Revisiting Options for Measuring Teacher and School Contributions to Student Growth with End-of-Course Assessments

Student Growth Implementation Committee Meeting

February 11, 2015

AIR®
American Institutes for Research®

Agenda

- 9:00-9:15 Welcome, Agenda Overview: Jason Gaitanis, *FLDOE,* Ronda Bourn, *Chair, SGIC*
- 9:15-10:00 FLDOE Updates: Jason Gaitanis, *FLDOE*
- 10:00-11:00 End of Course Model Options: Eric Larsen, *AIR*
- 11:00-11:15 Break
- 11:15-12:00 End of Course Models Continued
- 12:00-12:30 Next steps

History/Background – Going Beyond FCAT Value-Added Models

- February 2012: SGIC first considered Algebra 1 EOC model (using covariate adjustment approach)
- December 2012: SGIC considered Algebra 1 EOC models by grade along with SAT10 models
- February 2013: SGIC reviewed Algebra EOC and SAT 10 models again; also considered Biology, Geometry EOC models and AP Calculus and AP English
  - Recommended use of Grade 9 Algebra model; Grade 8 Algebra model optional
- September 2013: SGIC reviewed results of Algebra 1 2012-13 analysis
- December 2013: AIR re-analyzed 12-13 EOC data using additional approaches
  - SGIC did not take any action at that time
- February 2015: SGIC to revisit options for measuring student growth with assessments beyond FCAT/FSA

Goals of a Value-Added Model (VAM)

- Goal is to control for “sorting” of students into classes
- Necessary because students are not randomly assigned into future classes
- If sorting is not controlled for, teachers will have an advantage or disadvantage based on **who** they teach
  - Referred to as <u>selection bias</u>
- To measure teacher contributions to student learning, analysis should control for sorting to mitigate any effects associated with non-random assignment (“level the playing field”)

General Evaluation Criteria for VAM Models

- Questions to guide evaluation of the models:
  - Do the models implement a statistical approach that reasonably estimates teacher contributions to student learning?
    - The first question will be evaluated via your judgment--we will provide a model description along with benefits and risks of the different approaches
  - Do the statistical results (e.g., R squared) indicate good model fit and conform to expectations?
    - To be evaluated through data summarizing the model: size of variance components, r-squared, precision
  - Do the results of the models show differences across different classroom populations?
    - To be evaluated on the basis of statistical impact data

Summary of Feb 2013 Model Analysis

Summary of Other Models

	Biology EOC (2011-12)	Geometry EOC (2011-12)	SAT-10 Math (2010-11)	A.P. (2010-11)
Type of statistical model	Covariate adjustment	Covariate adjustment	Covariate adjustment, no meas error control	Ordered probit, no school component
Prior score data used	FCAT science; FCAT science + up to 2 prior FCAT math; FCAT science + up to 2 prior FCAT reading	Alg 1 EOC; up to 2 prior FCAT math; Alg 1 EOC + up to 2 prior FCAT math	Grade 1 SAT-10	AP English: grades 9 and 10 reading FCAT scores; AP Calculus AB: grades 7 and 8 math FCAT scores
Other covariates	Same as FCAT model	Same as FCAT model	Same as FCAT model	Same as FCAT model
Grades	8-12	8-12	2	All available

Summary of Other Models

Current Grade	Biology EOC (2011-12)	Geometry EOC (2011-12)	SAT-10 (2010-11)	A.P. (2010-11)
R-squared	0.61 to 0.63	0.62 to 0.65	0.62	N/A
Variance Components	Teacher > School	Teacher > School	Teacher > School	N/A
Impact Data: Mean Prior Score	0.18 to 0.20	0.23 to 0.26	0.15 (0.07 w/o school comp.)	0.38 (Calculus) 0.61 (English)
Impact Data: Percent ED	-0.21 to -0.22	-0.26 to -0.31	-0.27 (0.15 w/o school comp.)	-0.38 (Calculus) -0.54 (English)

Summary of 2013-14 Results

Current Grade	Math FCAT (2013-14)	Reading FCAT (2013-14)	Algebra I EOC Grade 9 (2013-14)	Algebra I EOC Grade 8 (2013-14)
R-squared	0.61 (grade 4) to 0.71	0.66 (grade 4) to 0.74	0.48	0.48
Variance Components	Teacher > School Math 6 & 7: School > Teacher	Teacher > School	Teacher > School	Teacher = School
Impact Data: Mean Prior Score	0.05	-0.02	0.07	0.17
Impact Data: Percent ED	-0.07	-0.04	-0.02	-0.09

Summary of New EOC Models Presented December 2013

AMERICAN INSTITUTES FOR RESEARCH®

New EOC Analysis (December 2013)

- Other EOC models (beyond Algebra grades 8 and 9) were originally not acted on by SGIC
  - We observed “reversals” in the variance component patterns
  - Impact data showed very high correlations between teacher scores and classroom composition
  - R-squared values and precision were relatively low
- To address these issues, AIR experimented with new models that analyze the data in different ways
- The aim is to determine if a different modeling strategy can improve on the approach that has been used in Florida to date

Objectives

- Following the SGIC's direction, we implemented 6 different analyses with EOC data to see if new methods can improve on previous approach
  - Models 1-3: Enhanced covariate adjustment models; Model 4: Z-score; Model 5: Percent proficient; Model 6: Probability of proficiency
- Focus on grade 9 Algebra EOC in order to make comparisons to implemented/recommended model
- The primary aim is to determine if other models can improve on the covariate adjustment model approach previously used for EOC assessments

Models 1-3: Enhanced Covariate Adjustment Model

- Some researchers have proposed that high school students are often sorted into different academic tracks
- If this tracking is correlated with sorting, then it would be necessary to control for course tracking to mitigate selection bias
- In Models 1-3, we control for students' prior math courses in addition to their prior test scores

Grade 8 Math Courses of Students Taking the Algebra I EOC in Grade 9

- Algebra I (3.4%)
- Algebra I Honors (2.5%)
- Algebra Ia (5.6%)
- M/J Intensive Mathematics (MC) (11.4%)
- M/J Mathematics 3 (56.4%)
- M/J Mathematics 3, Advanced (19.3%)
- Others (1.4%)

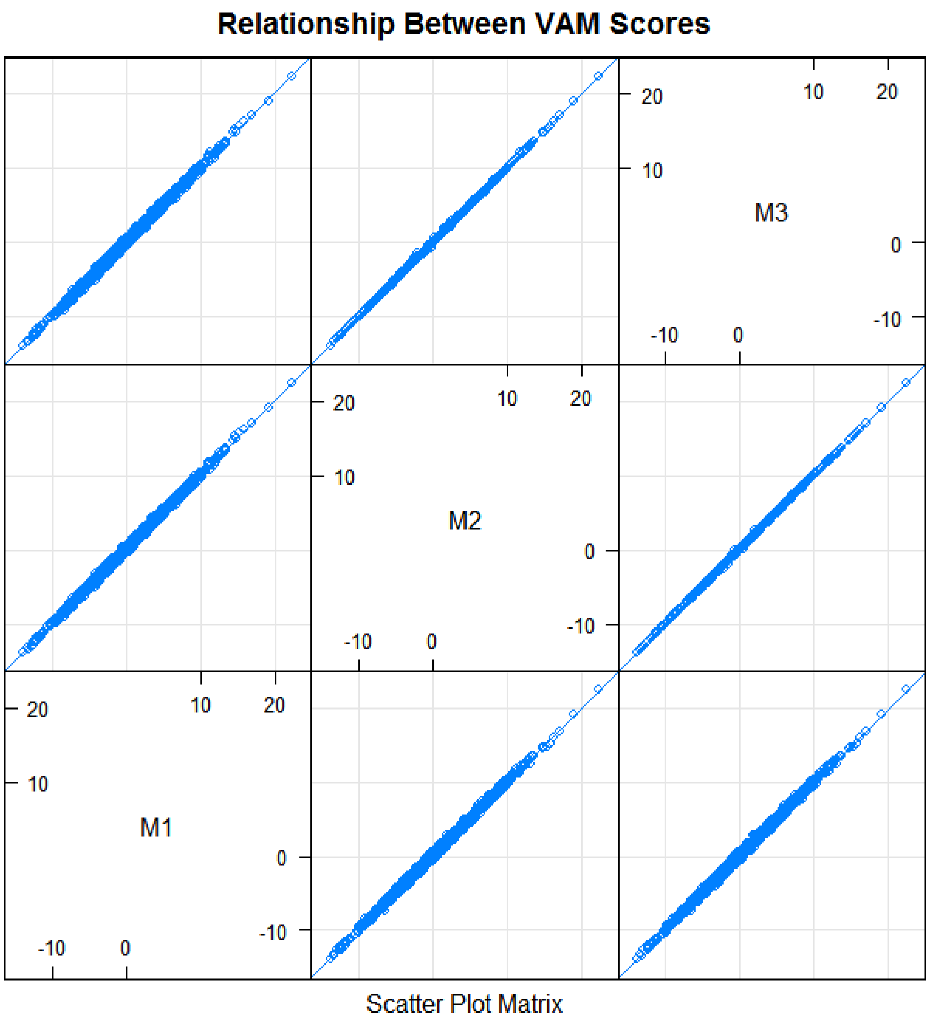
Summary of Models 1-3

- Model 1:
  - Control for two prior test scores
  - Control for mean prior score of students in class
  - School and teacher random effects
- Model 2:
  - Model 1 + prior course random effects
- Model 3:
  - Model 1 + prior course fixed effects

Summary of “Beta” EOC Models

	2012-13 Grade 9 Algebra Model	Baseline	Random Effects	Fixed Effects	Z-Score	Pct. Prof	Prob. Prof
R-squared	0.52	0.49	0.49	0.50			
Variance Components	Teacher > School	Teacher > School	Teacher > School	Teacher > School			
Impact Data: Mean Prior Score	0.058	0.093	0.095	0.095			
Impact Data: Percent ED	-0.043	-0.087	-0.086	-0.086			

Correlation Between Teacher VAM Scores: Models 1-3



Similar Models Were Implemented for the Geometry EOC

- Models were implemented separately for grade 9 and grade 10
- Three models were run for each grade
  - The baseline model including only prior scores as covariates
  - A model that includes course histories as random effects
  - A model that includes course histories as fixed effects
- The conclusions form these models were the same as for the Algebra I EOC: controlling for course history adds almost no explanatory power to the models

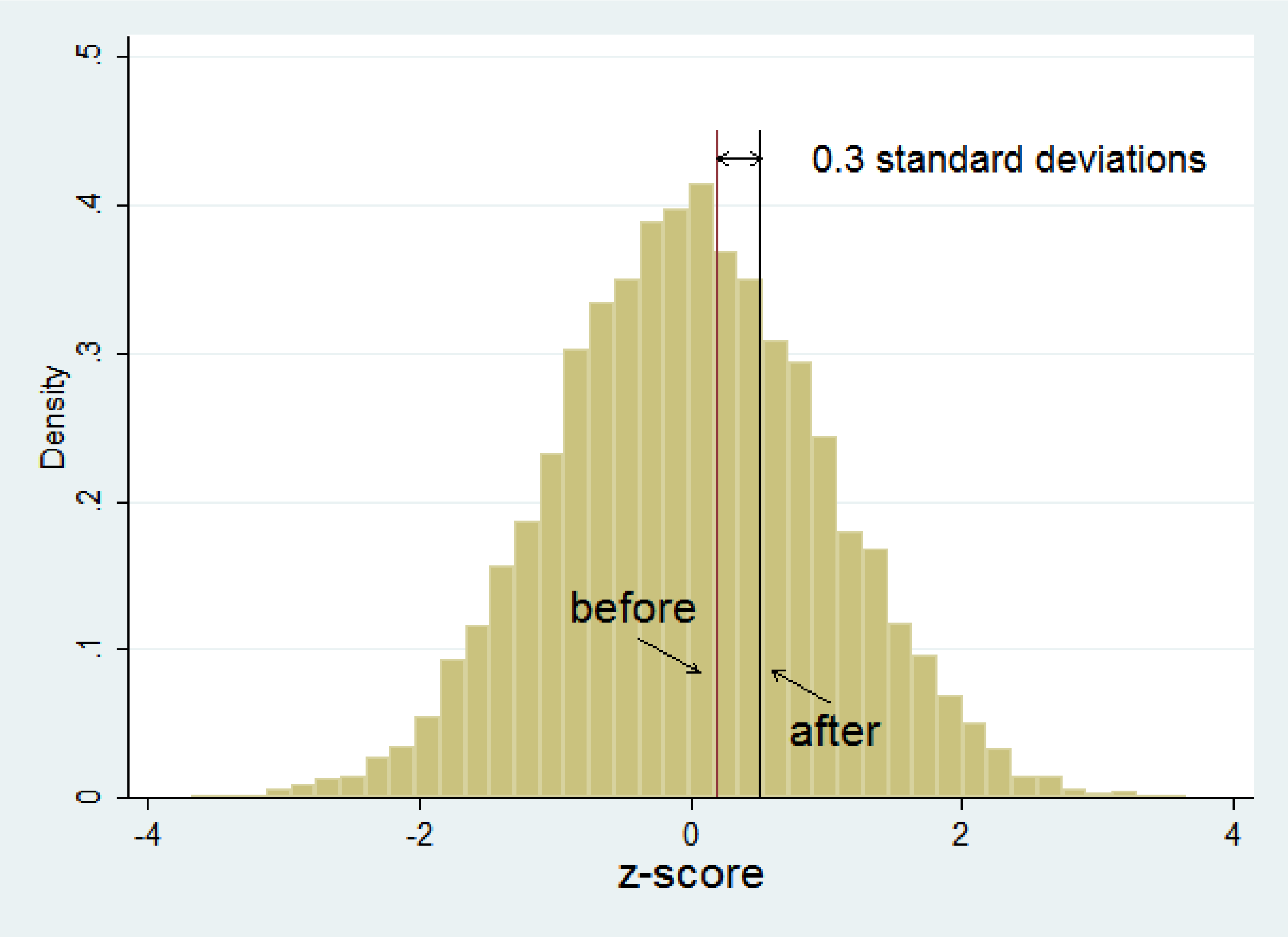
Introduction to Models 4-6

- These models are different from the linear covariate adjustment models used for FCAT and Algebra I
  - Statistical summaries previously presented do not necessarily apply since outcomes are different
- Model 4: Z-Score Model
  - How much do the teacher’s students move up/down relative to other students?
- Model 5: Percent of Students Achieving Proficiency
- Model 6: Probability of Proficiency
  - Measures impact of teacher on the probability the student achieves proficiency on Algebra I EOC

Model 4: Z-Score Model

- Measure where in the overall state distribution of student scores each student's grade 8 math score falls
- Measure where in the overall distribution of student scores each student's Algebra I EOC score falls
- Compare the two for each student to determine how much the student moved up or down in the overall distribution of student scores
  - Positive: moved up in the distribution
  - Negative: moved down in the distribution
  - Zero: stayed in the same place relative to other students

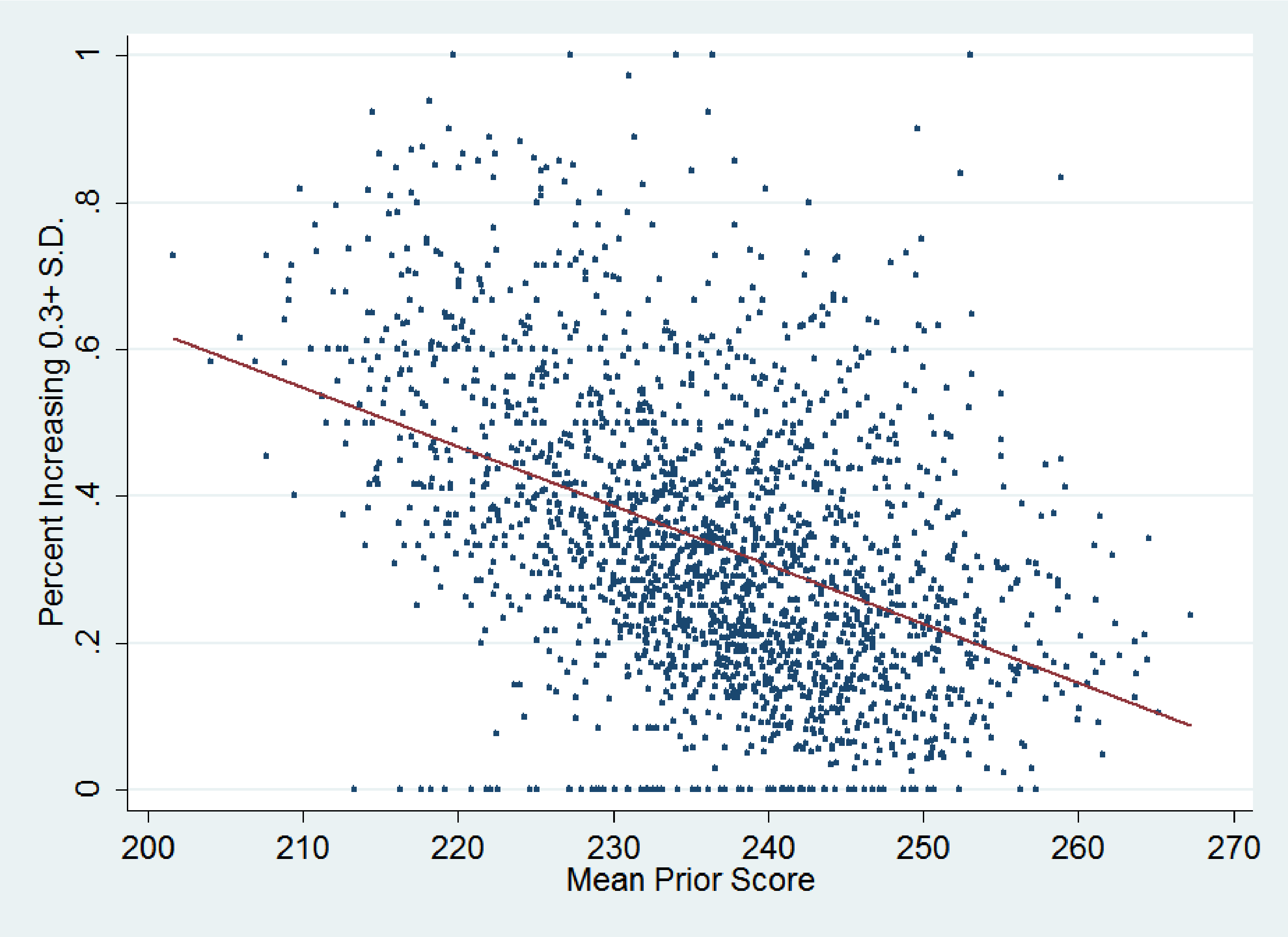
Student EOC Scores Converted to Z-Scores



Model 4: Z-Score Model

- Teacher's score = share of students who move up more than 0.3 standard deviations (s.d) in the distribution
  - Moving from the mean to 0.3 s.d. above the mean on 2012-13 Algebra I EOC is equivalent to moving up 13 percentile points in the distribution.
- Assumes all students are equally likely to move up 0.3 s.d. conditional on their prior scores.
- Relatively difficult for students with high grade 8 scores to move up 0.3 s.d. (due to measurement error/regression to mean)
- Relatively easy for students with very low grade 8 scores to move up 0.3 s.d. (due to measurement error/regression to mean)
- Unlike Model 5 (percent achieving proficiency), Model 4 puts teachers of students with high grade 8 scores at a disadvantage

Model 4: Z-Score Model



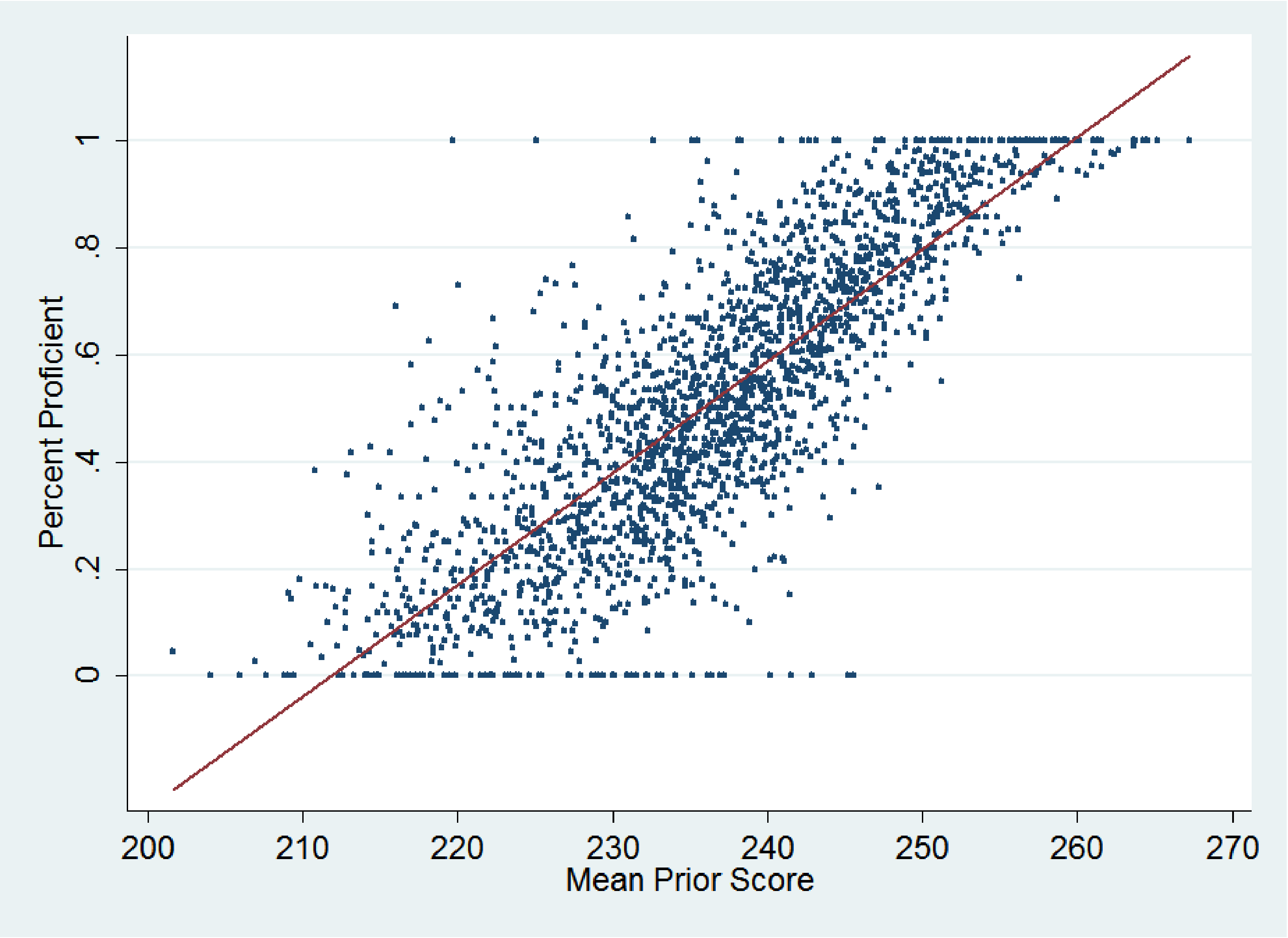
Summary of “Beta” EOC Models

	2012-13 Grade 9 Algebra Model	Baseline	Random Effects	Fixed Effects	Z-Score	Pct. Prof	Prob. Prof
R-squared	0.52	0.49	0.49	0.50	N/A		
Variance Components	Teacher > School	Teacher > School	Teacher > School	Teacher > School	N/A		
Impact Data: Mean Prior Score	0.058	0.093	0.095	0.095	-0.402		
Impact Data: Percent ED	-0.043	-0.087	-0.086	-0.086	0.173		

Model 5: Percent Achieving Proficiency

- Approach commonly associated with AYP
- Teacher rating is the share of students achieving proficiency (scoring above 399)
- Does not control for sorting
- Assumes students are randomly distributed across schools
- Does not control for prior test scores or any other covariates

Model 5: Percent Achieving Proficiency



Model 5: Percent Achieving Proficiency

- Teacher scores are highly correlated with students' prior scores
- Models such as this are useful in accountability systems when the emphasis is primarily based on identification of classrooms where students achieve a passing score
- These models typically provide different information about classrooms than is observed with growth models, but the percentage of students achieving proficiency is still a valuable outcome

Summary of “Beta” EOC Models

	2012-13 Grade 9 Algebra Model	Baseline	Random Effects	Fixed Effects	Z-Score	Pct. Prof	Prob. Prof
R-squared	0.52	0.49	0.49	0.50	N/A	N/A	
Variance Components	Teacher > School	Teacher > School	Teacher > School	Teacher > School	N/A	N/A	
Impact Data: Mean Prior Score	0.058	0.093	0.095	0.095	-0.402	0.807	
Impact Data: Percent ED	-0.043	-0.087	-0.086	-0.086	0.173	-0.378	

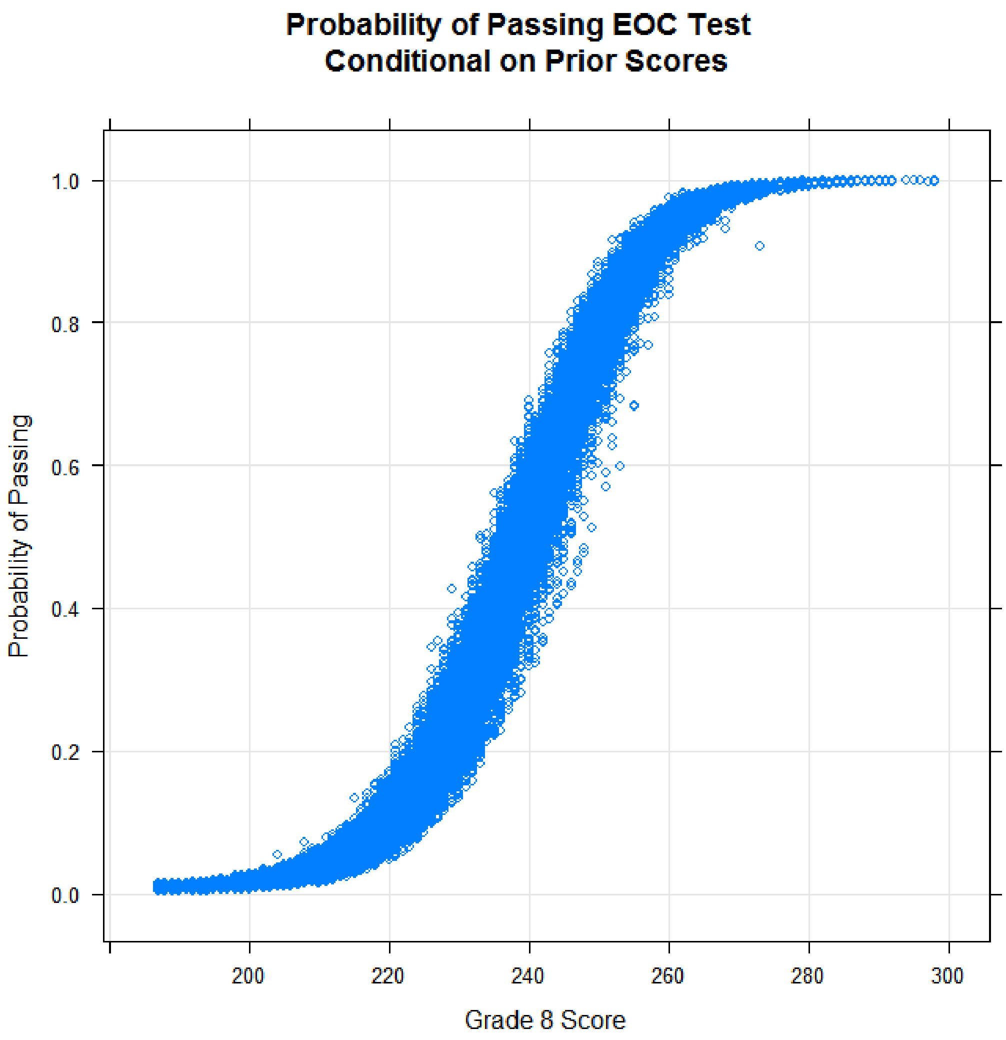
Model 6: Probability of Proficiency

- Use a student's prior test scores to estimate the probability the student will score above the proficiency cut-point
- Students with higher prior test scores have a higher predicted probability of passing
- Other covariates (SWD status, ELL status, prior course history, etc.) can be included in the model as well
- Conditional on a student's prior test scores (and possibly other covariates), we can determine whether some teachers' students are more likely to pass than other teachers' students

Model 6: Probability of Proficiency

- Model assumes that conditional on prior test scores and other included covariates, students are randomly distributed across teachers and schools
- If on average a teacher's students had a low probability of passing, but many of these students passed the cut-off, that teacher would receive a high score
- If a teacher's students pass or do not pass about as expected, that teacher would receive an average score
- If fewer of a teacher's students passed than was expected, based on their prior test scores, that teacher would receive a low score

Probability of Proficiency Model



Compare Actual to Predicted

- Share of outcomes correctly predicted is one measure of model fit
- Model correctly predicts passage for 77% of students

Pass Rates		Actual	
		Not Pass	**Pass**
Predicted	**Not Pass**	34678 (36.1%)	12620 (13.1%)
	Pass	9117 (9.5%)	39612 (41.3%)

Summary of "Beta" EOC Models

	2012-13 Grade 9 Algebra Model	Baseline	Random Effects	Fixed Effects	Z-Score	Pct. Prof	Prob. Prof
R-squared	0.52	0.49	0.49	0.50	N/A	N/A	(Correctly Predicts 77%)
Variance Components	Teacher > School	Teacher > School	Teacher > School	Teacher > School	N/A	N/A	N/A
Impact Data: Mean Prior Score	0.058	0.093	0.095	0.095	-0.402	0.807	0.243
Impact Data: Percent ED	-0.043	-0.087	-0.086	-0.086	0.173	-0.378	-0.127

Summary of Models 4-6

- Model 4 (Z-Score):
  - Rewards teachers whose students make significant growth in the overall distribution of student scores
  - Disadvantages teachers whose students have *high* math 8 scores
- Model 5 (Percent of Students Achieving Proficiency):
  - Measures share of students who achieve proficiency
  - Similar to AYP
  - Disadvantages teachers whose students have low math 8 scores
- Model 6 (Probability of Proficiency):
  - Measures teachers’ impact on the probability a student achieves proficiency
  - Has advantages similar to covariate adjustment model

Correlations Between “Beta” EOC Models

	Baseline	Random Effects	Fixed Effects	Z-Score	Pct. Prof	Prob. Prof
Baseline	1	0.999	0.999	0.569	0.423	0.618
Random Effects	0.999	1	0.99	0.567	0.424	0.616
Fixed Effects	0.999	0.99	1	0.567	0.424	0.616
Pct. Prof	0.423	0.424	0.424	0.007	1	0.721
Z-Score	0.569	0.567	0.567	1	0.007	0.489
Prob. Prof	0.618	0.616	0.616	0.489	0.721	1

Summary

- Controlling for students' prior courses does little to improve predictive power of covariate adjustment models
- "Percent achieving proficiency" and z-score models do not control for sorting
- The benefits of the "probability of proficiency" models come close to those of the covariate adjustment models

Summary of “Beta” EOC Models

	2012-13 Grade 9 Algebra Model	Baseline	Random Effects	Fixed Effects	Z-Score	Pct. Prof	Prob. Prof
R-squared	0.52	0.49	0.49	0.50	N/A	N/A	(Correctly Predicts 77%)
Variance Components	Teacher > School	Teacher > School	Teacher > School	Teacher > School	N/A	N/A	N/A
Impact Data: Mean Prior Score	0.058	0.093	0.095	0.095	-0.402	0.807	0.243
Impact Data: Percent ED	-0.043	-0.087	-0.086	-0.086	0.173	-0.378	-0.127

Summary of 2013-14 Results

Current Grade	Math FCAT (2013-14)	Reading FCAT (2013-14)	Algebra I EOC Grade 9 (2013-14)	Algebra I EOC Grade 8 (2013-14)
R-squared	0.61 (grade 4) to 0.71	0.66 (grade 4) to 0.74	0.48	0.48
Variance Components	Teacher > School Math 6 & 7: School > Teacher	Teacher > School	Teacher > School	Teacher = School
Impact Data: Mean Prior Score	0.05	-0.02	0.07	0.17
Impact Data: Percent ED	-0.07	-0.04	-0.02	-0.09

Summary of Other Models

Current Grade	Biology EOC (2011-12)	Geometry EOC (2011-12)	SAT-10 (2010-11)	A.P. (2010-11)
R-squared	0.61 to 0.63	0.62 to 0.65	0.62	N/A
Variance Components	Teacher > School	Teacher > School	Teacher > School	N/A
Impact Data: Mean Prior Score	0.18 to 0.20	0.23 to 0.26	0.15 (0.07 w/o school comp.)	0.38 (Calculus) 0.61 (English)
Impact Data: Percent ED	-0.21 to -0.22	-0.26 to -0.31	-0.27 (0.15 w/o school comp.)	-0.38 (Calculus) -0.54 (English)

Summary of Other Models

- Evidence suggests FCAT and Grade 9 Algebra EOC models control effectively for selection
- Evidence is not as strong for other EOCs, SAT-10, APs
  - The problem does not appear to be the functional form of the models
  - There are not strong predictors for these EOCs, SAT-10, and AP that effectively control for non-random sorting

Discussion

AMERICAN INSTITUTES FOR RESEARCH®

Discussion

- Are these models better than alternatives available to districts?
- Does the SGIC recommend consideration of a covariate adjustment or other modeling approach for any of the following assessments?
  - Geometry
  - Biology
  - U.S. History
  - Civics
  - FCAT Science
- If yes, what type of approach is recommended and with what parameters (e.g. what prior scores)?