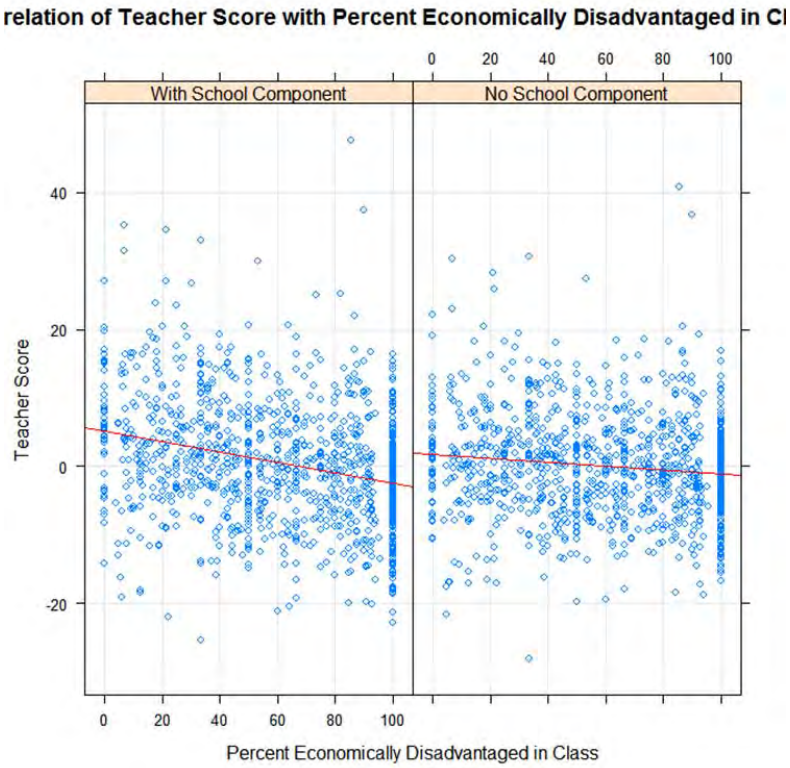
SAT-10

Teacher Value-Added and Percent Gifted



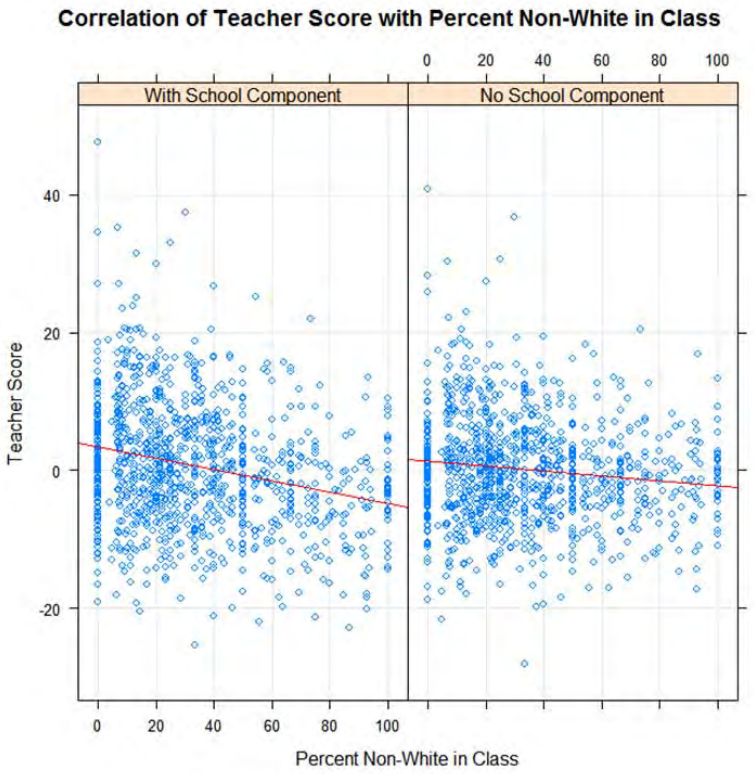
SAT-10

Teacher Value-Added and Percent Economically Disadvantaged



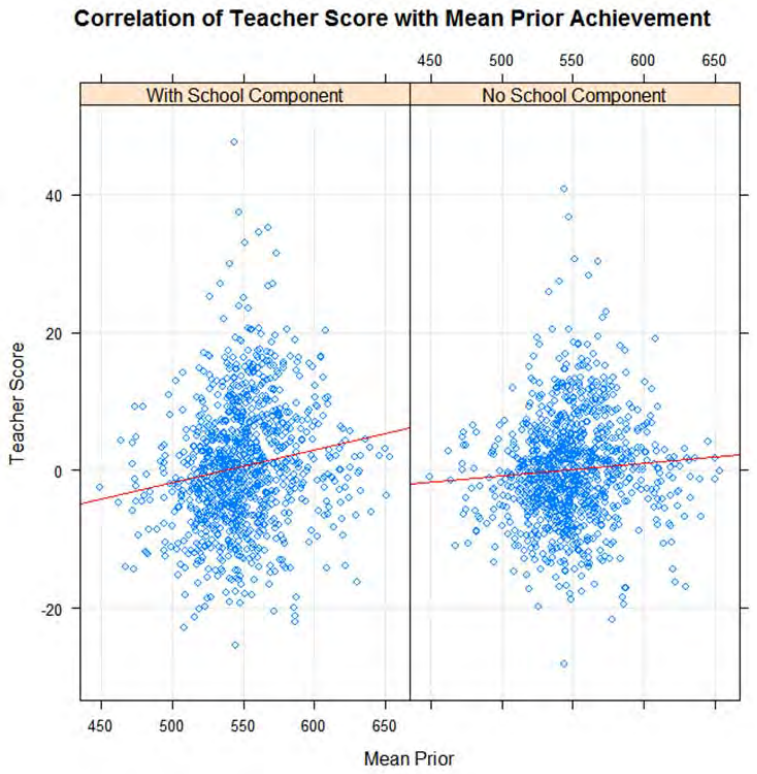
SAT-10

Teacher Value-Added and Percent Non-White



SAT-10

Teacher Value-Added and Mean Prior SAT-10 Score



SAT-10

Observed Correlations with Teacher Value-Added Scores

Model	No School Component	With School Component
Mean Prior	0.07	0.15
%ED	-0.12	-0.27
%SWD	-0.03	-0.05
%ELL	-0.04	-0.07
%Gifted	0.02	0.04
%Non-White	-0.12	-0.24

Impact Data Summary

- The impact data correlations are larger when the teacher score includes some of the school component.
- In this instance, it suggests that the school component adds back some of the systematic differences between schools that a VAM is trying to account for.

Optional Value-Added Models: AP English and AP Calculus

AIR®
American Institutes for Research®

AP

Advanced Placement Background Information

- Unlike the FCAT, SAT-10, and EOC exams, AP scores are categorical and not continuous, ranging from 1 to 5.
- A categorical model known as an ordered probit is used instead of a multilevel linear model.

AP Advanced Placement Background Information

- There is often only one AP teacher per school. This makes it impossible to estimate teacher effects and school effects separately. Therefore, the teacher value-added score includes only a teacher component and does not include a school component.
- Because student grade level is not reported with AP scores, models do not include grade-level covariates.

AP

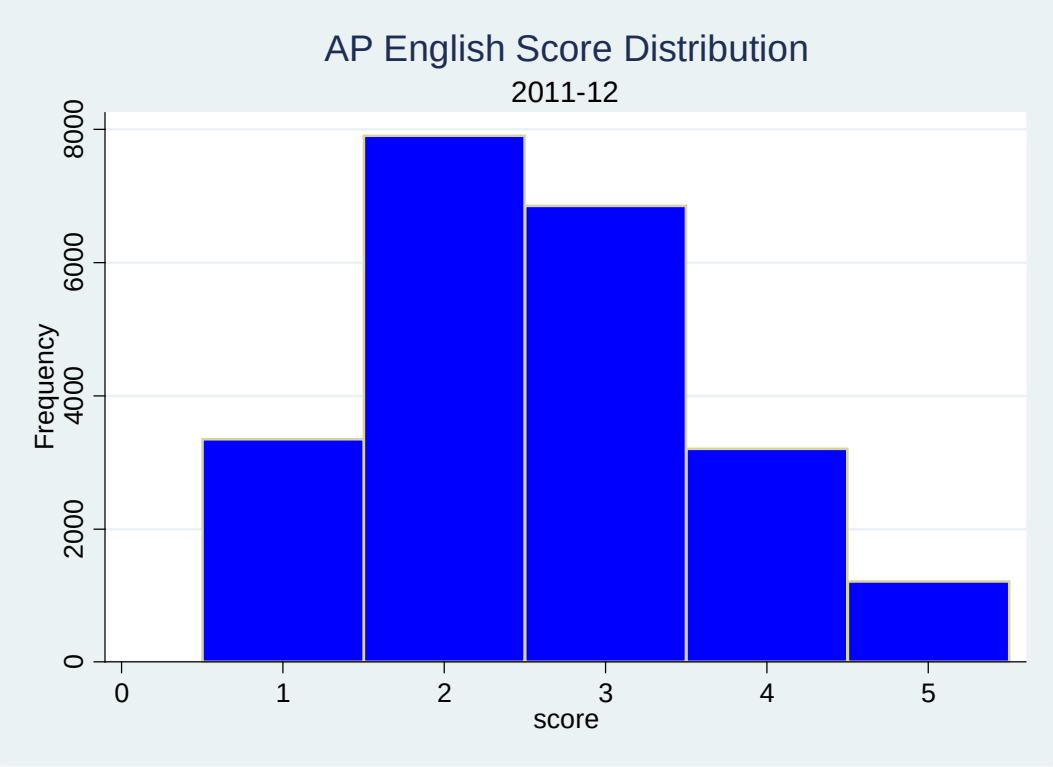
Three Times as Many Students Take AP English as Take AP Calculus AB

Model	N
AP English	**22,518**
AP Calculus AB	**7,330**

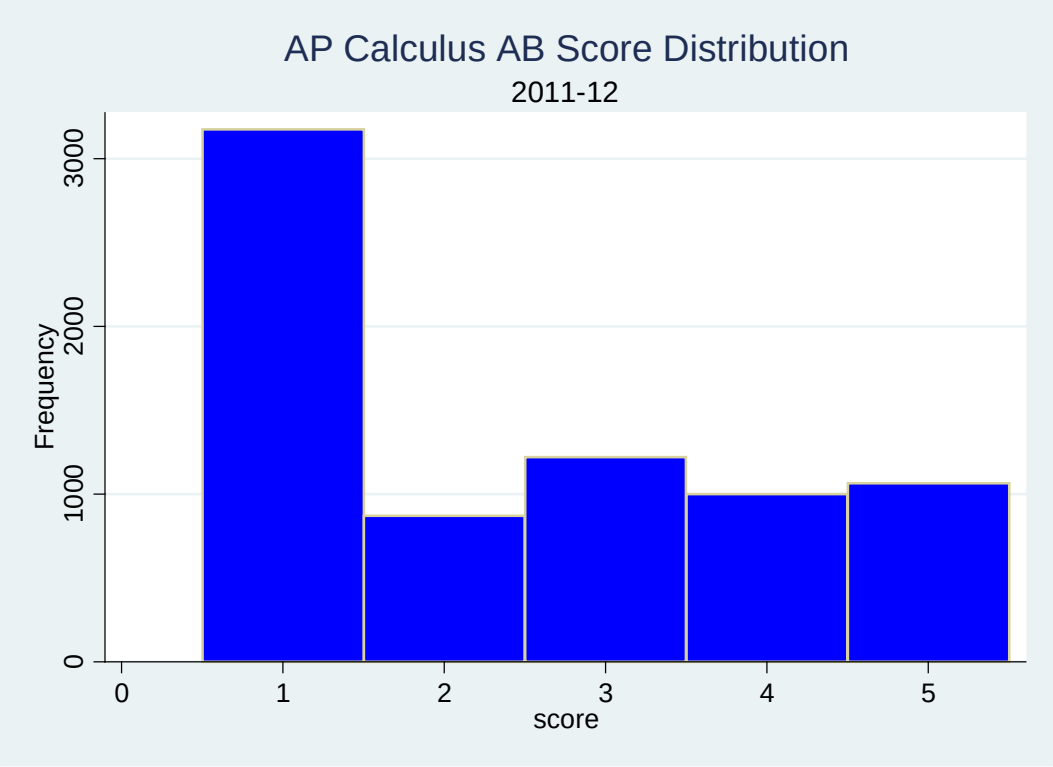
FCAT Scores Are Used as Prior Test Scores

- AP English: Grade 9 and 10 English FCAT scores
- AP Calculus: Grade 7 and 8 Math FCAT scores

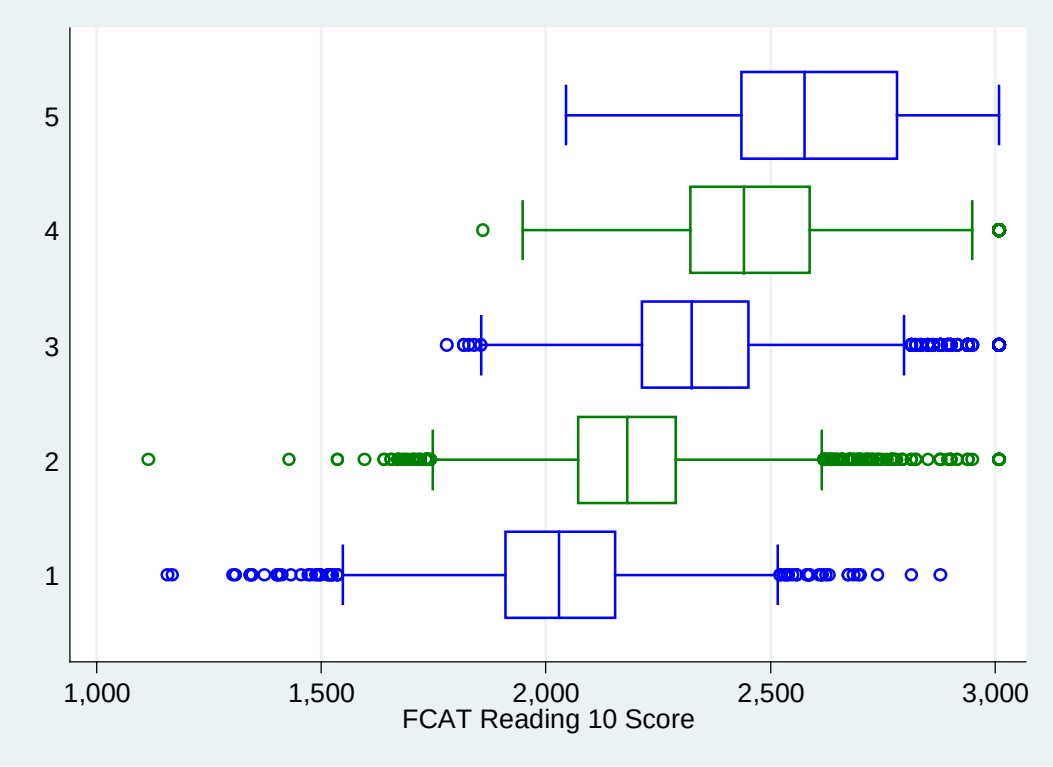
Distribution of AP English Scores



Distribution of AP Calculus AB Scores

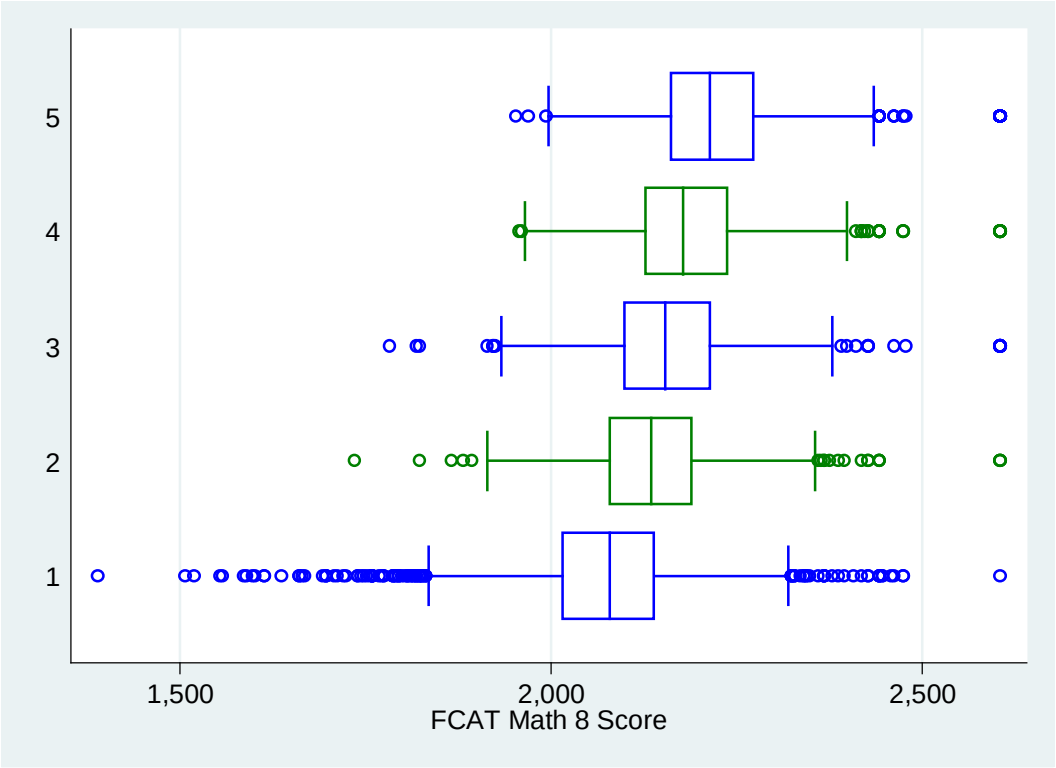


Distribution of FCAT Reading 10 Scores by AP English Score



AP

Distribution of FCAT Math 8 Scores, by AP Calculus AB Score



Both Models Are Able to Identify More and Less Effective Teachers

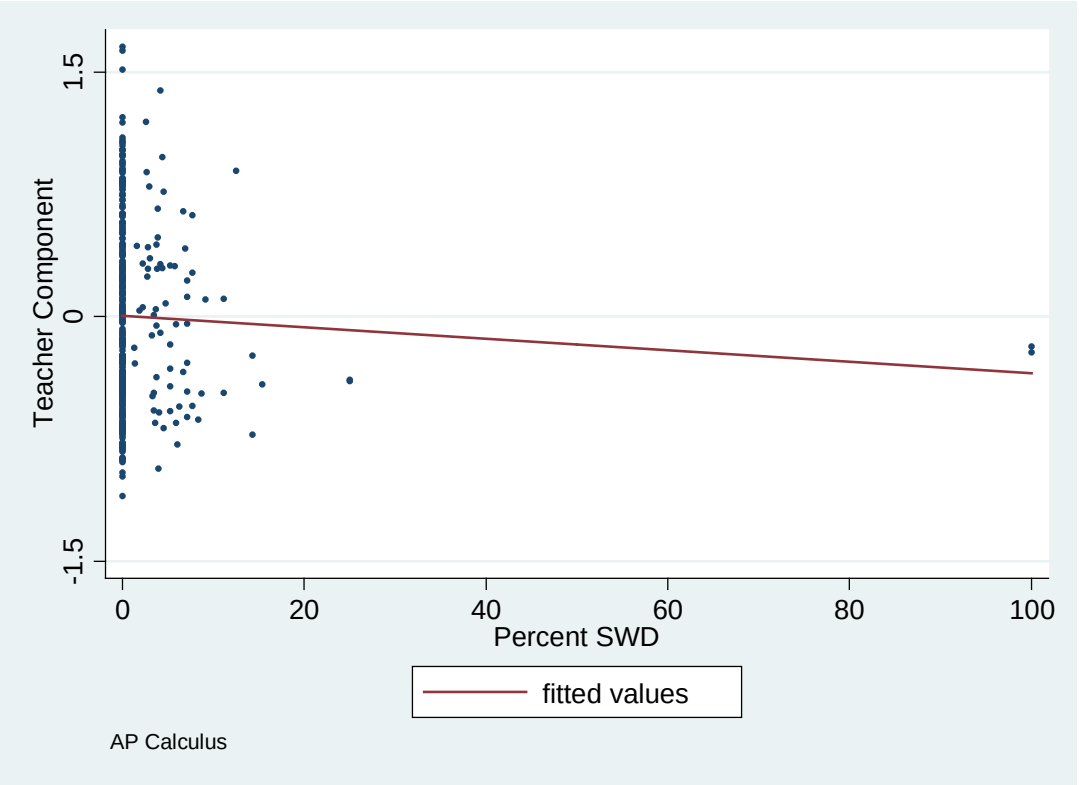
- AP English: 82 (20%) teachers are significantly above average, and 67 (17%) are significantly below average.
- AP Calculus: 126 (21%) teachers are significantly above average, and 112 (19%) are significantly below average.

Precision of the Teacher Estimates Is Uncertain

- Reliability Ratios:
  - AP English: 0.55
  - AP Calculus AB: 0.48
- Estimates are relatively precise.
- We are not able to account for measurement error, so the precision may be overstated.

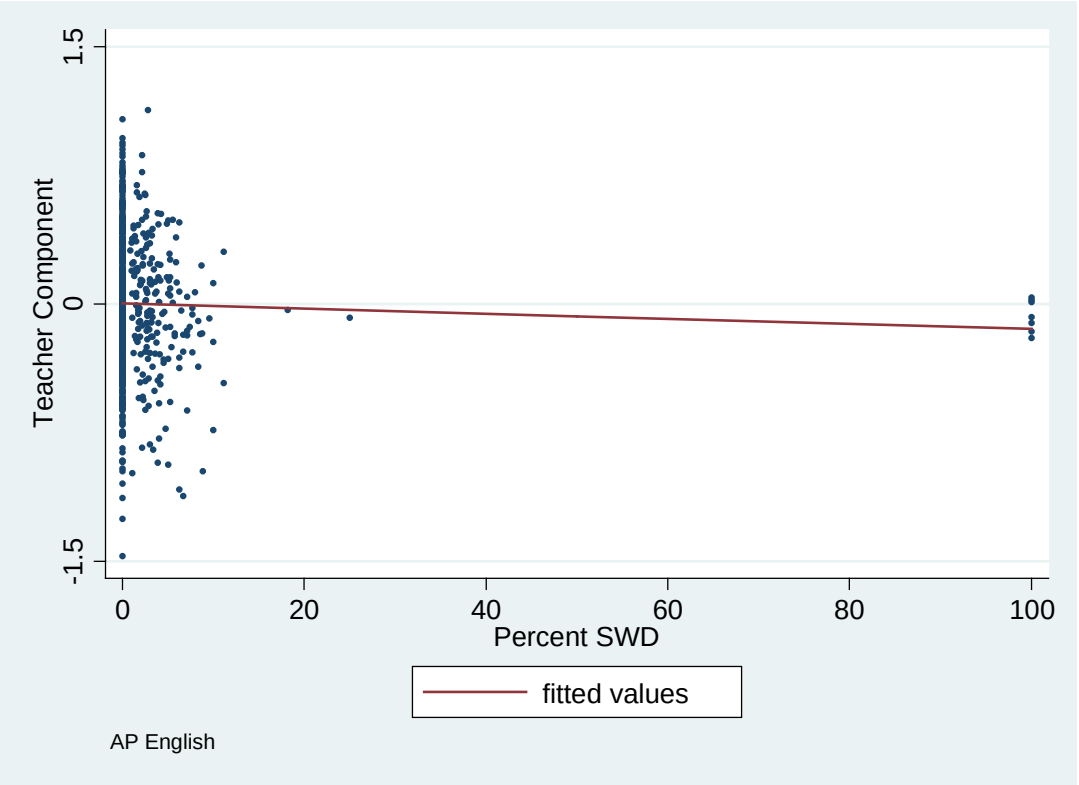
AP

Teacher Component and Percent Students with Disabilities: Calculus



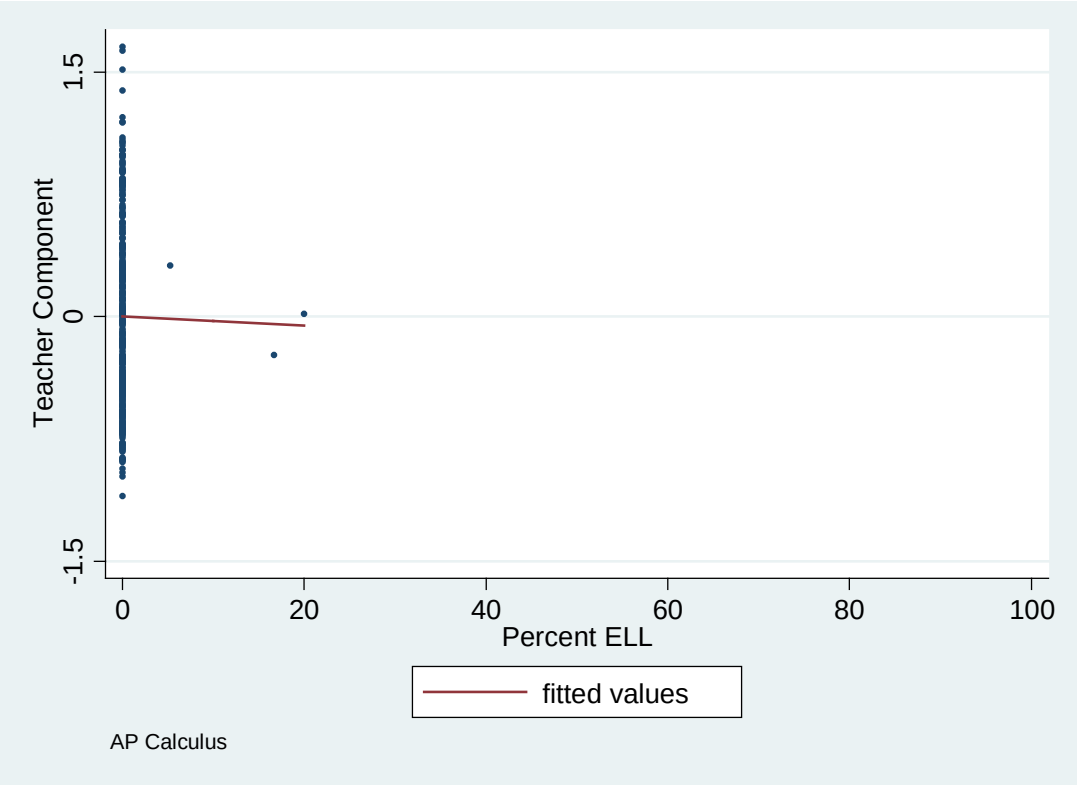
AP

Teacher Component and Percent Students with Disabilities: English



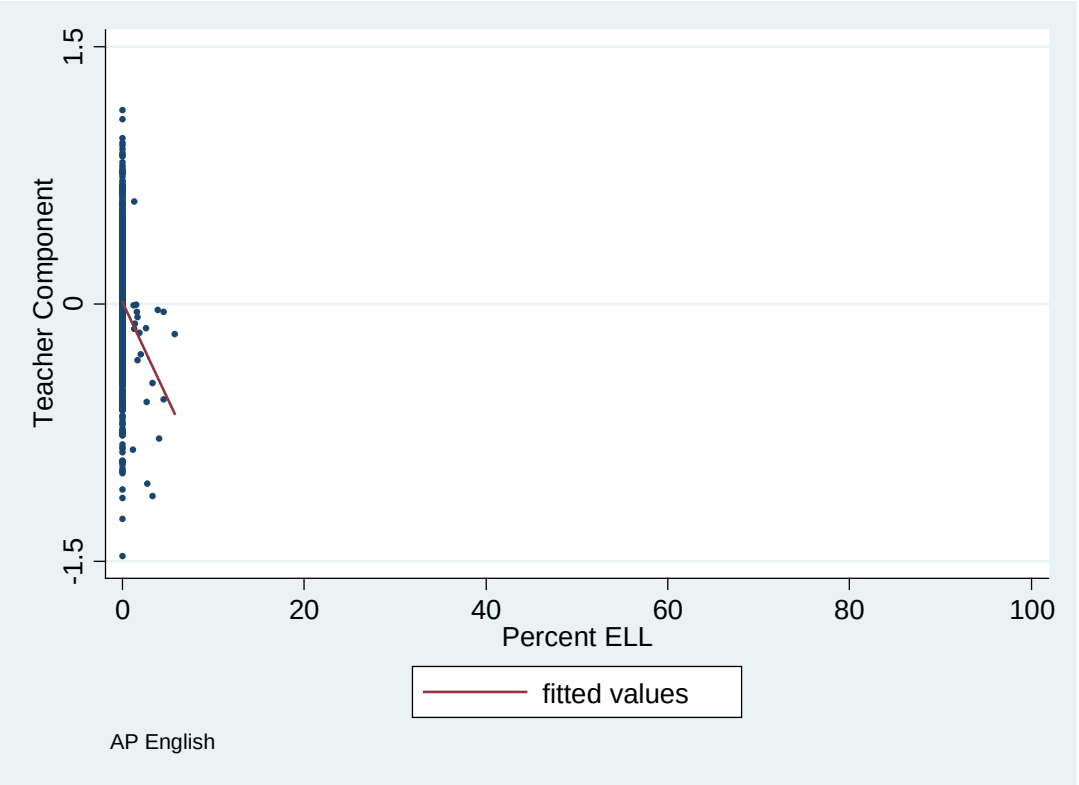
AP

Teacher Component and Percent English Language Learners: Calculus



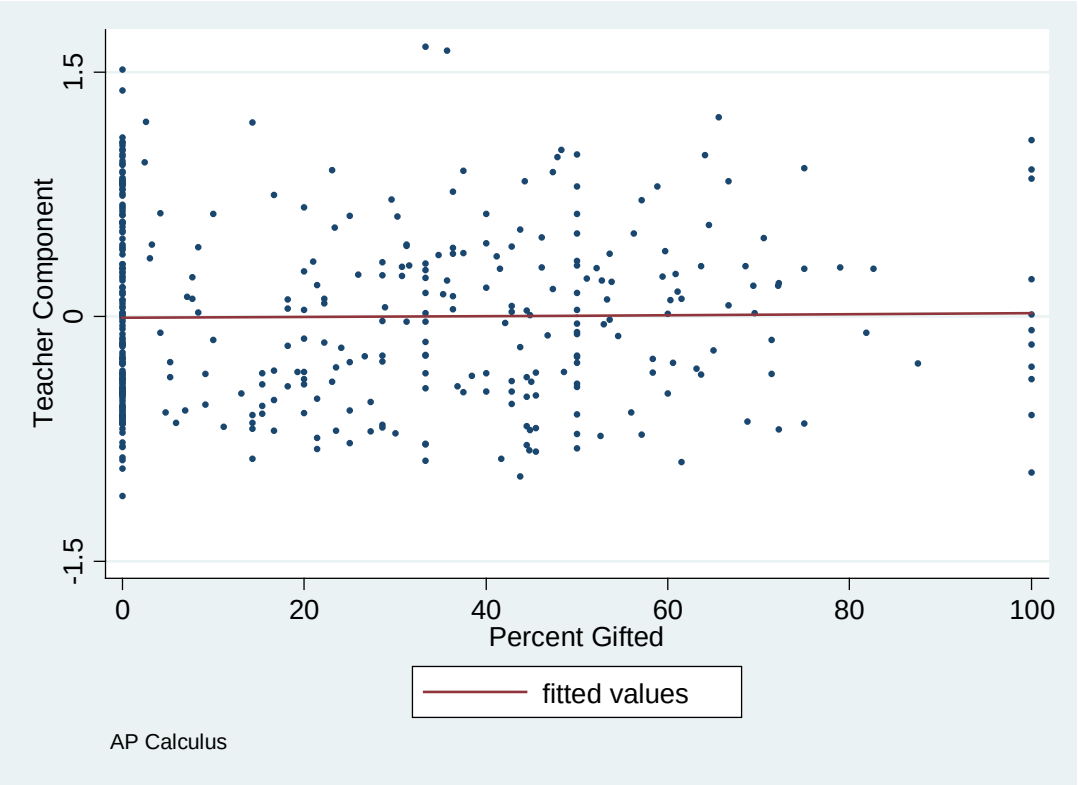
AP

Teacher Component and Percent English Language Learners: English

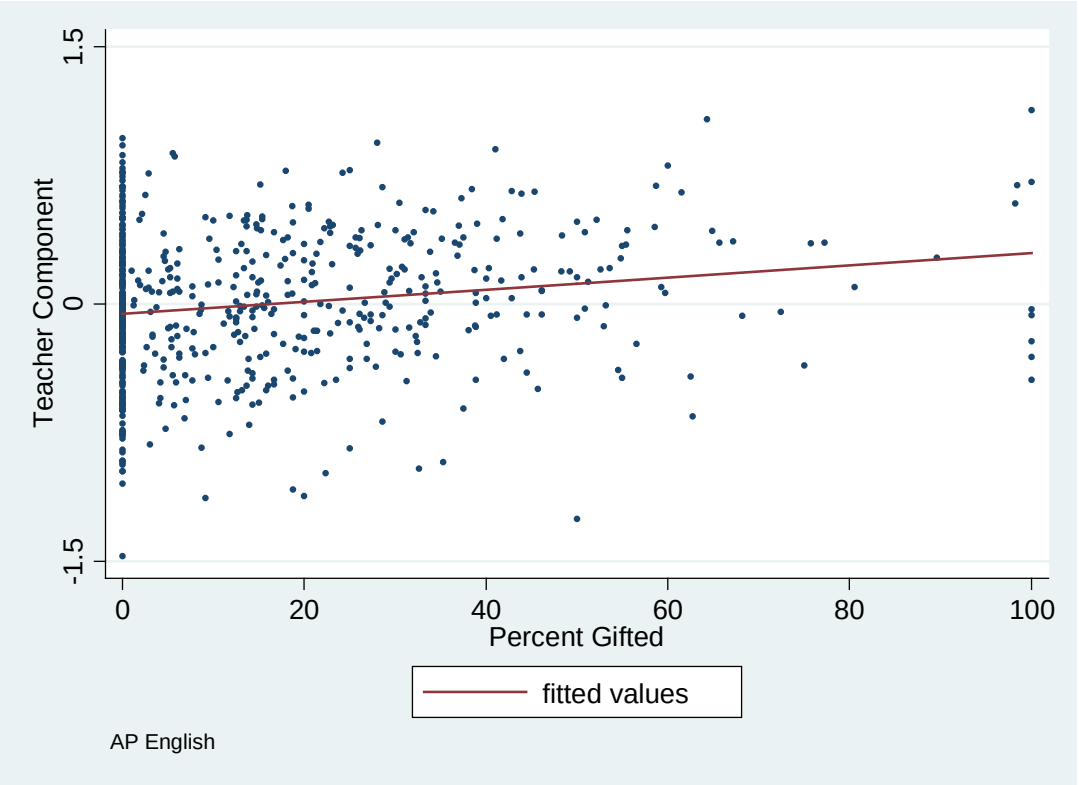


AP

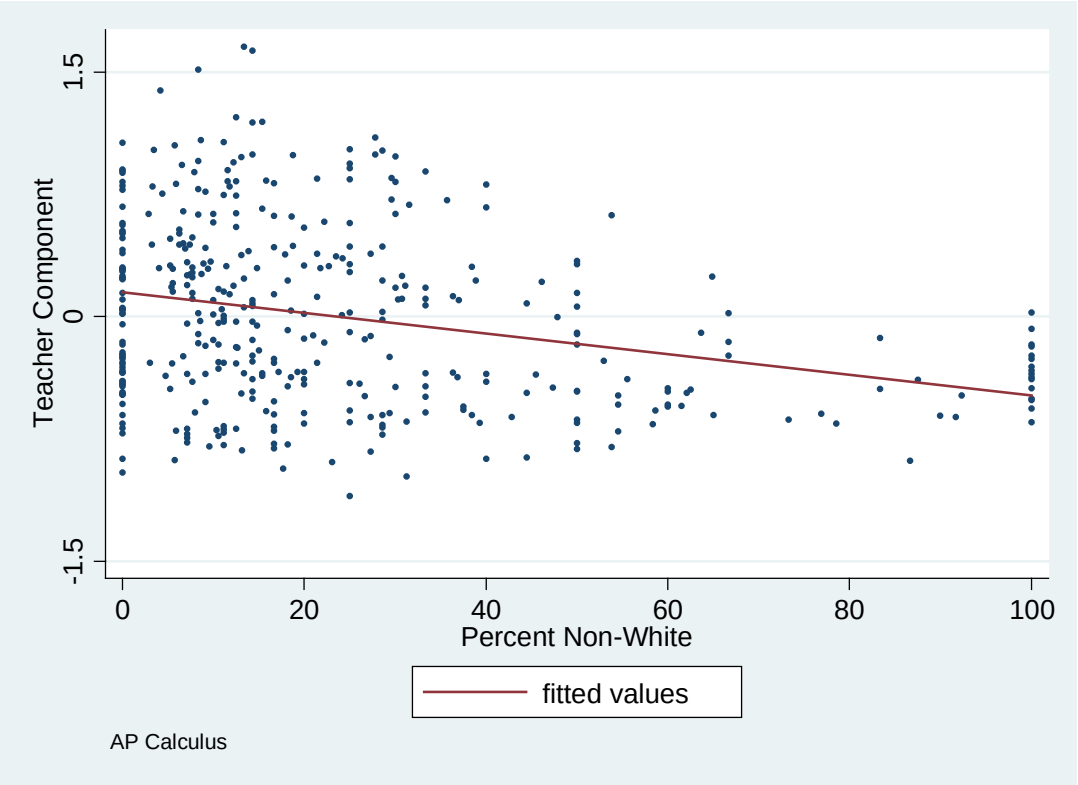
Teacher Component and Percent Gifted: Calculus



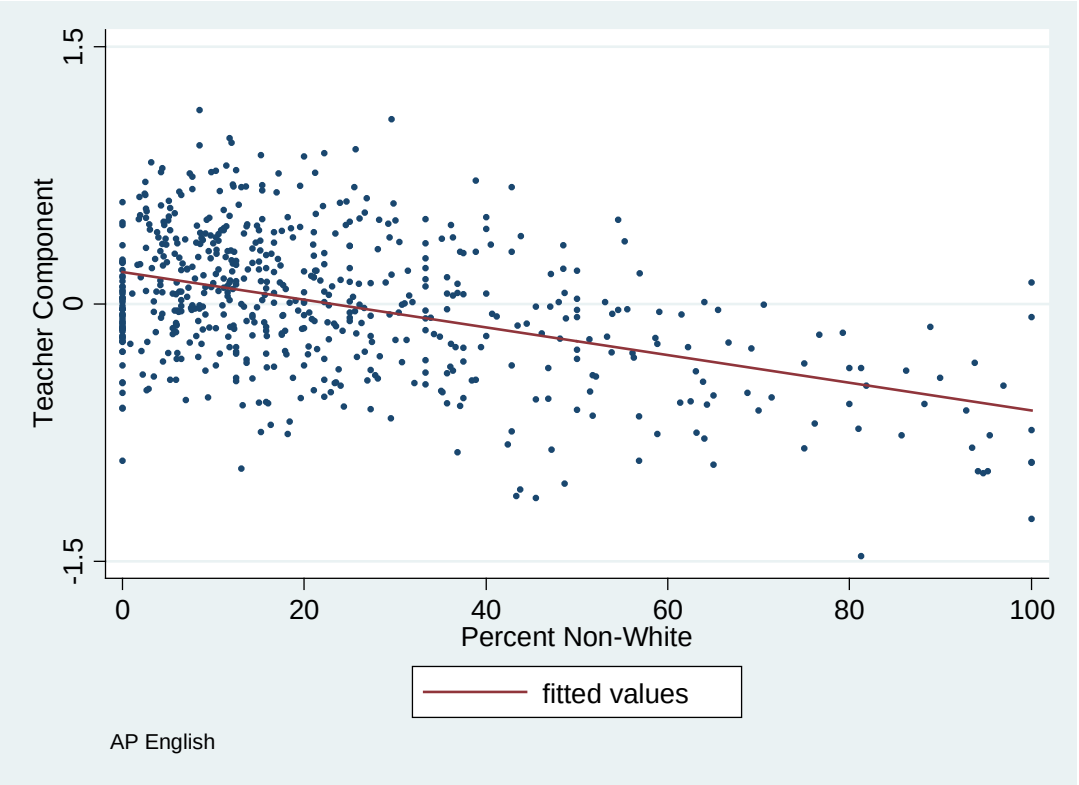
Teacher Component and Percent Gifted: English



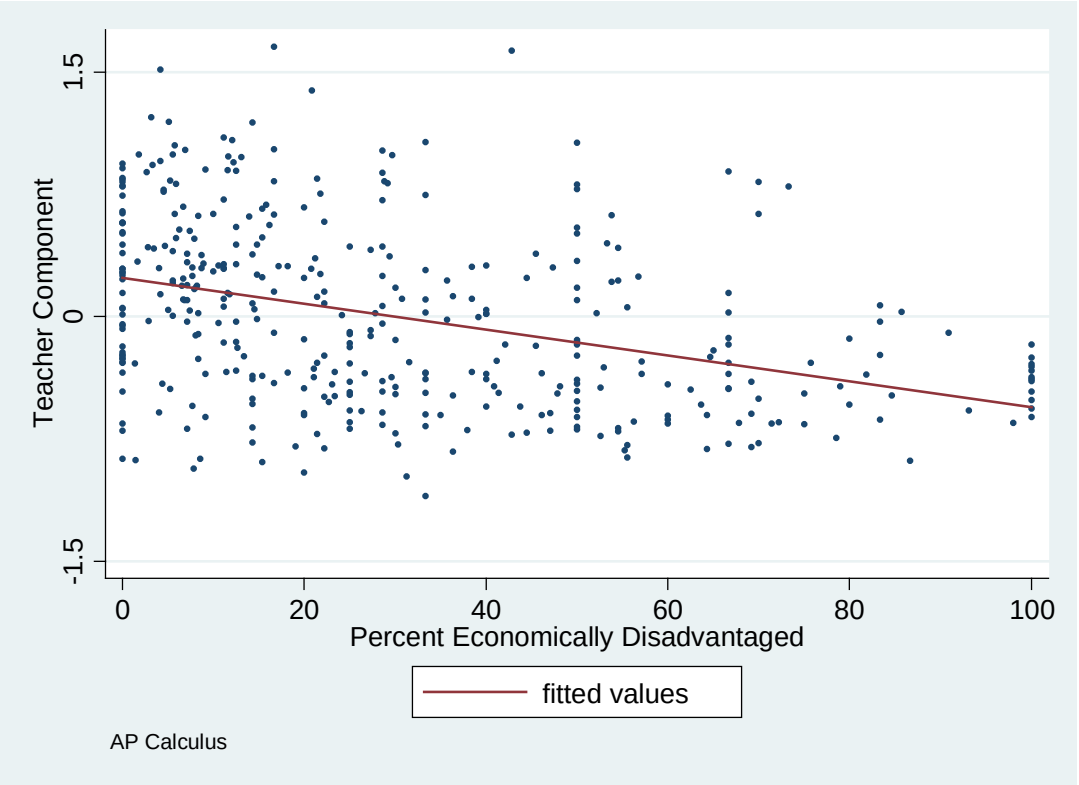
Teacher Component and Percent Non-White: Calculus



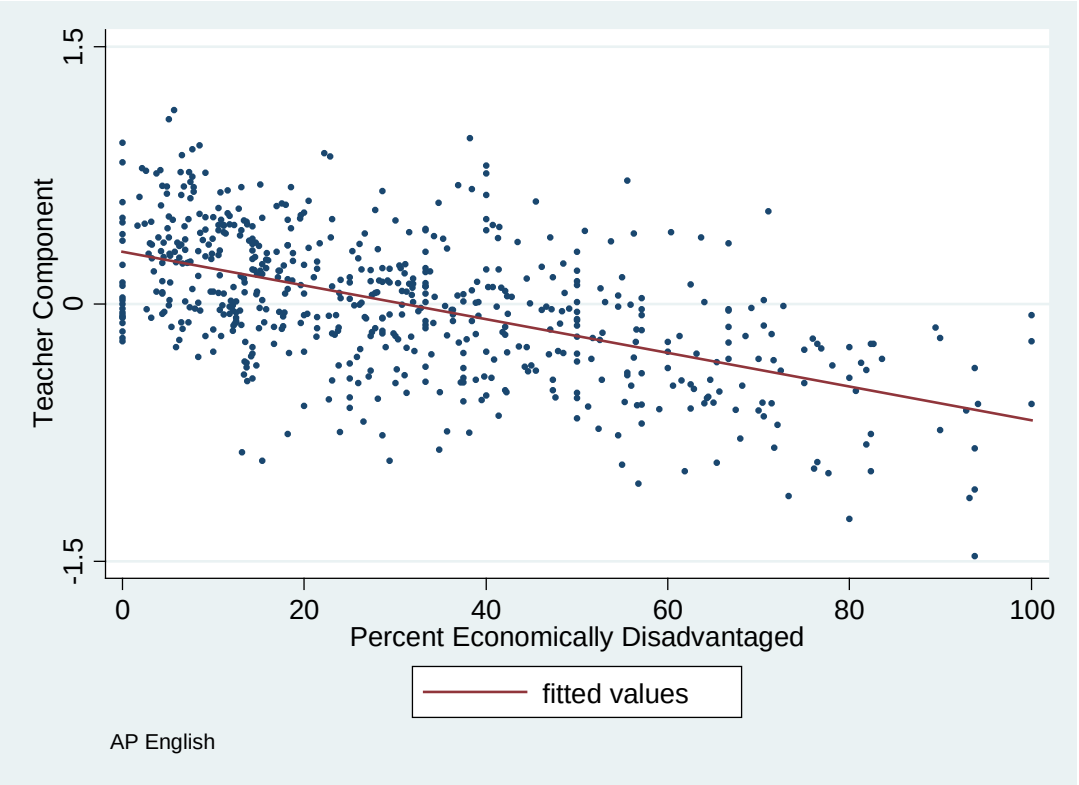
Teacher Component and Percent Non-White: English



Teacher Component and Percent Economically Disadvantaged: Calculus

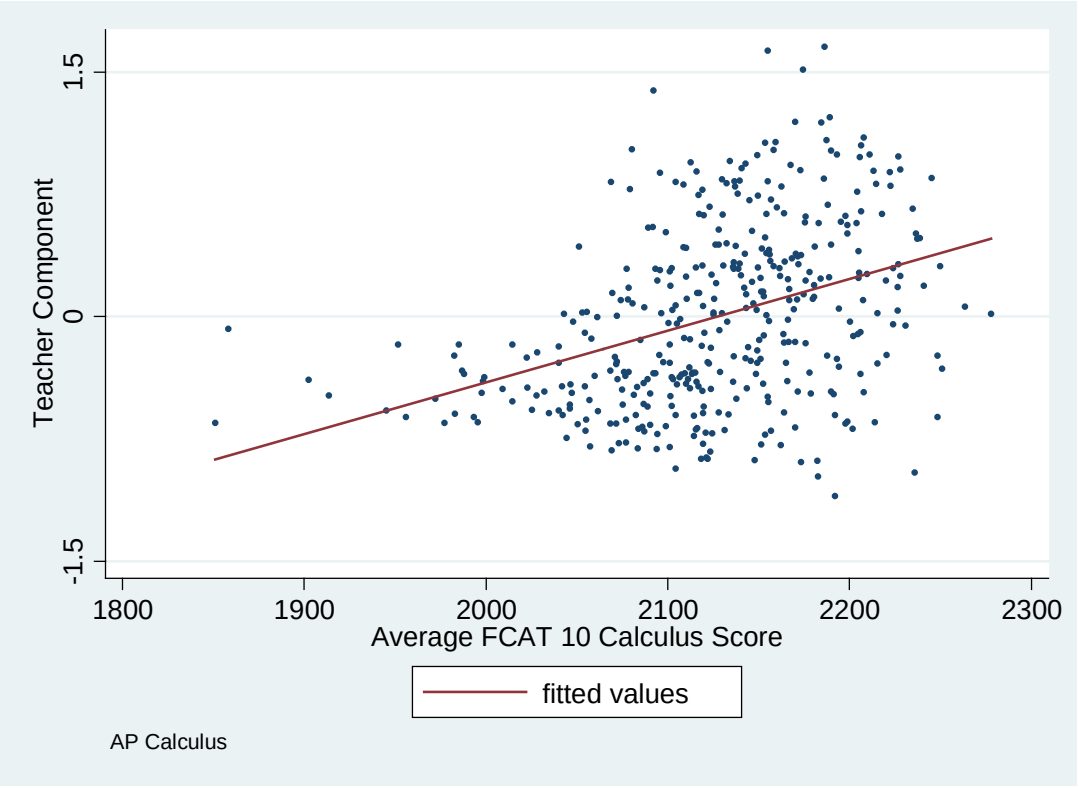


Teacher Component and Percent Economically Disadvantaged: English

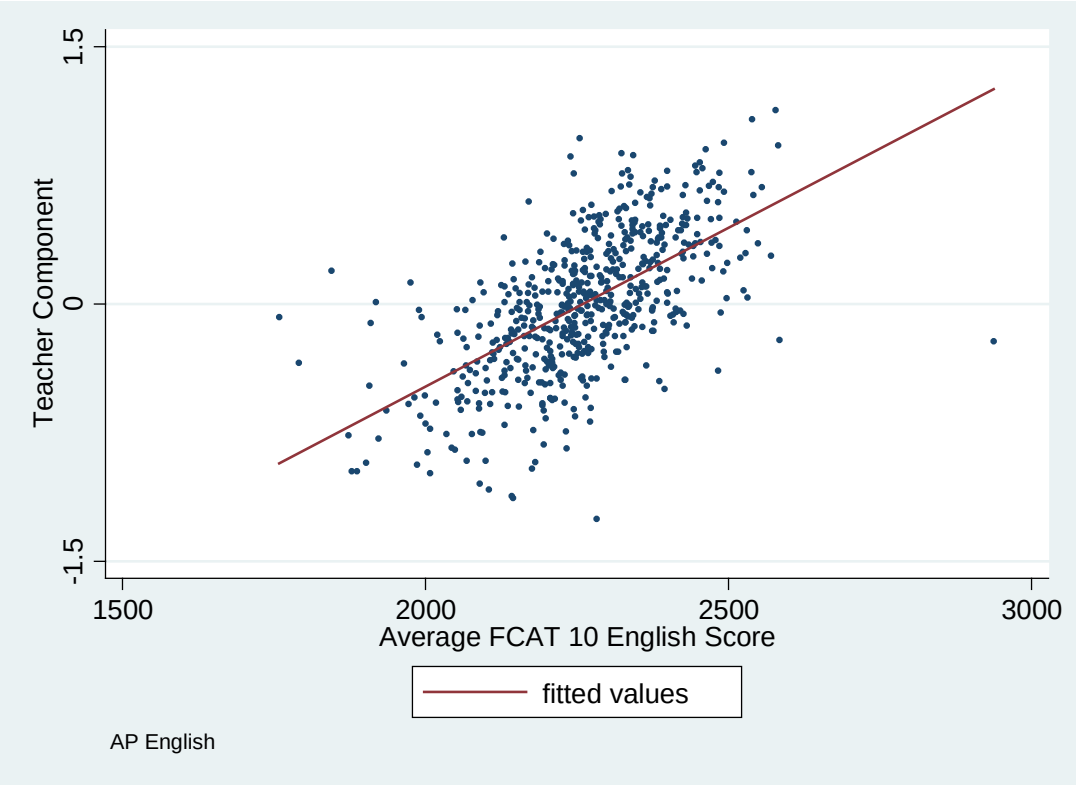


AP

Teacher Component and Average Prior FCAT Math 8 Score: Calculus



Teacher Component and Average Prior FCAT English 10 Score: English



AP

Observed Correlations with Teacher Value-Added Scores

Model	AP Calculus	AP English
Mean Prior	0.38	0.61
%ED	-0.38	-0.54
%SWD	-0.05	-0.04
%ELL	-0.01	-0.15
%Gifted	0.01	0.15
%Non-White	-0.29	-0.43

Discussion of Impact Analysis

- The impact of mean prior score, percent ED, and percent non-white is larger than the impact of other classroom characteristics.
- These correlations are larger than those we see in the other models.

Summary of Models: R-Squared and Reliability

- R-squared is similar across models (0.61 to 0.65), although the Algebra EOC models that subset by grade have a lower R-squared than the other models (0.53 to 0.54).
- Reliability is best in Geometry (0.81 to 0.84) and similar in other models (0.89 to 0.98).
- AP reliabilities are 0.48 and 0.55, perhaps due to measurement error.

Summary of Models: Variance Components

- Relative magnitudes of teacher and school variance are as expected in Algebra EOC models that exclude grades 6–8, Geometry and Biology EOC models, and SAT-10 model.
- AP models exclude school effect.

Summary of Models: Impact Data

- Correlation between percent of students who are economically disadvantaged and teacher component/teacher value-added is less than –10 across all models.
- Correlation with mean prior score is greater than 10 in Biology EOC and Geometry EOC models, Algebra EOC model 1b, and AP models.
- AP models have the largest correlations.
- Impact of other characteristics varies considerably across models.