COVID-Sim: Simulating Spring 2020 Assessment Scores

Prepared for: Tennessee Department of Education

credo

The Center for Research on Education Outcomes
Stanford University
credo.stanford.edu

INTRODUCTION

Motivation for Project

- CREDO wanted to support state education agency partners dealing with impacts of COVID-19.
- Loss of Spring 2020 assessments affects more than ESSA accountability.
- CREDO holds recent data from many states.
- CREDO has methodological expertise to conceive and execute simulations for Spring 2020 scores.
- Ultimately 20 states will receive proxies – two different sets
  - Full-year proxies – can be used for planning and gap analysis
  - Proxies adjusted for COVID-impacted learning loss – to help educators plan for student return in 2020-2021.

Table of Contents

There are many uses of assessment data in State and District operations.

Assessment Data Use Cases	SEA	LEA
Performance frameworks / Accountability / School report cards	X	X
USDE reporting	X	
SEA reporting		X
School Improvement programs	X	X
Need-based resource allocations	X	X
Curriculum review	X	X
Instructional improvement / professional development		X
Teacher compensation	X	X
Charter school authorizing / review	X	X
Program / policy evaluation	X	X
Research	X	X
Grant seeking	X	X
Economic / community development	X	X

Table of Contents

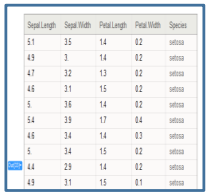
Project Overview

PHASE 1: Preparation

IDEA
Can we..??

CONCEPT
Mapped Options

PERMISSION
Partner Support



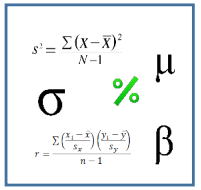
ORGANIZE TEST BED
14-15 to 18-19 Data

PHASE 2: Build Simulations

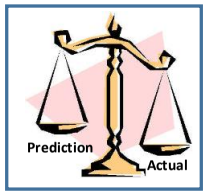
Based on 2017-2018

CODE
Program Scenarios

TSUNAMI !!
Tons of Output



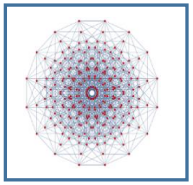
CHOOSE DIAGNOSTICS
Common Criteria



RUN DIAGNOSTICS
Compare Scenarios

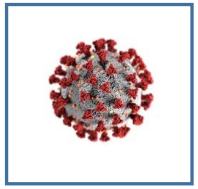
PHASE 3: Create Spring20 Proxies

SELECT SCENARIO
Best Predictor



PROJECT SPRING20
Full-Year Proxy

LEARNING LOSS
Based on MAP



ADJUST SPRING20
For COVID Loss

Table of Contents

General Approach

- 5 years of longitudinal student data serves as a test bed.
  - 2014-15 through 2018-19. For Tennessee, we took 2014-15 and 2016-17 scores to impute 2015-16 scores because the Tennessee test score data for 2015-16 were incomplete.
  - ESSA-mandated assessment data: 3rd-8th grade and HS achievement
  - Converted scale scores to standardized ones for simulation, then translate back for deliverables
- Pretend one year of scores is missing.
  - We used 2017-18
- Simulate the missing 2017-18 score from remaining data using various options.
  - Started out with 5 alternatives for non-Tennessee states; added more along the way. See Technical Appendices 1 and 3 for details.
  - Started out with 4 alternatives for Tennessee because the way in which 1516 scores are imputed makes the 5th alternative unfeasible; added more along the way.
- Test the predicted 2017-18 score against the actual 2017-18 score.
  - Overall accuracy
  - Focused accuracy – student subgroups, grade level results and school attributes
- Apply best performing simulation to recent years of data to estimate three sets of 2020 scores.
  - Full-year estimates as if COVID had not occurred
  - Adjusted for learning loss
    - Achievement at point of school building closure in March
    - Achievement adjusted for building closure and further learning loss by end-of year using NWEA-generated factors

Table of Contents

Student-Level Scenarios

- Do Nothing
  - Many states have chosen this approach
  - Best guess of 2020 scores is last year's grade-subject average score

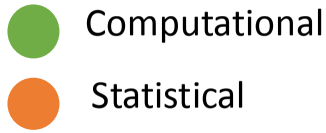
- Copy Past Year Score
  - Example: 4th grader gets same standardized score as received in 3rd
  - Needs a work-around for estimating 3rd grade scores in 2020
  - Where 9th grade is not tested, need a work-around for scores for 10th grade or higher in 2020

- Bridging
  - Take prior and future scores and impute current year value
  - Needs a work-around for estimating 3rd graders in 2020
  - Can test accuracy of approach with earlier data, but the 2020 imputation requires 2021 scores
  - Political issues with 2021 assessments increase the risk that this approach is infeasible

Table of Contents

Simulation Options - *Continued*

- Ordinary Least Squares Regression
  - Uses differential calculus on a set of variables to find the best fit to predict a known outcome.
  - The technique is geared to minimize the size of the error.
  - CREDO created estimates for the 2016-17 scores based on prior tests scores and demographics.
  - The model's results were then extrapolated to produce the estimates for 2017-18.
  - The same approach would be used to build 2020 proxies.
    - Regression #1: used one prior test score (3rd graders excluded in 2020 proxies)
    - Regression #2: used two prior test scores (3rd and 4th graders excluded in 2020 proxies)
      - "Regression #2" is not feasible for simulating 2017-18 for Tennessee because 2014-15, 2015-16 (imputed) and 2016-17 scores form a circular relationship in the model estimating 2016-17. However, we are able to apply "Regression #2" to the estimation of 2020 proxies for Tennessee.



Alternate Aggregated Scenarios

- Historical Average of Scores by School
  - Need for Example: Give all 2020 3rd Graders the average score for 3rd graders in 2018 and 2019.
  - Needed where students have no historical record or where gaps exist in student data (e.g., high school)
  - Examined 1-year, 2-year and 3-year averages. Found 2-year average was best.
  - Tests for accuracy are done at the grade level.

NOTE: Full technical descriptions of all the simulations that were tested is available in Technical Appendix 1 and Technical Appendix 3.

Table of Contents

SIMULATIONS OF 2017-2018 ACHIEVEMENT SCORES

Levels of Diagnostic Analysis

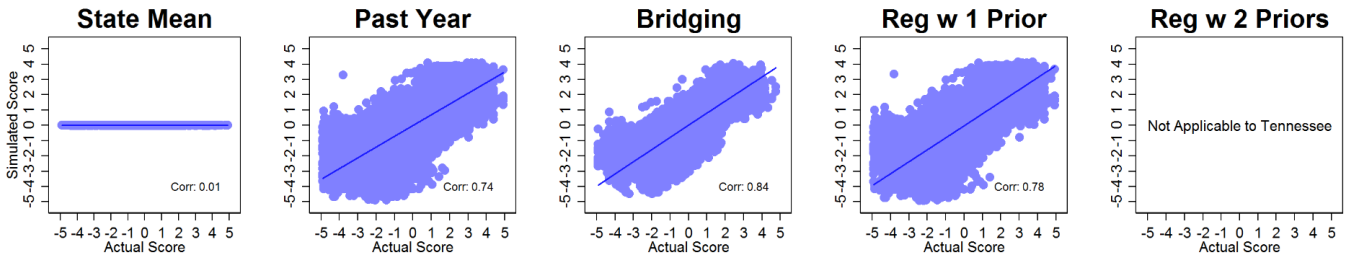
Simulation Option	Student-level Diagnostics	Grade within School Diagnostics	School-level Diagnostics
Do Nothing	X		
Copy Past Year's Score	X		
Bridging	X		
Historical Average Grade within School		X	X
Ordinary Least Squares	X	X	X

NOTE: Tests of accuracy at the grade-within-school or school level produce smaller errors than the same test at the student level, because the aggregated tests are based on averages. All the individual-level variation has been eliminated. This will produce tighter alignment between predicted and average, but also loses the precision of individual predictions.

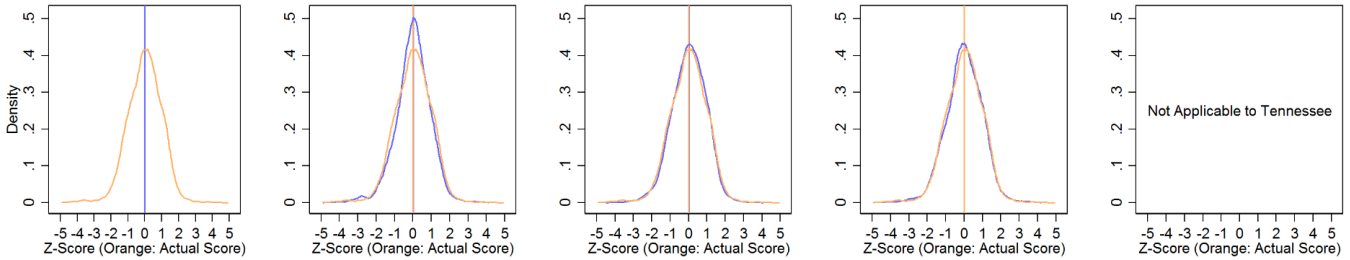
Table of Contents

Tennessee – Reading (1)
Diagnostics by Simulation Scenarios for 2017-18

Correlation:
Actual vs. Predicted Score



Distribution of
Actual and Predicted Score



State Mean Absolute Error

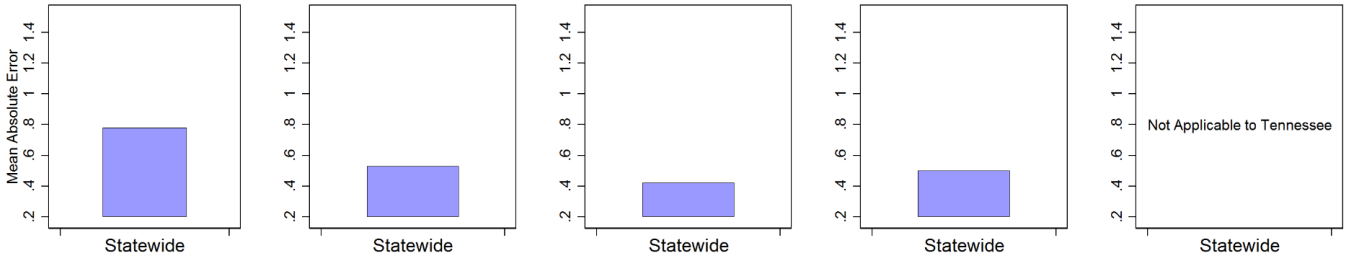


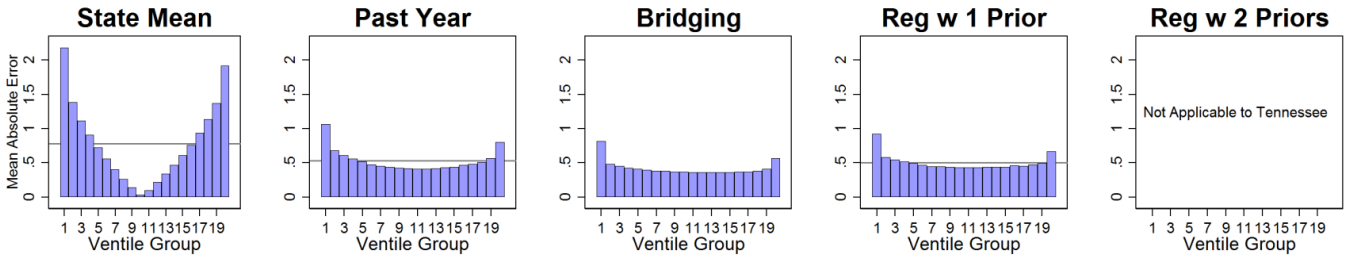
Table of Contents

Note: * denotes no statistical significance at 5% confidence level.

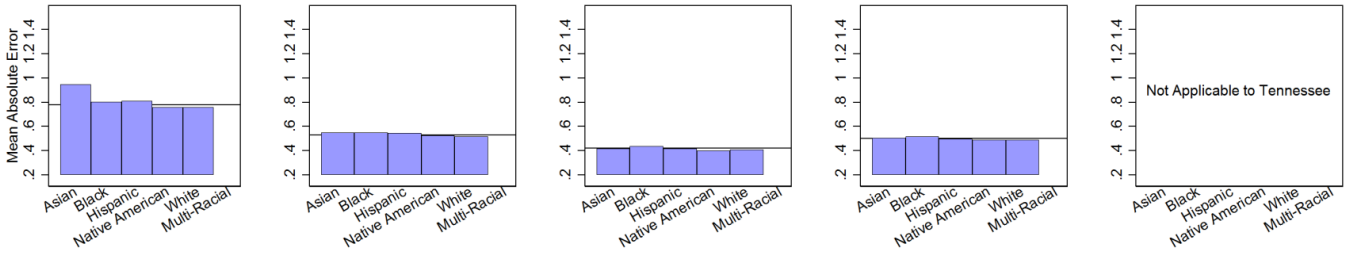
Tennessee – Reading (2)
Diagnostics by Simulation Scenarios for 2017-18

Simulation Results - *Continued*

Mean Absolute Error by Achievement Level



Mean Absolute Error by Student Subgroup - Race and Ethnicity



Mean Absolute Error by Student Subgroup - Poverty, ELL, and Special Ed

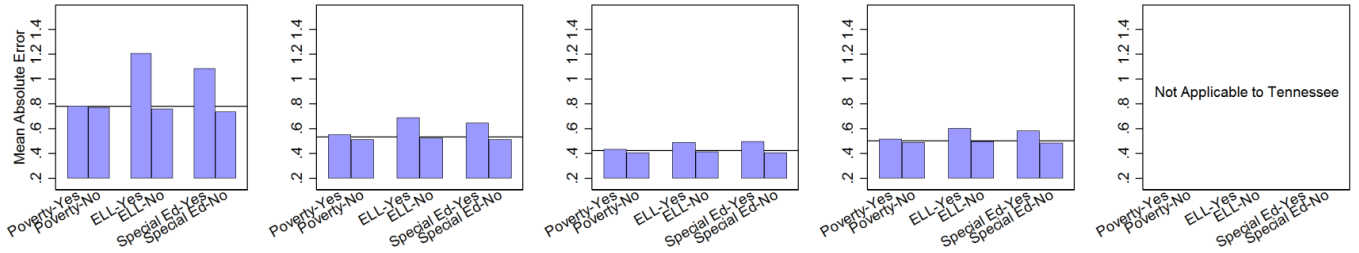


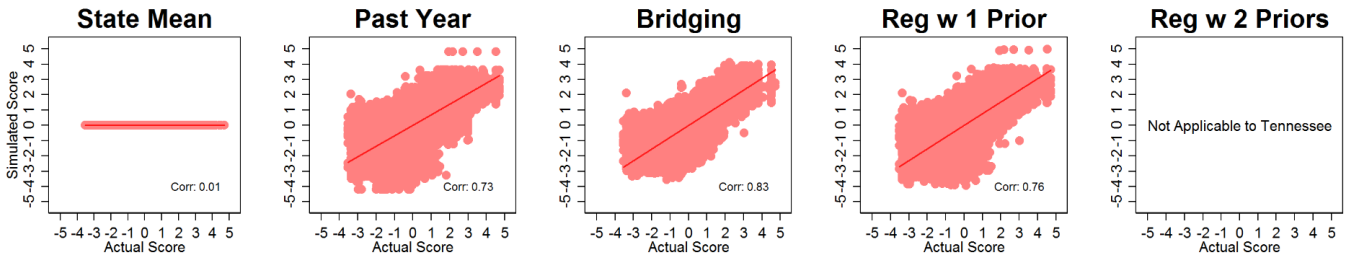
Table of Contents

Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide average absolute error.

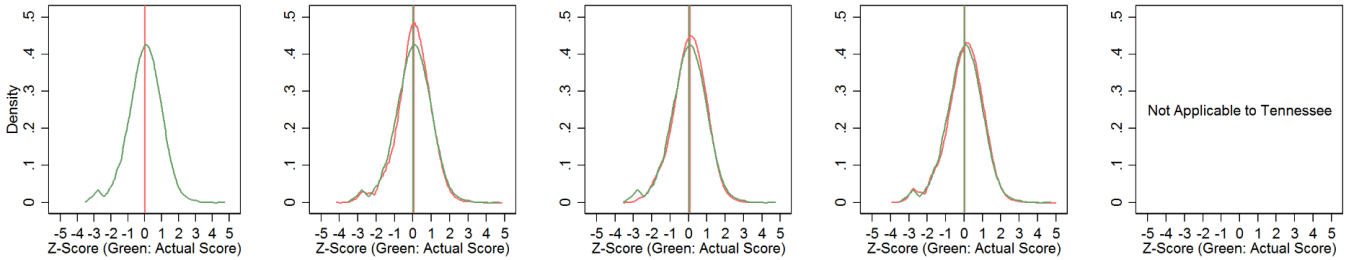
Tennessee – Math (1)
Diagnostics by Simulation Scenarios for 2017-18

Simulation Results - *Continued*

Correlation:
Actual vs. Predicted Score



Distribution of
Actual and Predicted Score



State Mean Absolute Error

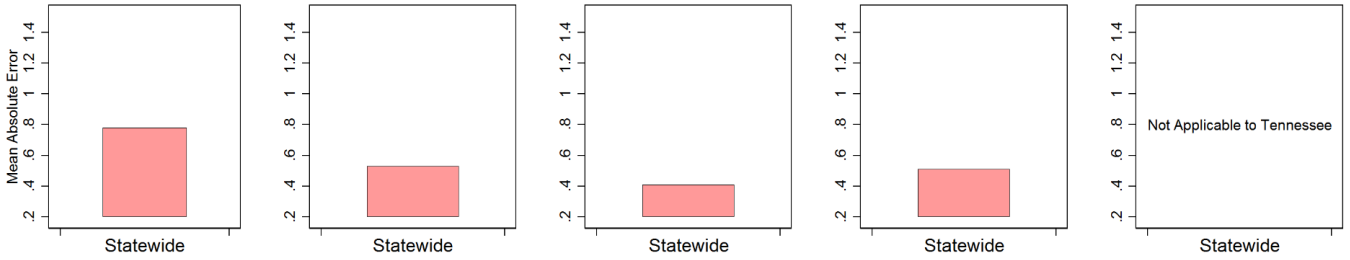


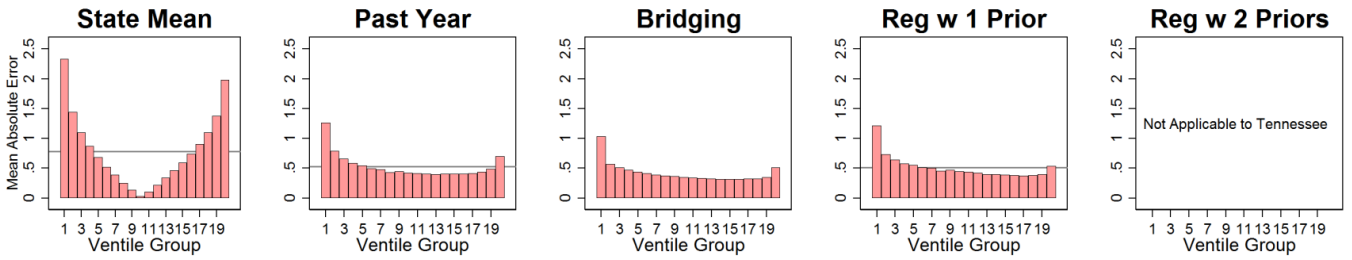
Table of Contents

Note: * denotes no statistical significance at 5% confidence level.

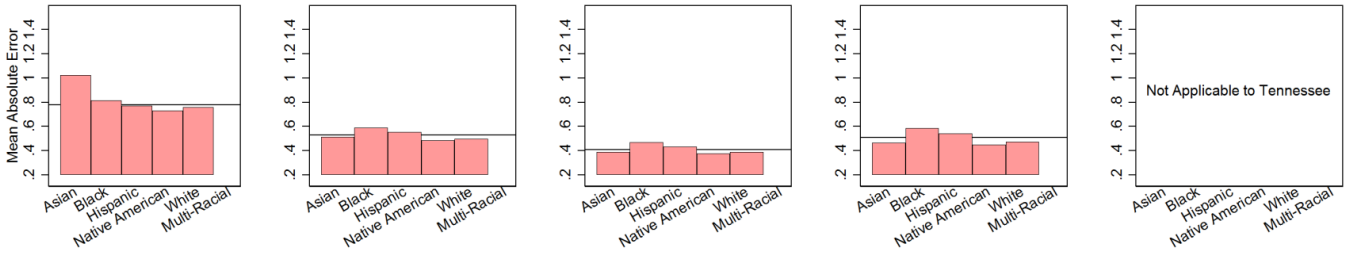
Tennessee – Math (2)
Diagnostics by Simulation Scenarios for 2017-18

Simulation Results - *Continued*

Mean Absolute Error by Achievement Level



Mean Absolute Error by Student Subgroup - Race and Ethnicity



Mean Absolute Error by Student Subgroup - Poverty, ELL, and Special Ed

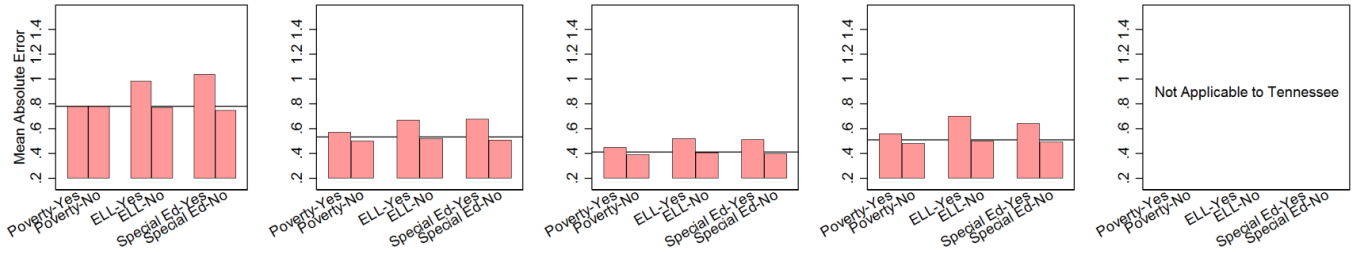


Table of Contents

Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide average absolute error.

CREATING 2020 PROXY ACHIEVEMENT SCORES

Principles for Building 2020 Proxy Scores

1. Student-level proxies are the first preference.
2. Best simulation is the one with smallest prediction error, even if it covers smaller numbers of students.
   - overall
   - by student groups
   - by school
3. "Closely related" models with similar approach and errors can be used to fill in observations if needed.
4. Will move to grade-level estimates only when no satisfactory student-level approach is available.

Table of Contents

Building Full-Year 2020 Proxies - *Continued*

- Best approach is Bridging, but it isn't available now and may never be.

- Second-best is Regression with Two Prior Scores.
  - Average Mean Error is very close to Bridging.
  - Errors by student demographics are evenly distributed for most groups.
  - Errors by school characteristics are evenly distributed.
  - Trouble spots:
    - High school scores have large errors, due to ESSA testing patterns, course pathways and end-of-course test requirements.
    - No 3rd or 4th grade proxies possible from this approach.

- Work-arounds give us "partial" solutions.
  - Wanted to improve predictions for some student subgroups.
  - Used 4th grade proxies from Regression with One Prior Score (3rd best results overall)
  - Gave all 3rd graders the average score for their school from prior two years
  - Gave high school tests the average of past two years' achievement scores.

Table of Contents

- The CREDO team created two different sets of adjustments to the Full-Year Proxies.

- The first adjustment, School Building Closure, measures the achievement of individual students at the point of school building closures as required by the Governor's Executive Order.
  - We estimated the number of classroom-based days of schooling that students did not experience.
  - Using CREDO's approach to estimating Days of Learning, we calculated the learning impact of lost classroom time.
  - This value was subtracted from each student's Full-Year Spring20 proxy.
  - The resulting estimate of achievement is the minimum bound of what students learned in 2019-2020. If no further learning occurred, the estimate represents the student's total learning for the year.

- The second adjustment, called Learning Slide, estimates the amount of learning that faded between school building closures and the end of the regular school year.
  - NWEA collaborated with CREDO to create Learning Loss factors (described on next page) at the grade by school level.
  - Using a conservative approach, the factors were applied to the grade average School Building Closure proxies.
  - The final proxy scores show the upper-bound learning loss for the 2019-2020 year.

Table of Contents

- NWEA created grade-within-school estimates of COVID-related learning loss students experienced. Their models included aggregated student characteristics supplied by CREDO. For further details, please refer to Technical Appendix 2.

- If more than 10 percent of schools in a state took MAP assessments, NWEA created state-specific estimates. The set of MAP-taking schools is not assured to be representative of the full student population in the state.

- In all other states, NWEA built a national set of estimates using a sample of MAP test takers.

- CREDO rejected NWEA estimates for two states, based on comparisons of the demography of the test takers versus the K-12 population in the state, and swapped in the national set.

- In Tennessee, the estimation of learning loss used Tennessee-specific estimates to produce proxies of achievement adjusted for learning loss from school closures.

Table of Contents

Caution is advised in the use of Learning-Loss-Adjusted Achievement proxies.

- Some student groups may be under-represented in the NWEA database.
- The adjustment factors are tailored to 504 different combinations of school demographics, grades and subjects. Some "rare" combinations may apply to small numbers of schools. The estimates that arise from these small samples may be more vulnerable to measurement error than more common combinations.
- The high schools that use MAP for interim assessment generally have substantial shares of remedial students; the factors that are derived may overestimate the degree of learning loss for high schools that have higher levels of achievement.

The best approach to the adjusted achievement proxies is to use them for relative comparisons within and across schools, not as absolute point-estimates of student readiness for 2020-2021.

Table of Contents

- The following graphs display the COVID-related learning loss for students in the 2019-2020 year by the component parts of loss.
- The graphs show each type of loss at the grade-by-school level, as described earlier.
- We use separate colors to distinguish the types of learning loss, as follows:

Source of Loss	Reading	Math
School Building Closure	Color is	Color is
Learning Slide	Color is	Color is

Table of Contents

Tennessee – Reading (1)
Estimated 2019-20 Learning Loss by Achievement Level

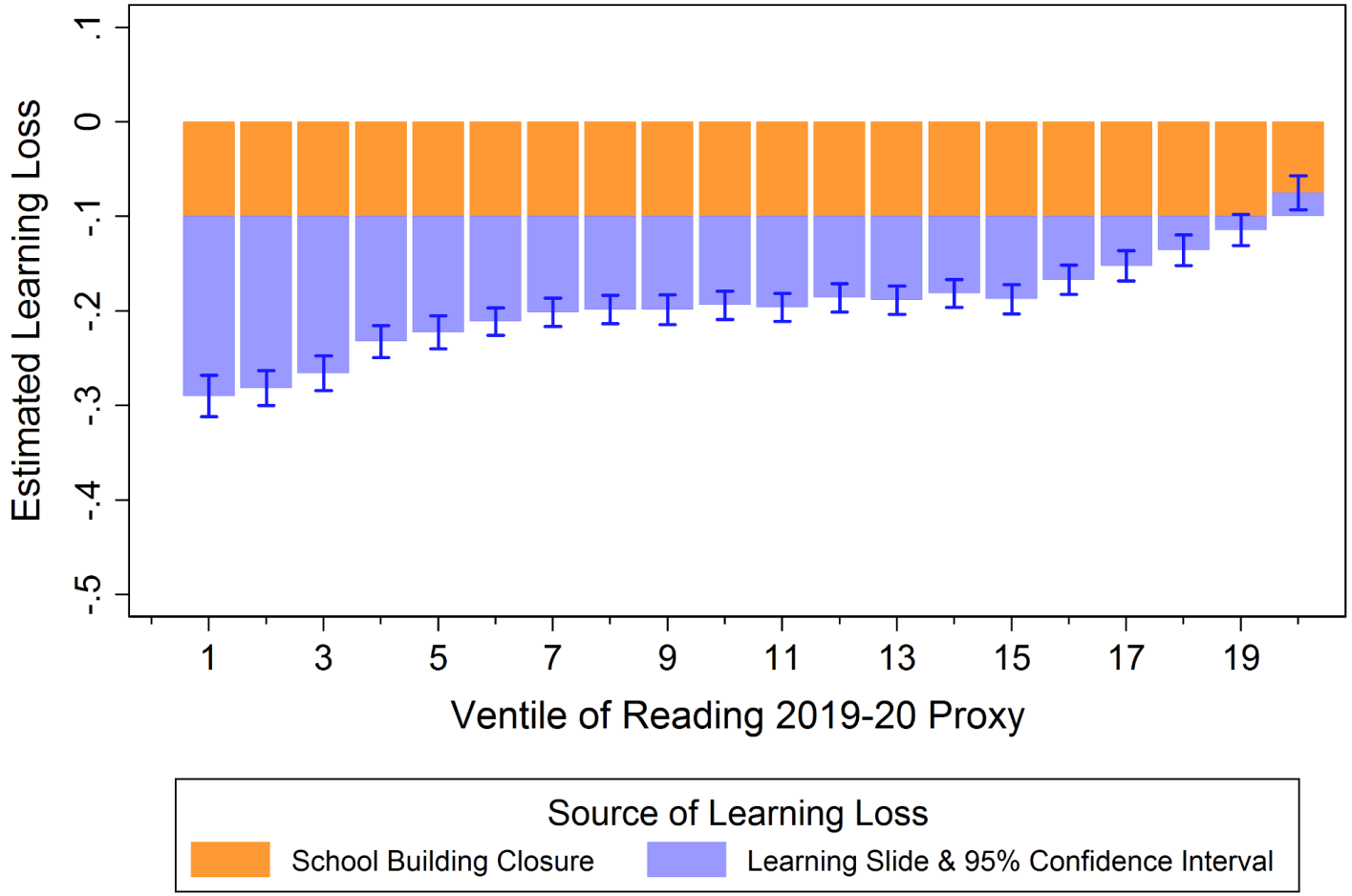


Table of Contents

Note: Learning loss is calculated at the school-by-grade level.

Tennessee – Reading (2)
Estimated 2019-20 Learning Loss by Achievement Level and Grade/EOC

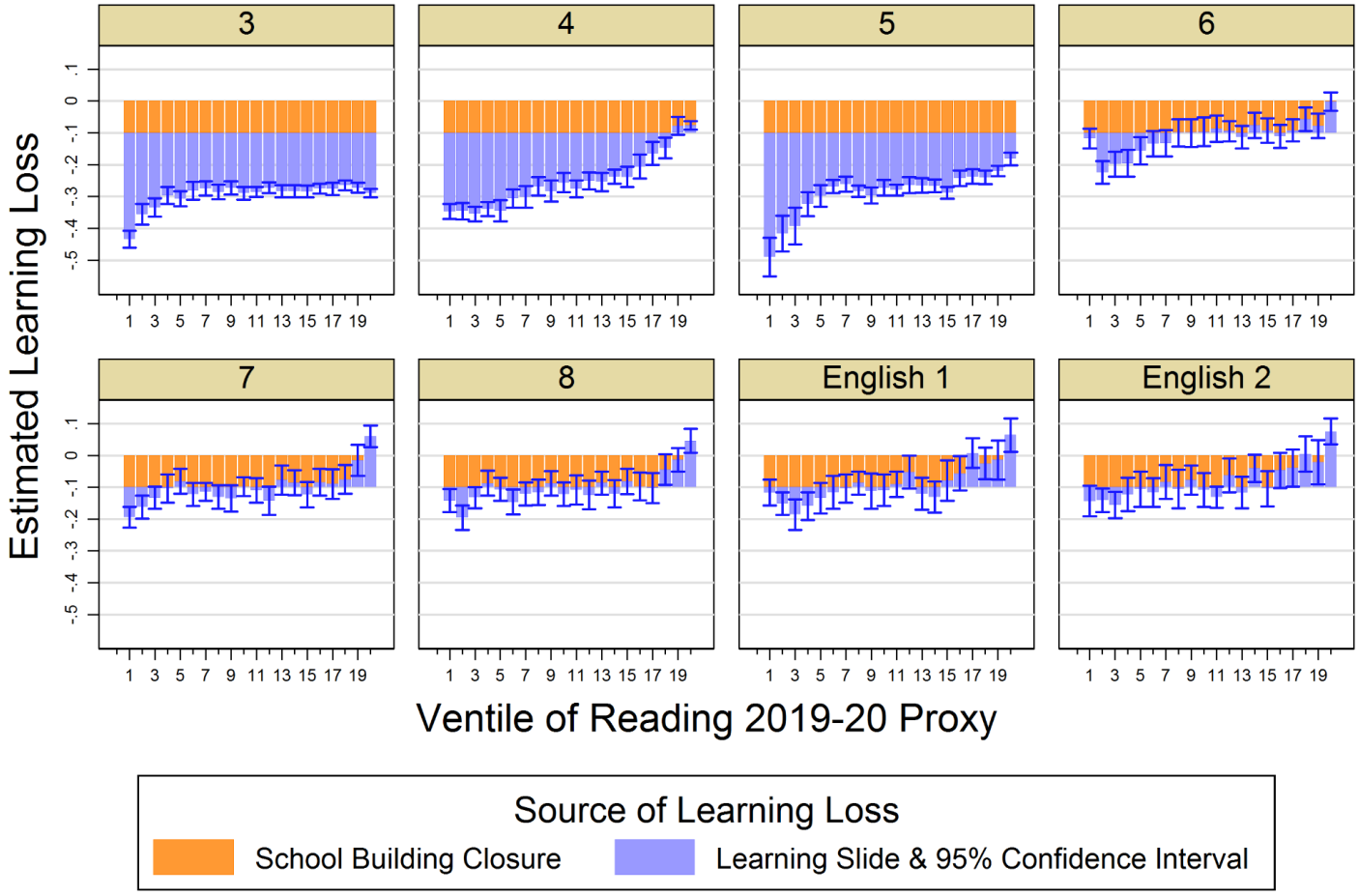


Table of Contents

Note: Learning loss is calculated at the school-by-grade level.

Tennessee – Math (1)
Estimated 2019-20 Learning Loss by Achievement Level

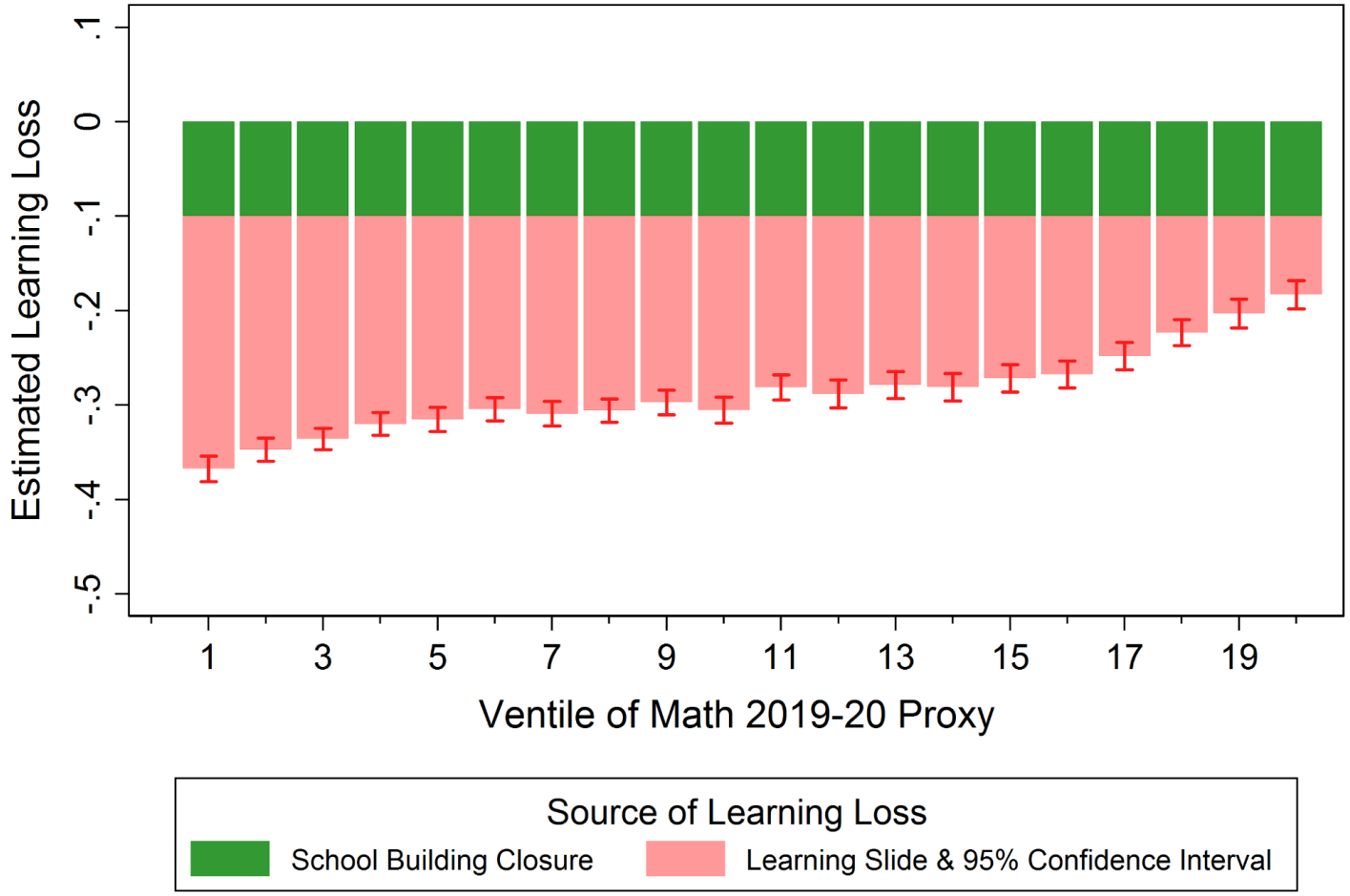


Table of Contents

Note: Learning loss is calculated at the school-by-grade level.

Tennessee – Math (2)
Estimated 2019-20 Learning Loss by Achievement Level and Grade/EOC

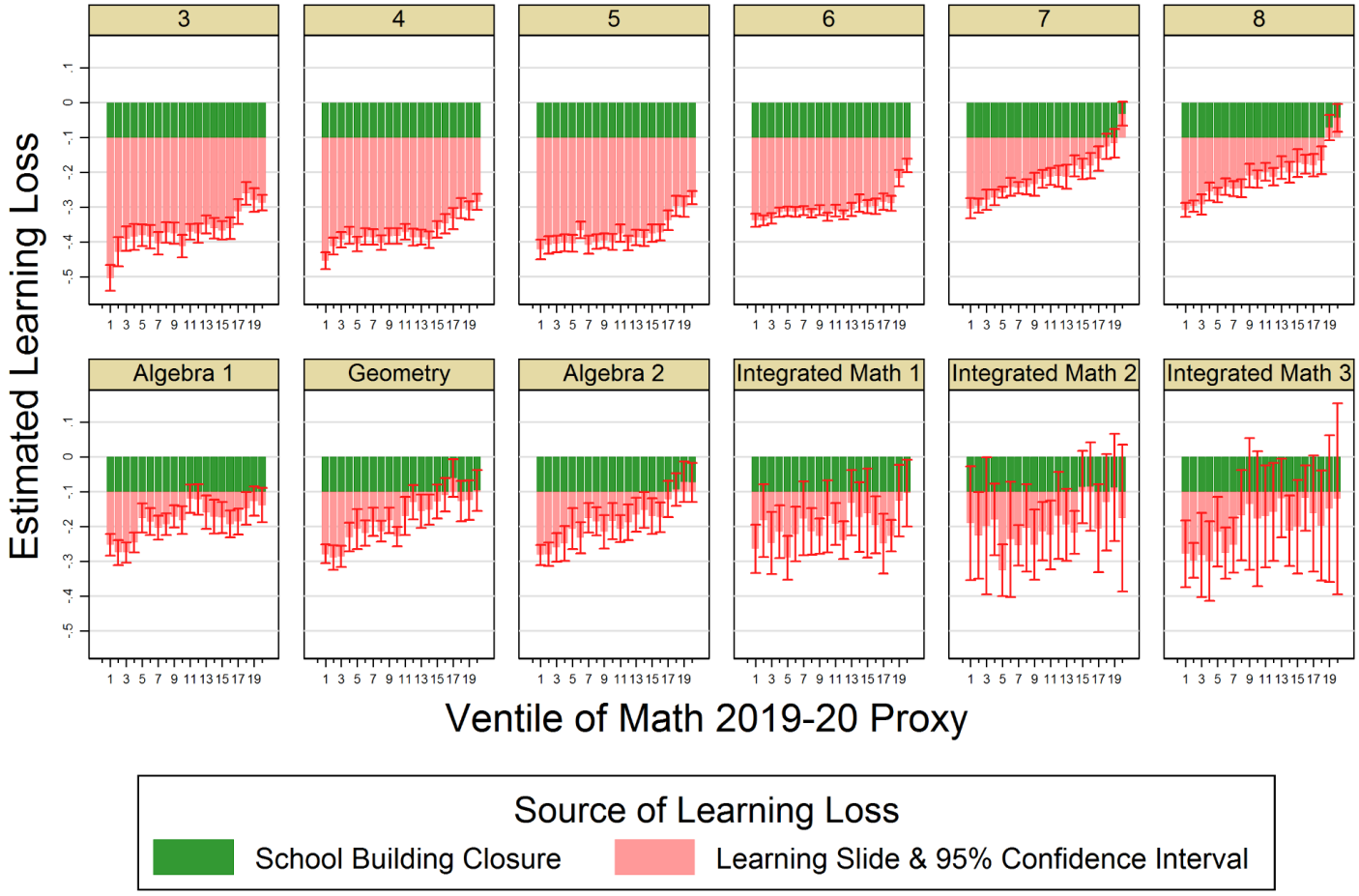


Table of Contents

Note: Learning loss is calculated at the school-by-grade level.

What's Next?

- SEA Opportunities
  - Review results of COVID-Sim Project
  - Consider suitability for various use cases (based on Slide 6)
  - Consider value of adjusted proxies for SEA / LEA plans for 2020-2021

- CREDO Opportunities
  - Deliver briefing on COVID-Sim Project and results
  - Support review and application of proxies, as requested
  - Build a summary memo for public release to share lessons learned

Table of Contents

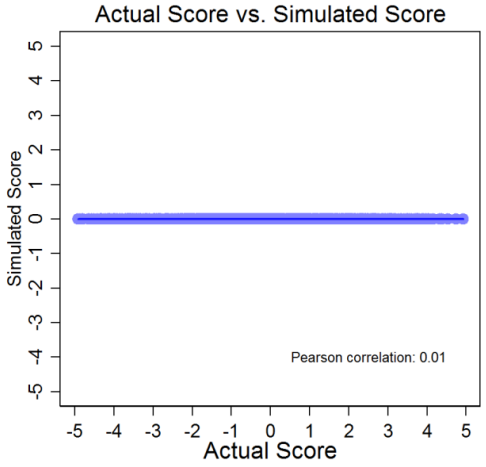
THANK YOU!!

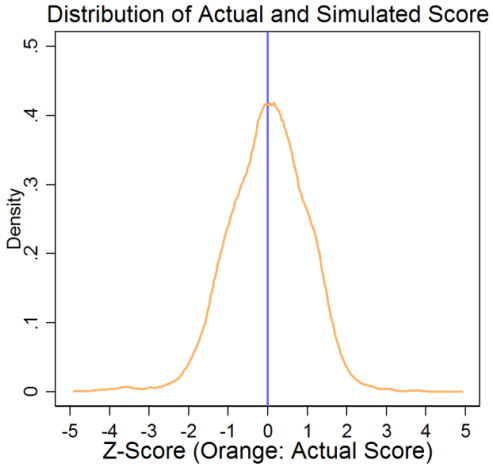
APPENDICES

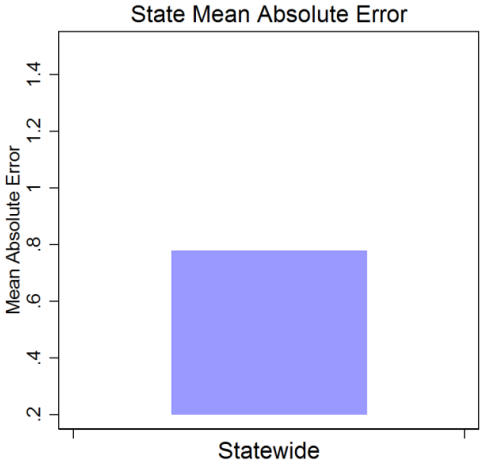
Appendix 1

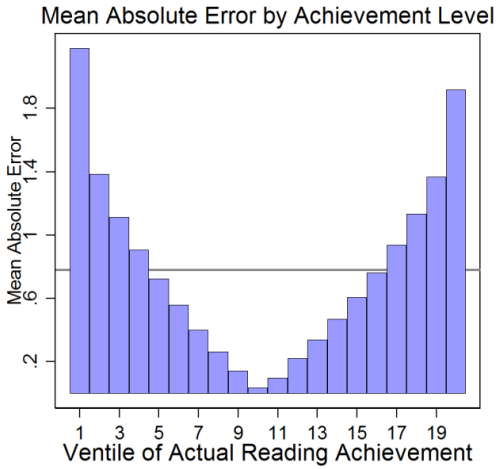
Full-Sized Graphs of 2017-18 Simulations

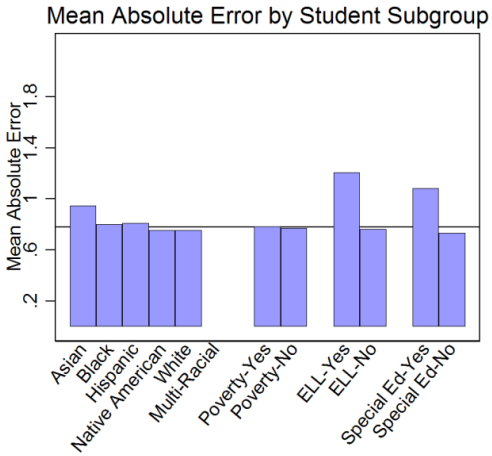
[Tennessee - Reading] Simulation Scenario: State Mean











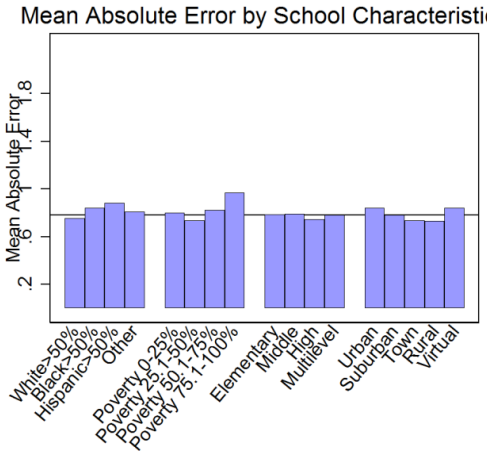
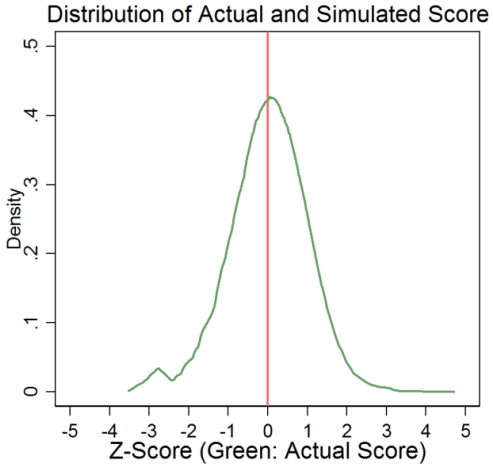
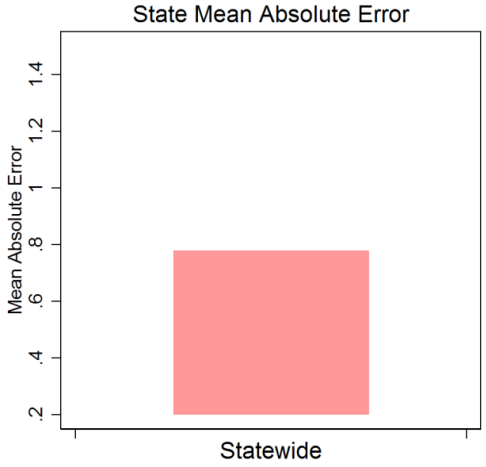


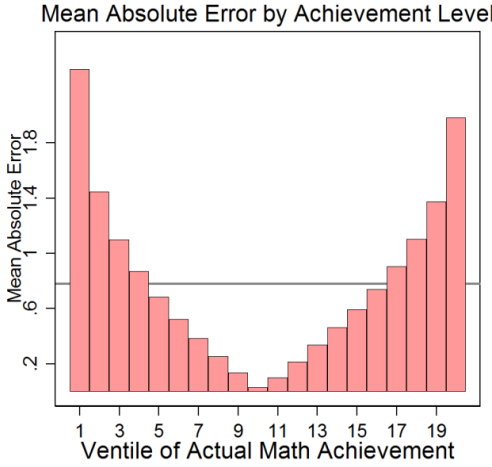
Table of Contents

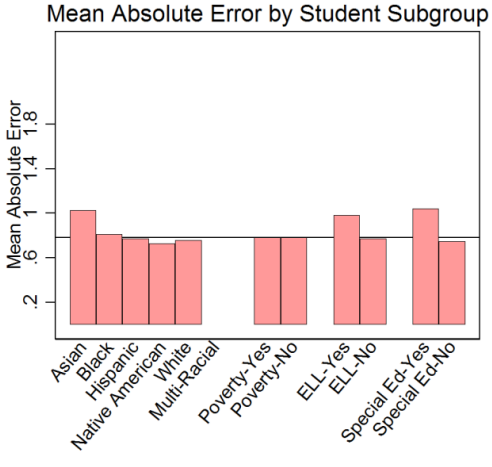
Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide mean absolute error.

[Tennessee - Math] Simulation Scenario: State Mean









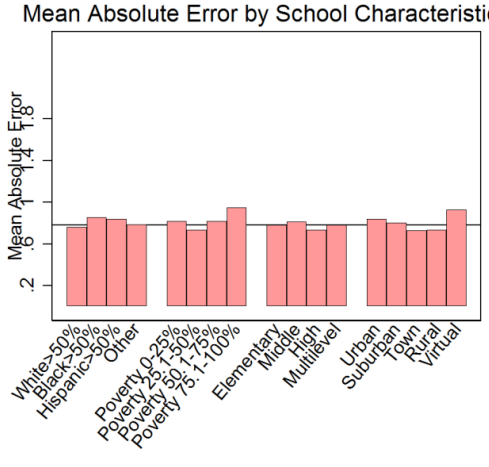
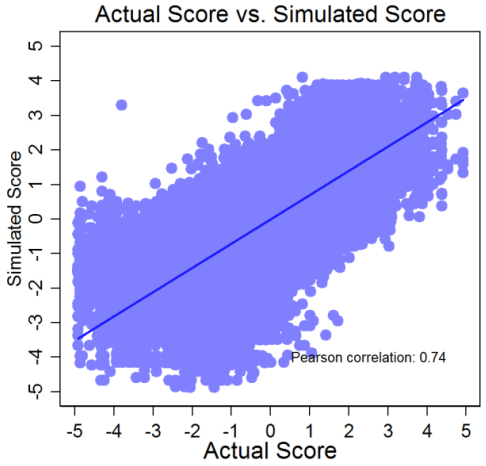
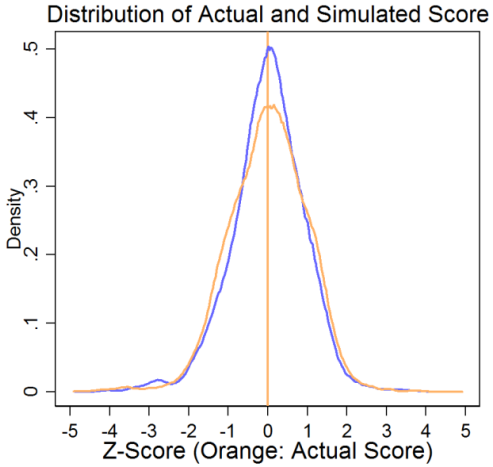


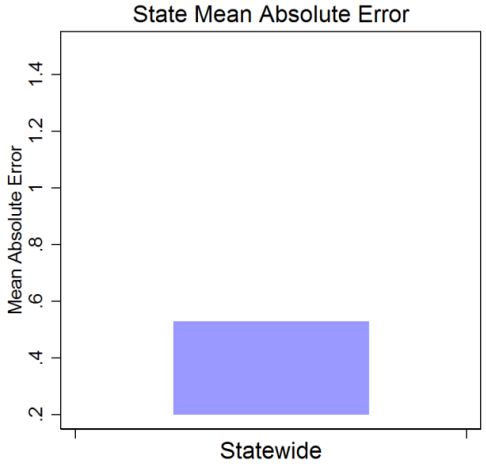
Table of Contents

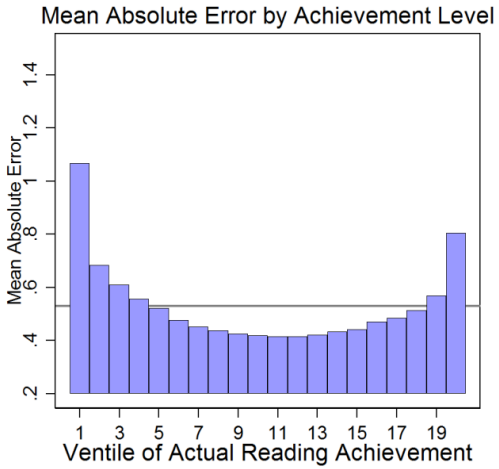
Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide mean absolute error.

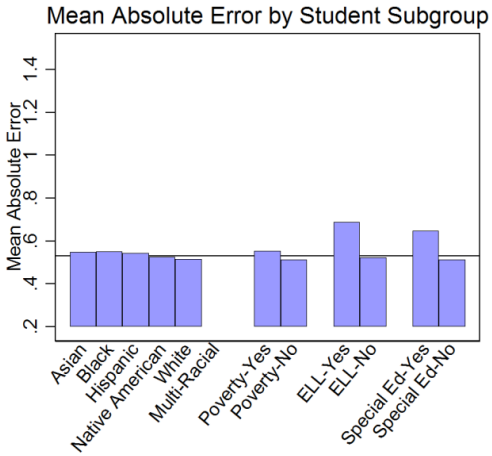
[Tennessee - Reading] Simulation Scenario: Past Year's Score











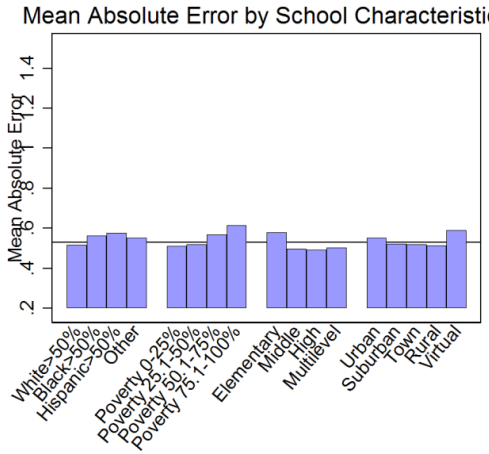
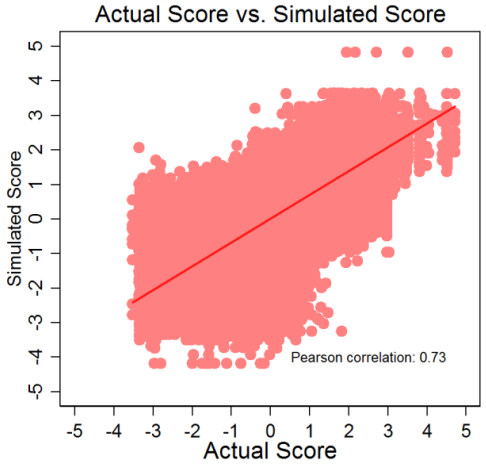
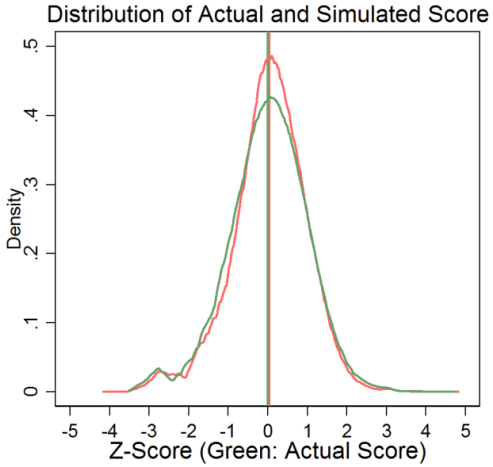


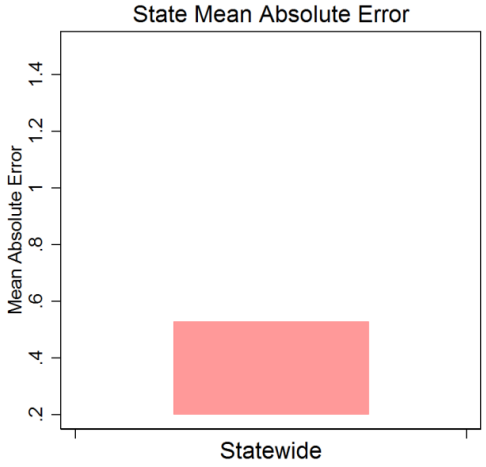
Table of Contents

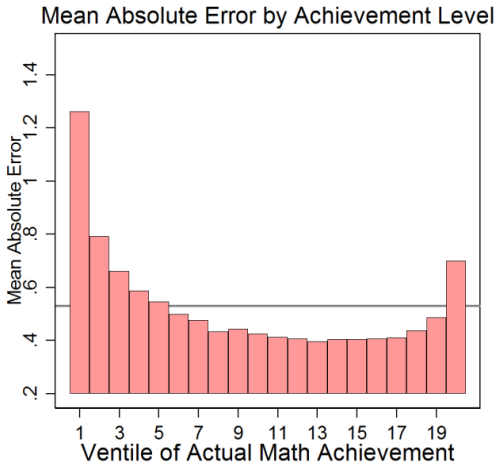
Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide mean absolute error.

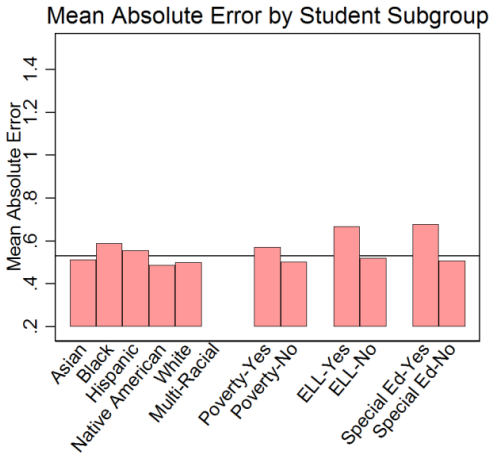
[Tennessee - Math] Simulation Scenario: Past Year's Score











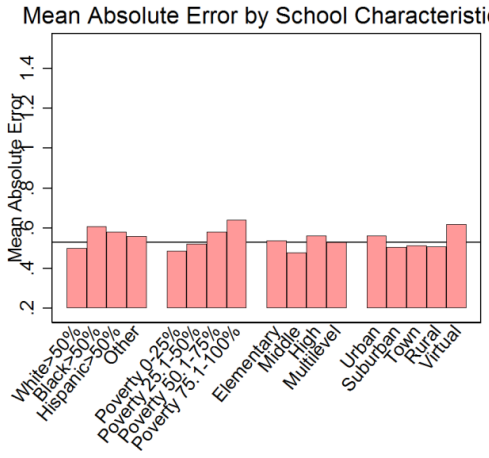
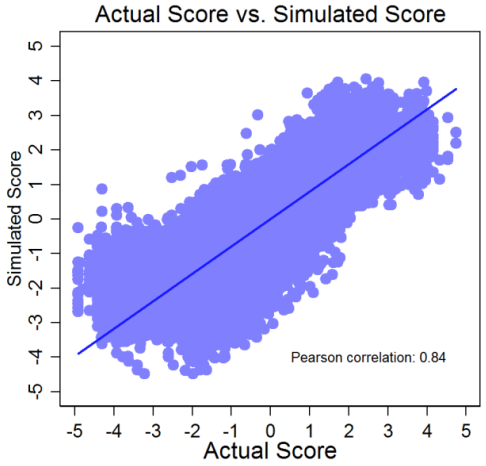
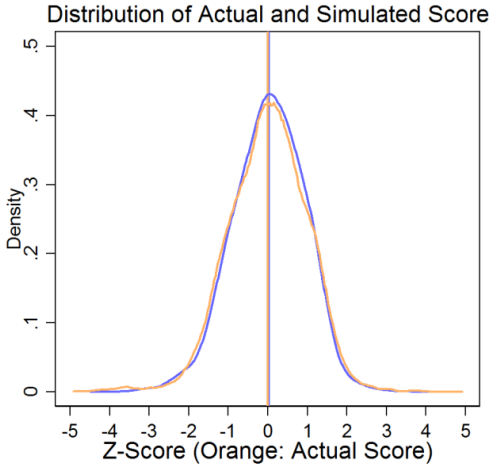


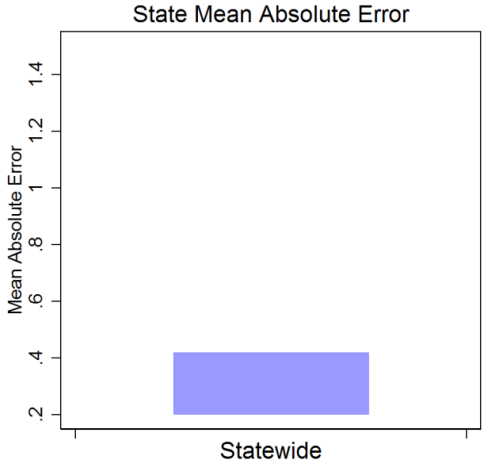
Table of Contents

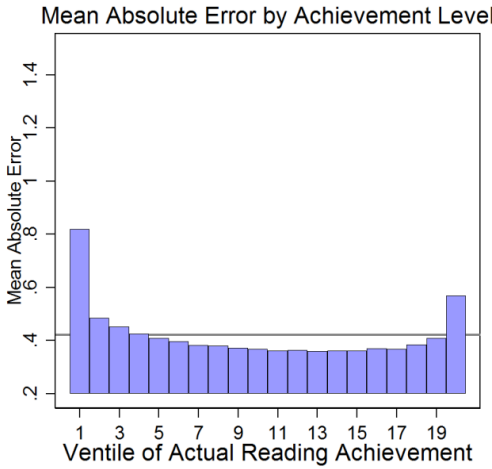
Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide mean absolute error.

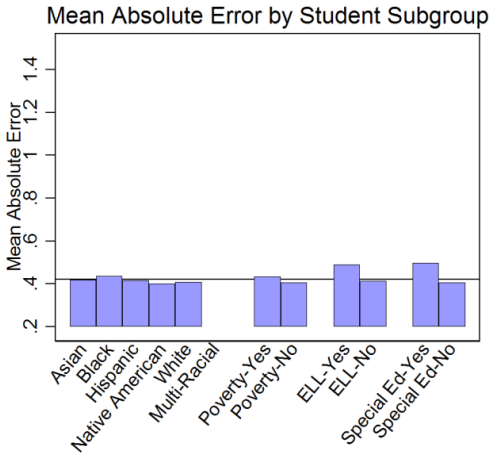
[Tennessee - Reading] Simulation Scenario: Bridging











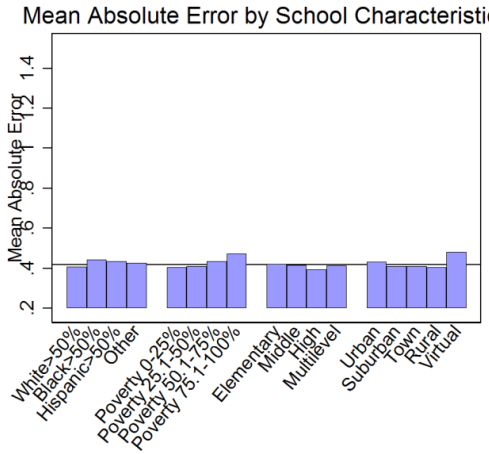
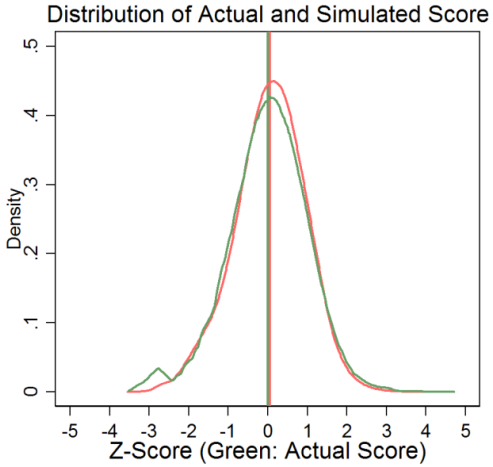
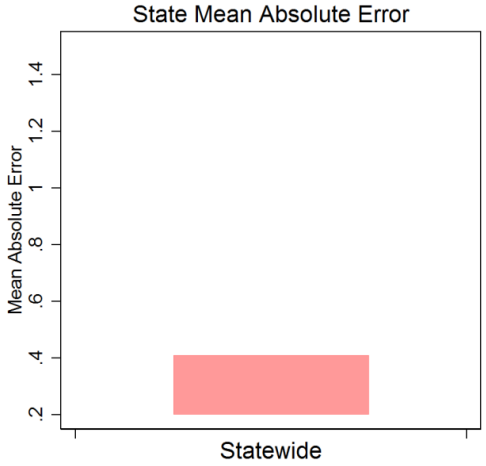


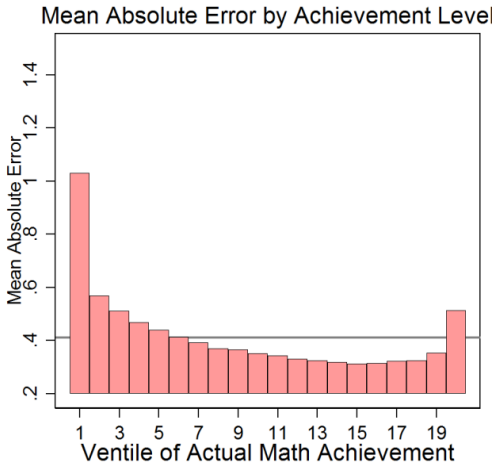
Table of Contents

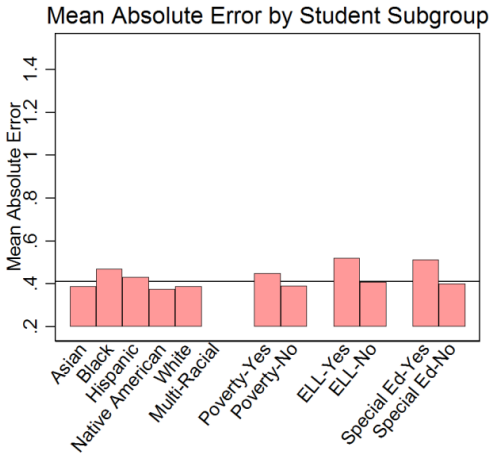
Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide mean absolute error.

[Tennessee - Math] Simulation Scenario: Bridging









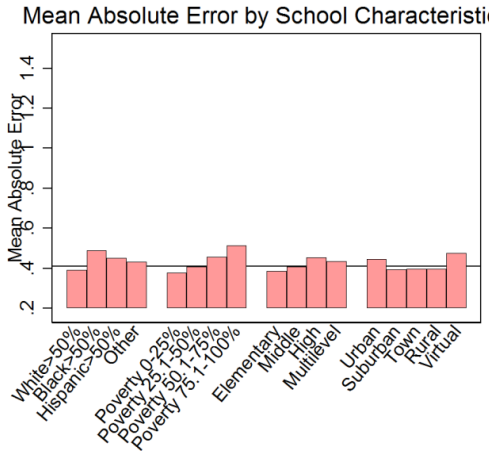
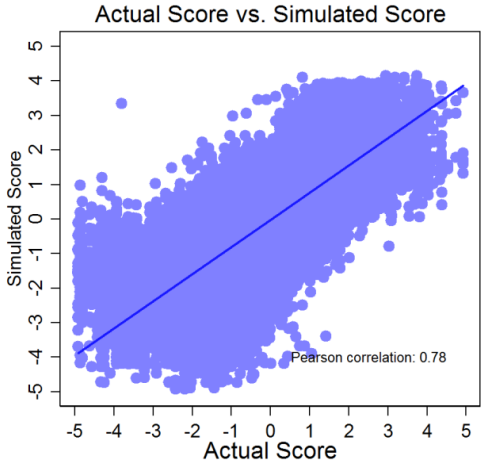
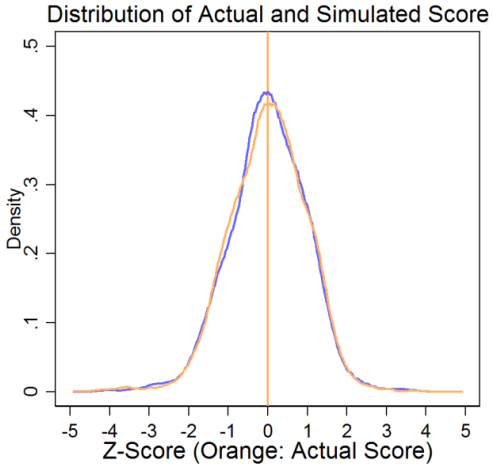


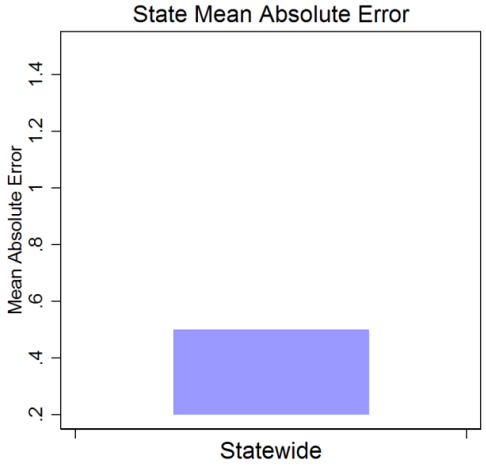
Table of Contents

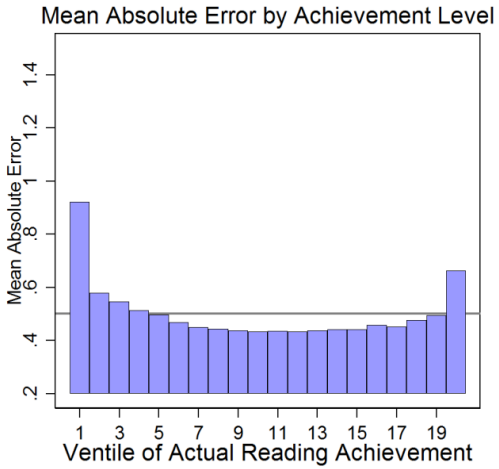
Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide mean absolute error.

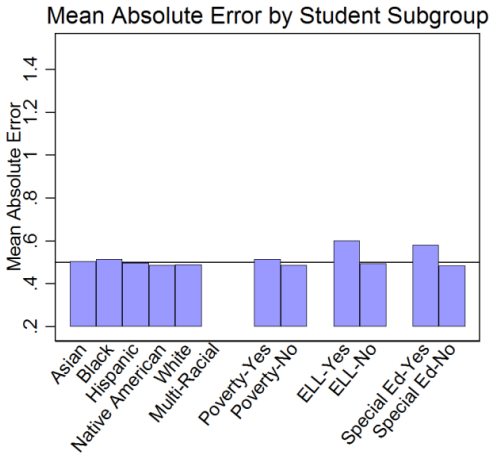
[Tennessee - Reading] Simulation Scenario: Regression with 1 Prior











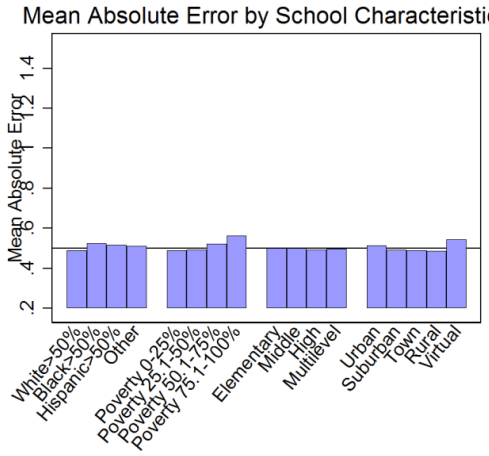
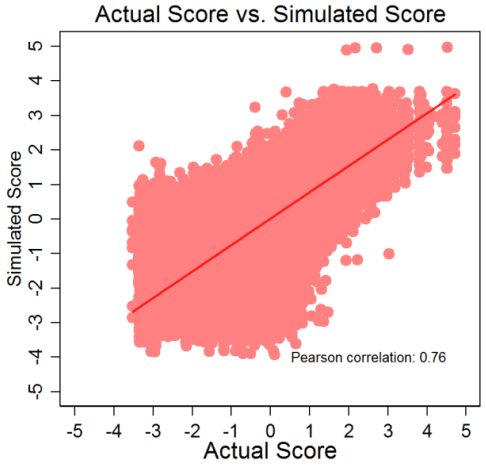
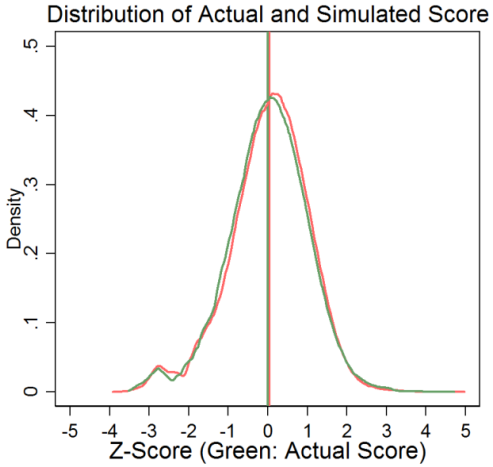


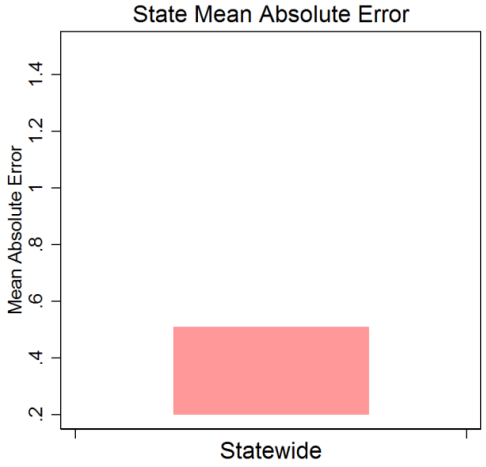
Table of Contents

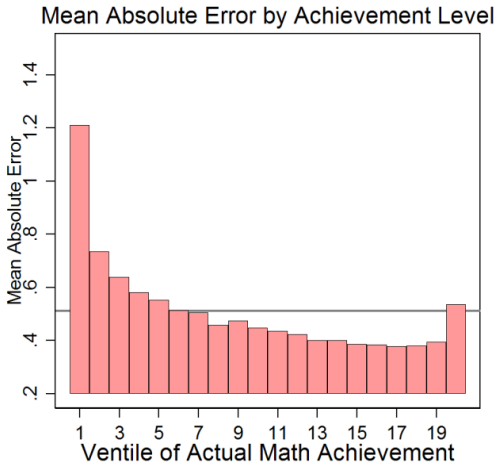
Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide mean absolute error.

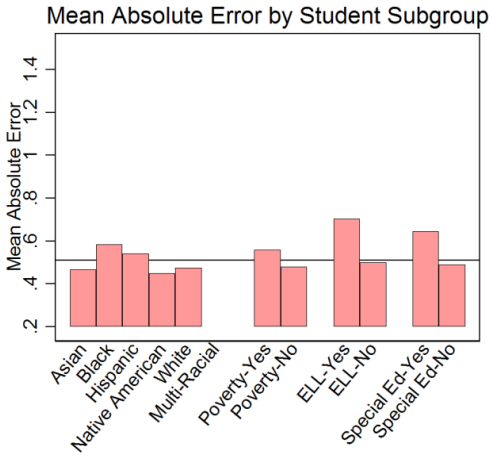
[Tennessee - Math] Simulation Scenario: Regression with 1 Prior











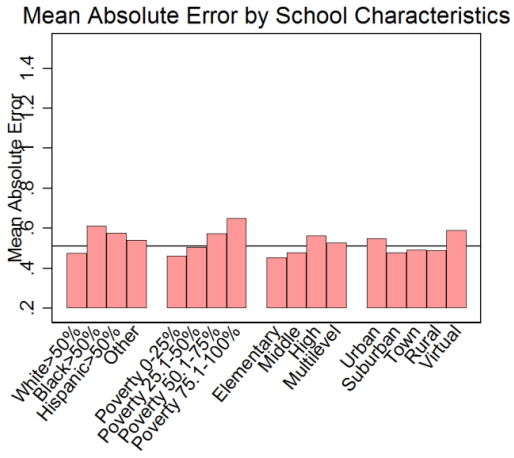


Table of Contents

Notes: (1) * denotes no statistical significance at 5% confidence level. (2) Horizonal gray line is statewide mean absolute error.

[Tennessee – Reading] Simulation Scenario: Regression with 2 Priors

We took 2014-15 and 2016-17 scores to impute 2015-16 scores because the Tennessee test score data for 2015-16 were incomplete. Hence, the "Regression with 2 Priors" approach is not feasible for simulating 2017-18 for Tennessee because 2014-15, 2015-16 and 2016-17 scores form a circular relationship in the model estimating 2016-17 scores. However, we are able to apply the "Regression with 2 Priors" approach in the proxy estimation for 2019-20 for Tennessee.

Table of Contents

[Tennessee – Math] Simulation Scenario: Regression with 2 Priors

We took 2014-15 and 2016-17 scores to impute 2015-16 scores because the Tennessee test score data for 2015-16 were incomplete. Hence, the "Regression with 2 Priors" approach is not feasible for simulating 2017-18 for Tennessee because 2014-15, 2015-16 and 2016-17 scores form a circular relationship in the model estimating 2016-17 scores. However, we are able to apply the "Regression with 2 Priors" approach in the proxy estimation for 2019-20 for Tennessee.

Table of Contents

Appendix 2

Table of 2017-18 Simulation Results

Tennessee: Simulation Diagnostics for 2017-18 Achievement in Reading

Simulation Approach		State Mean	Past Year's Score	Bridging	Regression with One Prior	Regression with Two Priors
Count of Student Records Simulated		637,219	603,537	391,311	455,035	n.a.
Percentage of Student Records Simulated		100.0%	94.7%	61.4%	71.4%	n.a.
Overall Mean Absolute Simulation Error (in SD)		**0.777**	**0.528**	**0.415**	**0.497**	**n.a.**
Mean Absolute Simulation Error by Student Subgroup						
Gender	Males	0.794	0.541	0.426	0.511	n.a.
	Females	0.759	0.514	0.405	0.482	n.a.
Race/Ethnicity	Asian/PI	0.947	0.548	0.418	0.505	n.a.
	Black	0.802	0.549	0.436	0.515	n.a.
	Hispanic	0.810	0.544	0.416	0.498	n.a.
	Native American	0.756	0.524	0.401	0.487	n.a.
	White	0.756	0.516	0.408	0.490	n.a.
	Multi-racial					n.a.
Poverty	No	0.773	0.514	0.405	0.487	n.a.
	Yes	0.784	0.552	0.434	0.515	n.a.
Special Education	No	0.735	0.511	0.405	0.486	n.a.
	Yes	1.084	0.646	0.497	0.582	n.a.
English Language Learners	No	0.761	0.523	0.413	0.495	n.a.
	Yes	1.206	0.689	0.490	0.603	n.a.
Tested Grade in 2017-18	3	0.799	0.730	0.605		n.a.
	4	0.792	0.523	0.427	0.824	n.a.
	5	0.787	0.506	0.423	0.507	n.a.
	6	0.791	0.522	0.429	0.523	n.a.
	7	0.790	0.482	0.414	0.483	n.a.
	8	0.778	0.482	0.399	0.484	n.a.
	English 1	0.739	0.478	0.394	0.479	n.a.
	English 2	0.773	0.484	0.485	0.485	n.a.
	English 3	0.731	0.521	0.521	0.516	n.a.

Back Table of Contents

Tennessee: Simulation Diagnostics for 2017-18 Achievement in Reading – *Continued*

Simulation Approach		State Mean	Past Year's Score	Bridging	Regression with One Prior	Regression with Two Priors
Count of Student Records Simulated		637,219	603,537	391,311	455,035	n.a.
Percentage of Student Records Simulated		100.0%	94.7%	61.4%	71.4%	n.a.
Overall Mean Absolute Simulation Error (in SD)		**0.777**	**0.528**	**0.415**	**0.497**	**n.a.**
Mean Absolute Simulation Error by School Characteristic						
School Locale	Urban - Large	0.864	0.567	0.443	0.526	n.a.
	Urban - Midsize	0.828	0.539	0.423	0.506	n.a.
	Urban - Fringe	0.794	0.520	0.406	0.491	n.a.
	Suburban	0.783	0.521	0.411	0.493	n.a.
	Town	0.735	0.518	0.410	0.490	n.a.
	Rural	0.730	0.514	0.406	0.487	n.a.
	Virtual Campus	0.841	0.591	0.481	0.544	n.a.
School Grade Span	Elementary	0.785	0.580	0.423	0.501	n.a.
	Middle	0.793	0.499	0.416	0.499	n.a.
	High	0.747	0.493	0.395	0.492	n.a.
	Multi-level	0.780	0.503	0.412	0.497	n.a.
School Size	Small	0.773	0.548	0.424	0.506	n.a.
	Medium	0.780	0.534	0.416	0.496	n.a.
	Large	0.773	0.503	0.409	0.495	n.a.
School Sector	District	0.776	0.527	0.415	0.496	n.a.
	Charter	0.806	0.540	0.429	0.517	n.a.

Table of Contents

Tennessee: Simulation Diagnostics for 2017-18 Achievement in Math

Simulation Approach		State Mean	Past Year's Score	Bridging	Regression with One Prior	Regression with Two Priors
Count of Student Records Simulated		636,641	599,745	434,317	449,730	n.a.
Percentage of Student Records Simulated		100.0%	94.2%	68.2%	70.6%	n.a.
Overall Mean Absolute Simulation Error (in SD)		**0.780**	**0.526**	**0.411**	**0.506**	**n.a.**
Mean Absolute Simulation Error by Student Subgroup						
Gender	Males	0.811	0.543	0.423	0.523	n.a.
	Females	0.748	0.509	0.400	0.489	n.a.
Race/Ethnicity	Asian/PI	1.024	0.512	0.389	0.466	n.a.
	Black	0.814	0.588	0.469	0.584	n.a.
	Hispanic	0.771	0.555	0.430	0.541	n.a.
	Native American	0.730	0.486	0.374	0.449	n.a.
	White	0.760	0.499	0.388	0.474	n.a.
	Multi-racial					n.a.
Poverty	No	0.780	0.502	0.391	0.479	n.a.
	Yes	0.781	0.570	0.449	0.558	n.a.
Special Education	No	0.746	0.508	0.400	0.491	n.a.
	Yes	1.039	0.677	0.512	0.644	n.a.
English Language Learners	No	0.772	0.521	0.408	0.500	n.a.
	Yes	0.982	0.668	0.519	0.702	n.a.
Tested Grade/EOC in 2017-2018	3	0.791	0.719	0.604		n.a.
	4	0.788	0.455	0.380	0.786	n.a.
	5	0.789	0.455	0.391	0.457	n.a.
	6	0.780	0.471	0.378	0.472	n.a.
	7	0.805	0.468	0.408	0.469	n.a.
	8	0.801	0.511	0.455	0.501	n.a.
	Algebra 1	0.772	0.499	0.414	0.508	n.a.
	Geometry	0.733	0.548	0.456	0.544	n.a.
	Algebra 2	0.734	0.602	0.544	0.600	n.a.
	Integrated Math 1	0.795	0.513	0.422	0.518	n.a.
	Integrated Math 2	0.767	0.583	0.494	0.583	n.a.
	Integrated Math 3	0.814	0.650	0.681	0.646	n.a.

Table of Contents

Tennessee: Simulation Diagnostics for 2017-18 Achievement in Math – *Continued*

Simulation Approach		State Mean	Past Year's Score	Bridging	Regression with One Prior	Regression with Two Priors
Count of Student Records Simulated		636,641	599,745	434,317	449,730	n.a.
Percentage of Student Records Simulated		100.0%	94.2%	68.2%	70.6%	n.a.
Overall Mean Absolute Simulation Error (in SD)		**0.780**	**0.526**	**0.411**	**0.506**	**n.a.**
Mean Absolute Simulation Error by School Characteristic						
School Locale	Urban - Large	0.863	0.599	0.474	0.596	n.a.
	Urban - Midsize	0.800	0.532	0.413	0.508	n.a.
	Urban - Fringe	0.820	0.504	0.400	0.479	n.a.
	Suburban	0.802	0.507	0.393	0.480	n.a.
	Town	0.730	0.513	0.397	0.491	n.a.
	Rural	0.732	0.508	0.398	0.489	n.a.
	Virtual Campus	0.927	0.619	0.475	0.589	n.a.
School Grade Span	Elementary	0.784	0.538	0.386	0.455	n.a.
	Middle	0.815	0.480	0.407	0.479	n.a.
	High	0.733	0.563	0.454	0.562	n.a.
	Multi-level	0.782	0.531	0.435	0.529	n.a.
School Size	Small	0.769	0.535	0.414	0.508	n.a.
	Medium	0.781	0.525	0.408	0.499	n.a.
	Large	0.781	0.524	0.419	0.518	n.a.
School Sector	District	0.778	0.523	0.408	0.502	n.a.
	Charter	0.835	0.590	0.476	0.589	n.a.

Table of Contents