A Meta-Analysis of Simulations of 2020 Achievement Assessments in 19 States

The Center for Research on Education Outcomes

Stanford University

October 2020

Center for Research on Education Outcomes
Stanford University
Stanford, CA

http://credo.stanford.edu

CREDO, the Center for Research on Education Outcomes at Stanford University, was established to improve empirical evidence about education reform and student performance at the primary and secondary levels. CREDO at Stanford University supports education organizations and policymakers in using reliable research and program evaluation to assess the performance of education initiatives. CREDO's valuable insight helps educators and policymakers strengthen their focus on the results from innovative programs, curricula, policies and accountability practices.

Acknowledgements

CREDO gratefully acknowledges the support of the State Education Agencies and School Districts who contributed their data to this partnership. Our data access partnerships form the foundation of CREDO's work, without which studies like this would be impossible.

Contents

Table of Figures

Table of Tables

Appendices

Introduction

Cancellation of state standardized testing in Spring 2020 due to school closures during the coronavirus pandemic was prudent and logistically necessary. Still, loss of data on student learning directly impacts state and local education efforts, ranging from reporting on mandated programs to school accountability to student support service planning and delivery to evaluation of curriculum and instructional programs. While the immediate focus of all educators and policy makers throughout the pandemic has been on restoring instruction and supporting students and teachers, the gaps in student-level data are certain to be eventually recognized.

The Center for Research on Education Outcomes (CREDO) at Stanford University was uniquely positioned to offer a solution. CREDO holds recent student-level data for over 30 states as part of our research consortium. CREDO also has the methodological expertise to simulate proxies for the missing Spring 2020 achievement scores from state mandated assessments. We devised a multi-step process to create estimates of student-level achievement. Ultimately, we provided estimates of student academic learning for 20 states, of which 19 are included in this meta-analysis.[1]

For each state, we developed three sets of proxies. The first consisted of full-year achievement proxies for students – as though COVID had not happened. This series can be used for planning purposes and to serve as a first-order target for academic recovery efforts. In the second set, CREDO estimated the achievement for each state's students at the point of school building closures in March 2020, the last point of classroom-based instruction. Finally, with support from the Northwest Evaluation Association (NWEA), the third set estimated the full effect of COVID-related effects to produce student-level estimates of "achievement after COVID slide" by the end of the 2019-2020 academic year for each state in the project.

> **Achievement measures** are level or status indicators of what a student knows at a particular point in time. Several factors influence a student's individual achievement: the amount of prior schooling (which in turn relates to school quality); individual innate abilities, motivation and other non-cognitive assets; family support for education and the individual pace of maturation. During normal conditions, it is difficult to tease out the interplay of these influences. With the coronavirus pandemic and other disruptions, it is even more challenging.

CREDO ultimately identified three simulations that produced similar achievement estimates with sufficient accuracy that achievement proxies could be computed for 2020 as if COVID had not happened. Additional adjustments for lost days of classroom instruction and other COVID-related impacts produced student-level estimates of learning loss by the end of the 2019-2020 school year. Looking at state averages, the magnitude of losses in Reading ranged from slightly less than -.1 standard deviations, about 57 days of learning, to -.316, which equates to 183 days of schooling. In Math, the average losses ranged from -.235, or about 136 days of learning, to -.402, equal to 232 days of learning.[2]

[1] We treat New York City separately from the rest of New York State.

[2] CREDO regularly converts education impacts into Days of Learning to facilitate comprehension.

This briefing paper shares details on the overall simulation effort including the findings from the COVID – Sim exercise. With the uncertain fate of assessments in the future, these lessons may prove helpful beyond the 2019-2020 school year. At the least, this report may contribute to an anthology of response efforts during the coronavirus pandemic of 2020 in the United States.

This paper is organized in two parts. The first part describes our approach to building the assessment proxies and compares the results of several simulation options. The simulations produced equivalent results across all the participating states, which strengthens our confidence that the techniques are consistent. Of the many simulations we investigated, two techniques emerged as clearly superior and produced results that were close to each other in every state. Another positive outcome was that the two preferred approaches delivered equally precise estimates for all student subgroups and all school aggregations. Having all subgroups served equally well in a simulation is an important criterion for success. The winning simulations were combined to create the three sets of 2020 assessment proxies.

The second part of the paper zeros in on the estimates of learning loss both within and across states. These estimates point to troubling disparities in the severity of learning impact associated with the coronavirus pandemic across schools and across student groups. Quantifying these differences and identifying where needs are greatest for the 2020-2021 school year can help state and local education agencies plan for reopening and chart pathways for recovery.

Building Simulations for 2020 Assessment Scores

General Approach and Performance Criteria

Drawing on student-level data obtained through data sharing agreements with our state education agency partners, CREDO produced separate analyses for each state. We used five years of longitudinal student data as a test bed to build and test a number of possible simulations. We used data from the 2014-15 through 2018-19 school years. In each year, we included the scores from ESSA-mandated achievement assessments for 3rd through 8th grade and high school assessments. To enable the computations, we converted scaled test scores to standardized ones for simulation, then transformed back once the actual projections for 2020 assessments were completed.

With the five years of test scores, we pretended that the scores from 2017-2018 were missing. We then used the remaining data to simulate the missing 2017-2018 scores, using various approaches; we began with five possible approaches but eventually expanded to consider more than twenty. (See Appendix 1 for details.) Each approach produced a prediction for the 2017-2018 student achievement scores which we evaluated against the real, known scores.

The real score for 2017-2018 was a common standard for judging the accuracy of the various alternative simulations. We evaluated each simulation in two ways. The first was overall accuracy, gauged by the correlation of predicted and actual scores and by the average absolute error of the predicted scores. The second way focused on how well the simulations evidenced equivalent accuracy across student subgroups, grades and school attributes; for each grouping we used the average absolute error for each subset of students in the group.

Once the best performing approaches were identified for the 2017-2018 tests, we moved to apply the approaches to more recent data to estimate three sets of 2020 student achievement scores, as shown in Figure 1.

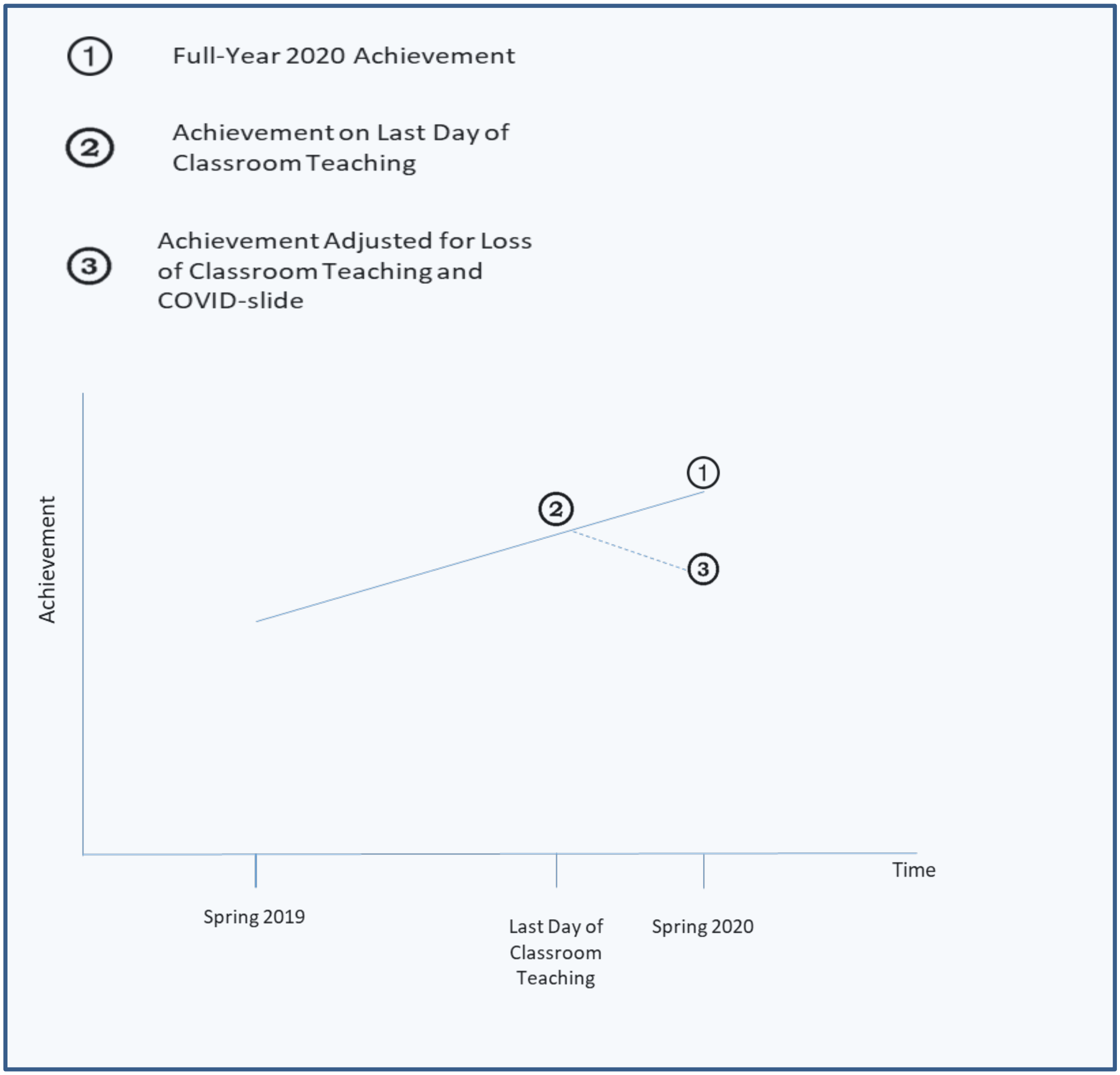
We developed proxies to estimate what students' achievement scores would have been in Spring 2020 if the pandemic had not occurred, shown as ①.

The second series, shown as ②, estimated students' achievement at the point in mid-March 2020 when school buildings were ordered closed. This series reflects the loss of classroom instruction for the remainder of the year and adjusts the full-year student achievement estimates accordingly.

To build estimates of the loss from March 2020 until the end of the school year, NWEA used historical sets of student-level records of assessments using their Measuring Academic Progress (MAP) interim assessment instruments which they aggregated to detailed student groups. CREDO supplemented NWEA's data with student profile information classified at the school-grade level. This approach assumes that the learning decay that students experienced between the closure of school buildings and the end of the school year would occur at the same pace as the learning loss NWEA has observed for students over the summer months. It also assumes that schooling effectively stopped for the year in mid-March. The pace of decay differs across student groups; for each state, there are 504 possible combinations of student characteristics based on academic grade, subject, percentof students in poverty, and concentrations of English learners or Special Education students. In a few detailed groups, there may be additional growth, as was seen in a small handful of schools that managed the pivot to remote learning smoothly and effectively.

Using the school-level student profiles, NWEA estimated what the loss in learning would be from ② to ③. Where NWEA had sufficient data, state-specific estimates were produced; otherwise, they produced a set of estimates from a national sample of students to use in the remainder of the participating states. CREDO then applied each school-grade level value to the individual student achievement proxies at the point of school building closure. These resulting proxies are the estimates of 2019-2020 student achievement adjusted fully for the impacts of the coronavirus pandemic, which combines loss of classroom instruction and learning loss over the rest of the school year. This series is the third set of estimates, shown in Figure 1 as ③. A full explanation of the COVID learning loss adjustments is available in Appendix 2.

Figure 1: Schematic of CREDO Simulations of Student Achievement Scores



The Simulations

We used technical and practical considerations in choosing the scenarios to investigate. On the practical side, we observed what many states elected to do about the missing Spring 2020 assessments and chose to model those. In addition, we looked at a range of predictive techniques used in other fields of public policy such as health, welfare and labor. By the end of the project, we had explored more than twenty different possible approaches.

The focus quickly centered on five "short-listed" student-level scenarios; each is described briefly below. Along the way, we realized that no approach would be able to produce proxies for certain groups of students for whom prior achievement information was either unavailable or temporally remote. Examples include students in 3rd grade in 2020, or high school geometry students who have no clear set of prior scores to use to produce individual proxies. Work-arounds were needed to address those situations: we computed 2-year running averages of grade or course performance at the grade-within-school or course-within-school level to act as a substitute for a individual-specific proxy.

- Do Nothing / State Mean for All

  Many states have chosen to deal with their lack of Spring 2020 assessment data by doing nothing. From a simulation viewpoint, this scenario sought to discern what the best guess of a student's score would be under this arrangement. The most logical choice would be to give all students the statewide grade-subject average from the most recent year of assessments. As an example, students in 6th grade in 2019-2020 would be assigned the state average score from the 2018-2019 math and reading assessments.

- Copy Past Year Scores

  We learned from our discussions with state officials that some intended to use the prior year assessments as the proxies for missing Spring 2020 scores. This approach requires the scores to be standardized so that the relative distributions of students in each grade cohort can be preserved. The prior year's scaled scores for the grade the student was enrolled in in 2020 brings the proxy into 2020 focus. This approach needs the work-arounds mentioned above for students enrolled in 2020 as 3rd graders or high school students in grades or courses that lacked a prior grade assessment score.

- Bridging

  Drawing on simulation practices in other policy fields, the bridging approach takes prior and future scores to impute the missing year value. We were able to test the accuracy of the bridging approach with earlier data, but imputing 2020 scores would require 2021 assessments to provide the future anchor. There are technical and political challenges with this approach. Since the 3rd grades in 2020 will not have a prior test score from 2019 and the high school students in the last tested grade or course in 2020 will not have a future test score from 2021, a work-around was needed for them. Political uncertainties about conducting 2021 assessments increase the risk that this approach is infeasible.

- Ordinary Least Squares Regression with One Prior Achievement Score

  Using differential calculus, ordinary least squares regression uses the relationships among a set of observed variables to predict a known outcome. This statistical technique is geared to identifying the best fit that minimizes prediction errors. CREDO estimated a student-level achievement model of 2016-17 scores based as a function of one prior test score, grade and individual student demographics. The model's results were then extrapolated to produce predictions for 2017-18 achievement scores that could be evaluated against the actual student scores. Since a prior test score is needed, this simulation cannot deliver estimates for 3rd graders; as in other approaches, a work-around was needed.

- Ordinary Least Squares Regression with Two Prior Achievement Scores

  This simulation parallels the preceding regression but includes two prior achievement scores. This approach requires work-arounds for 3rd grade and 4th grade since neither has two prior scores.

The Results for Simulations of 2017 –2018 Achievement Scores

Building simulations of achievement scores for 2017-2018 allowed direct comparison of each scenario's predictions against the true scores for that year. It was then possible to compare the different approaches against each other to determine which simulation proved most accurate. The best performing simulations were considered for creating the 2020 student-level proxy achievement scores.

Before delving in to the results, a few comments will create some context for reviewing them. First, the accuracy of the projection techniques was independent of the window of years used to test them. We pressure-tested the results by using other "missing" years as the simulation target but the results were the same. Consistency across different data windows gives us confidence that the simulations are robust over time and therefore could be expected to continue being effective when we create the 2020 achievement proxies. Second, these efforts use either straight mathematical computations or time-series projections to create our predicted achievement scores. We can be confident that the school year was fairly typical until the pandemic forced school facilities to close, so these techniques are suitable for these purposes. They are not, however causal estimations, so we are unable to discern any of the underlying mechanisms that may have influenced the predictions we obtain. Third, since student achievement is assessed annually, there is a limited number of measures for any given student. This causes tension between how much of a student's history to use and the number of students that could be included in the simulation calculations. Using longer data series means fewer students would be included, potentially introducing bias in our results. We opted for greater inclusion; as a result, the amount of data that is available for this effort is fairly shallow, which helps explain the errors we observe.

Table 1 presents the simulations' average absolute error and range of error results for the five scenarios for each state for Reading, shown in standard deviation units. Table 2 shows the same results for Math. Based on the average absolute error, the comparison of scenarios in every state yielded the same results: in both subjects, the scenarios performed (from worst to best) in the same order:

- State Mean
- Past Year's Score
- Regression with 1 Prior Score
- Regression with 2 prior Scores
- Bridging

Table 1: Simulation Diagnostics for Shortlisted Scenarios for 2017-18 Student Achievement in Reading - by State										
	Scenario 1 State Mean		Scenario 2 Past Year's Score		Scenario 3 Bridging		Scenario 4 Regression with 1 Prior		Scenario 5 Regression with 2 Priors	
State	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**
Arizona	0.815	3.646	0.489	5.776	0.377	3.772	0.428	5.368	0.397	5.502
Arkansas	0.816	3.332	0.455	4.191	0.337	2.914	0.384	3.979	0.352	3.625
District of Columbia	0.816	3.323	0.487	3.076	0.353	2.430	0.396	2.384	0.372	2.127
Illinois	0.811	3.630	0.510	5.058	0.396	3.503	0.430	3.316	0.404	3.164
Indiana	0.792	4.985	0.502	5.068	0.391	3.866	0.425	4.027	0.394	3.591
Kentucky	0.780	4.836	0.560	5.392	0.437	4.258	0.475	4.637	0.434	4.128
Louisiana	0.802	3.743	0.538	4.077	0.414	3.090	0.466	3.413	0.431	3.404
Michigan	0.833	3.605	0.500	5.645	0.375	4.747	0.426	4.446	0.396	4.033
Missouri	0.787	4.979	0.501	5.864	0.386	5.205	0.430	5.614	0.401	5.700
New Jersey	0.804	3.031	0.476	4.023	0.357	2.999	0.414	3.843	0.391	3.821
New Mexico	0.810	3.992	0.513	5.185	0.409	3.114	0.450	3.029	0.423	2.996
New York - Upstate	0.781	4.646	0.544	6.027	0.424	4.662	0.460	4.896	0.425	4.803
New York City	0.786	4.974	0.545	4.995	0.419	4.152	0.452	4.360	0.416	4.414
North Carolina	0.808	3.407	0.497	5.420	0.382	5.195	0.444	5.065	0.402	5.132
Rhode Island	0.806	3.379	0.508	3.477	0.387	2.637	0.423	2.977	0.396	2.648
South Carolina	0.828	4.895	0.479	5.286	0.356	3.841	0.394	4.657	0.372	3.741
Tennessee	0.777	4.924	0.528	7.100	0.415	5.178	0.497	7.150	n.a.	n.a.
Utah	0.808	3.097	0.502	4.363	0.384	4.362	0.437	3.901	0.408	3.987
Wisconsin	0.801	4.962	0.488	5.761	0.376	5.480	0.428	5.113	0.396	4.449
Minimum	0.777	3.031	0.455	3.076	0.337	2.430	0.384	2.384	0.352	2.127
Maximum	0.833	4.985	0.560	7.100	0.437	5.480	0.497	7.150	0.434	5.700

Notes:

(1) Columns titled *Mean* and *95% Range* show the average and 95% error value range (97.5th percentile-2.5th percentile), respectively, of the *absolute simulation error* overall in standard deviation units for each shortlisted simulation scenario in each State.

(2) Row titled *Minimum (Maximum)* shows the minimum (maximum) of the absolute simulation error statistic represented in each column.

(3) Scenario 5: Regression with 2 priors is not applicable in Tennessee because of the lack of 2015-16 test scores and the associated simulation diagnostics are not available (n.a.).

(4) Shortlisted scenarios presented:

i. State Mean: student simulated scores in a specific grade are equal to past year’s grade-average score in the State.*

ii. Past Year's Score: each student's simulated score is equal to his or her score from the previous year.*

iii. Bridging: each student's simulated score is equal to the average of his or her actual achievement in the previous year and the following year.*

iv. Regression with 1 prior: each student's simulated score comes from a statistical model that includes past year's score.*

v. Regression with 2 priors: each student's simulated score comes from a statistical model that includes scores from two prior years.*

*For details and special cases, please see Appendix 1.

Table 2: Simulation Diagnostics for Shortlisted Scenarios for 2017-18 Student Achievement in Math - by State										
	Scenario 1 State Mean		Scenario 2 Past Year's Score		Scenario 3 Bridging		Scenario 4 Regression with 1 Prior		Scenario 5 Regression with 2 Priors	
State	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**	**Mean**	**Range**
Arizona	0.810	4.030	0.496	4.117	0.376	3.363	0.430	3.365	0.404	3.492
Arkansas	0.806	4.897	0.537	5.275	0.424	3.359	0.469	4.049	0.442	4.236
District of Columbia	0.810	4.026	0.518	3.391	0.388	2.470	0.438	3.482	0.428	3.083
Illinois	0.807	4.030	0.486	4.836	0.362	4.198	0.410	4.524	0.386	4.959
Indiana	0.768	4.717	0.458	6.754	0.345	4.075	0.395	5.641	0.370	4.879
Kentucky	0.785	4.851	0.527	6.107	0.403	5.653	0.449	5.642	0.412	4.305
Louisiana	0.805	4.600	0.526	5.438	0.404	3.309	0.453	4.421	0.422	3.805
Michigan	0.812	3.254	0.468	5.338	0.335	3.843	0.392	4.618	0.365	3.644
Missouri	0.766	4.983	0.529	6.575	0.396	4.510	0.448	6.287	0.421	6.489
New Jersey	0.808	4.134	0.483	4.323	0.379	3.420	0.414	3.632	0.387	3.225
New Mexico	0.802	4.510	0.547	3.845	0.443	3.056	0.479	3.312	0.453	3.291
New York - Upstate	0.777	4.938	0.530	5.595	0.386	4.437	0.439	4.949	0.408	4.988
New York City	0.792	4.718	0.549	5.197	0.418	4.646	0.465	4.850	0.441	5.097
North Carolina	0.823	3.578	0.511	4.160	0.399	3.556	0.450	3.771	0.418	3.867
Rhode Island	0.806	3.719	0.489	3.902	0.352	2.213	0.419	3.074	0.387	3.207
South Carolina	0.807	4.893	0.508	6.278	0.379	3.438	0.429	5.134	0.410	4.010
Tennessee	0.780	4.717	0.526	5.426	0.411	5.491	0.506	5.472	n.a.	n.a.
Utah	0.798	2.806	0.461	4.745	0.341	4.354	0.392	4.333	0.370	4.224
Wisconsin	0.778	4.480	0.487	5.222	0.391	3.884	0.433	4.664	0.407	3.931
Minimum	0.766	2.806	0.458	3.391	0.335	2.213	0.392	3.074	0.365	3.083
Maximum	0.823	4.983	0.549	6.754	0.443	5.653	0.506	6.287	0.453	6.489

Notes:

(1) Columns titled *Mean* and *95% Range* show the average and 95% error value range (97.5th percentile-2.5th percentile), respectively, of the *absolute simulation error* overall in standard deviation units for each shortlisted simulation scenario in each State.

(2) Row titled *Minimum (Maximum)* shows the minimum (maximum) of the absolute simulation error statistic represented in each column.

(3) Scenario 5: Regression with 2 priors is not applicable in Tennessee because of the lack of 2015-16 test scores and the associated simulation diagnostics are not available (n.a.).

(4) Shortlisted scenarios presented:

i. State Mean: student simulated scores in a specific grade are equal to past year’s grade-average score in the State.*

ii. Past Year's Score: each student's simulated score is equal to his or her score from the previous year.*

iii. Bridging: each student's simulated score is equal to the average of his or her actual achievement in the previous year and the following year.*

iv. Regression with 1 prior: each student's simulated score comes from a statistical model that includes past year's score.*

v. Regression with 2 priors: each student's simulated score comes from a statistical model that includes scores from two prior years.*

*For details and special cases, please see Appendix 1.

The best approaches – Bridging and Regression with 2 prior scores -- had an average error around .35 standard deviations, while the worst performing approach – using the state average scores on the most recent achievement tests -- had twice the error of the best. These values illustrate how inherently noisy achievement measurement is, since it blends a large number of influences such as personal aptitude, quality of current schooling, prior learning and personal circumstances into a single measure. Moreover, even with sophisticated multivariate computations that are geared to driving down prediction error to the greatest degree possible, the predicted scores still had a lot of random error. Given the inherent degree of error that these computations produced, it is no surprise that all the simulations had overall error that was statistically significant at the $p \leq .01$ level.

Appendix 3 includes additional details on how well the scenarios predicted scores for different demographic groups, for students in poverty and for students with particular learning needs. Those tables show that the scenarios created the same overall rankings as found for the full student population. More important, however, is that each scenario produced sub-group level errors that were generally equivalent across the subgroups. The size of the overall average absolute error is reflected in all the breakouts we examined.

Table 1 and Table 2 also present the range of the errors, measured in standard deviations between the predicted and the actual student score. The smallest errors hovered at zero, meaning the predicted score equaled the actual score. The values in Table 1 and Table 2 show the spread between the 2.5th percentile and the 97.5th percentile. For this measure, we discarded the extreme outliers, of which there were only a handful, because they more than tripled the range overshadowing the picture of dispersion for 95 percent of observations. The values for the 95 percent range follow those of the overall average error: where the overall average error is larger, as in Scenarios 1 and 2, the ranges are also wider. The preferred scenarios based on the average error also have the smallest range of estimated achievement scores. This result strengthens confidence that these simulation approaches are the best available options to pursue to create the 2020 achievement proxies.

In contrast, each scenarios performed consistently across the states. Looking at the minimum error and the maximum average absolute error for each scenario, the difference between the two was in the range of .056 - .114 standard deviations. These results give us confidence to build the 2020 proxy achievement series using the same approach for every state.

Tradeoffs in Building the 2020 Achievement Proxies

Considering the average absolute error values, we dropped from further consideration Scenario 1 – State Means and Scenario 2 – Past Year's Score and moved forward with the remaining three scenarios.

Scenario 3, Bridging, proved the best on historical simulations, but is of no practical value for the present. Further, considerable uncertainty surrounds the administration of the 2021 achievement tests, raising the question of whether the approach will ever have its day.

As discussed earlier, the regression models, Scenarios 4 (Regression with 1 prior) and 5 (Regression with 2 priors), respectively, each use scores on prior achievement tests in their computations of predicted scores. Building the estimated scores on known trajectories helps to reduce the error, but it also means that the early grades and high school students with test and data gaps will not be included in the estimation since they lack the requisite historical scores. For Scenario 4, only the 3rd grade is excluded, but 3rd and 4th grade are affected in Scenario 5.

Our final protocol for creating the full-year 2020 achievement proxies, shown as ① in Figure 1, ultimately involved a bit of patchwork. We used Scenario 5 as the approach for grades 5 through 8 and contiguously tested high school grades. We augmented that approach in three ways. First, we used Scenario 4 to create scores for 4th graders in 2020. The two scenarios have similar overall performance. (We considered an alternative that created a running average at the grade-within-school level, but found that the errors were larger than swapping in the results from Scenario 4.) For 3rd graders and high school students with test gaps, we used the school's average of the past two years of student scores to create school-specific averages. Finally, for high school end-of-course assessments with no prior tests, we created school-specific historical average scores by course. These estimates cannot be associated with individual students, since we did not have 2020 enrollment or course assignment information.

Using the full-year 2020 achievement proxies as the foundation, we created the second set of proxies by subtracting the amount of learning that the average student would have learned in the last 2 months of the school year. Using CREDO's Days of Learning transformation, the loss of 58 days of classroom based instruction translates to .1 standard deviations of achievement, which was deducted from each student' full-year proxy score. The resulting value corresponds to the achievement of students at the point of school building closures, shown as ② in Figure 1.

All three sets of student-level and school proxies were delivered to the participating states.

Estimates of Learning Loss in the 2019-2020 School Year

We present the average learning loss by state. Tables 3 and 4 present these estimates for Reading and Math, respectively. The values reflect the difference in achievement that would have occurred absent the pandemic and the estimated measure of student learning at the conclusion of the school year with the disruptions that did occur. Part of the loss, -.1 std, can be attributed to the lost class-based days of instruction and applies across the board. The remainder of the loss comes from the decay of learning or "slide" associated with out-of-school time.

Since the learning loss estimates are grounded in the projections of missing 2020 achievement scores, the individual student level estimates inevitably carry a higher degree of "noise" than if real assessment scores were used. When the estimates are aggregated, the noise is reduced, but may not be entirely eliminated. Accordingly, these values should be viewed as approximations, not precise point-estimates.

We can provide clearer insight into these values by considering how many school days of learning were lost. CREDO has routinely converted standard deviation units to Days of Learning based on progress on the National Assessment of Education Progress (NAEP). One standard deviation of achievement equates to 3.22 years of school, or about 580 days. For a typical 180-day school year a .1 std reflects 58 days of learning.

In Table 3, South Carolina is identified as having the largest average learning loss at -.316 std in Reading. North Carolina had the smallest learning loss with -.097 std. Converted to lost school days of learning, North Carolina lost 57 days of learning while South Carolina lost 183 (or a complete school year) in Reading. States and schools differ in estimated learning losses based on the variations in both historical school performance and the degree of achievement slide related to differences in student profiles across schools.

Table 4 shows learning losses in Math were greater than shown for Reading in every state. The average learning losses in Math were greatest in Illinois, where students lost .4 std in achievement from their full-year estimated values. Wisconsin posted the smallest average loss with -.235 std. Translated into days of learning lost, students in Illinois declined about 232 days (or more than a year) and Wisconsin students lost about 136 days.

The differences within states are also noteworthy. Tables 3 and 4 include columns that display the largest estimate of learning loss and the smallest in each state as well as the range between the two. Recall that these estimates are computed at the school-grade level and applied to every student's achievement proxy, so the offset is constant for all students in a school-grade even though the individual achievement proxies in ③ will vary.

Looking at the learning losses for Reading presented in Table 3, the states showed sharp differences in their within-state variation in learning loss. The largest estimated loss was located in Tennessee where in at least one school students faced a loss of -.734 std. North Carolina has a closely similar estimate of -.715 std. These contrast with Rhode Island or the District of Columbia, where the largest learning loss was -.267 std and -.278, respectively. The smallest estimates of learning loss were also varied across the states. Recalling that the adjustment for lost classroom instruction was -.1 std, most states were able to reverse that loss to some extent in some of their schools. In Arizona and Indiana, for instance, the "lost days" effect was almost eliminated in some schools. In contrast, South Carolina's best case school increased the learning loss slightly. Unexpectedly, NWEA's prior experience with learning slide revealed that in some situations, students not only don't lose learning, they actually gain over the period. Arkansas, Missouri, New Jersey, North Carolina, Tennessee and Wisconsin all had schools in which some grades had positive impact, with Arkansas having the largest at .534 std.

In Math, shown in Table 4, South Carolina had the largest estimated learning loss with -.961 std. The state whose worst learning loss was the smallest was Arkansas with -.482 std. When looking at the best-case estimates of learning loss in each state, most positive estimate of learning loss in Illinois was -.212 std. As with Reading, there were states where schools were estimated to make positive achievement gains despite the interruptions of the pandemic. Some Wisconsin schools posted positive gains of .124 std. Similar positive estimates were obtained for some school-grades in Arkansas (.095 std), North Carolina (.035 std.) and Tennessee (.028 std).

Looking at both the cross state differences in the average learning loss and the differences in dispersion around those averages, it is not surprising that the range of scores is as different across the state as shown in the final columns of Tables 3 and 4. For Reading, the largest spread (1.027 std in Arkansas) is more than six times larger than the smallest, seen in Rhode Island with .168 std. For Math, the largest range seen in South Carolina is not quite double the tightest range seen in Louisiana with .455 std.

Table 3: Estimated 2019-20 Pandemic-Related Learning Loss in Reading				
	Overall			
State	**Mean**	**Largest**	**Smallest**	**Range**
Arizona	-0.146	-0.393	-0.004	0.389
Arkansas	-0.107	-0.493	0.534	1.027
District of Columbia	-0.218	-0.278	-0.096	0.182
Illinois	-0.217	-0.408	-0.111	0.297
Indiana	-0.223	-0.574	-0.007	0.567
Kentucky	-0.209	-0.510	-0.055	0.455
Louisiana	-0.171	-0.278	-0.079	0.199
Michigan	-0.211	-0.615	-0.091	0.524
Missouri	-0.173	-0.459	0.163	0.623
New Jersey	-0.121	-0.291	0.059	0.351
New Mexico	-0.169	-0.278	-0.076	0.202
New York - Upstate	-0.180	-0.269	-0.073	0.196
New York City	-0.215	-0.278	-0.099	0.179
North Carolina	-0.097	-0.715	0.209	0.924
Rhode Island	-0.191	-0.267	-0.099	0.168
South Carolina	-0.316	-0.535	-0.123	0.412
Tennessee	-0.151	-0.734	0.183	0.917
Utah	-0.155	-0.278	-0.073	0.205
Wisconsin	-0.165	-0.478	0.225	0.703
Minimum	-0.316	-0.734	-0.123	0.168
Maximum	-0.097	-0.267	0.534	1.027

Notes:

(1) Columns titled *Mean, Largest,* and *Smallest* show the average, minimum value, and maximum value, respectively, of the estimated total learning loss in 2019-20 student achievement in standard deviation units in each State.

(2) Column titled *Range* shows the difference between the smallest and largest (smallest–largest). estimated total learning loss in 2019-20 student achievement in each State.

(3) Row titled *Minimum (Maximum)* shows the minimum (maximum) of the total learning loss statistic represented in each column.

(4) Estimated learning loss at the end of the school year includes both the effect of school building closures and the learning slide.

Table 4: Estimated 2019-20 Pandemic-Related Learning Loss in Math				
	Overall			
State	**Mean**	**Largest**	**Smallest**	**Range**
Arizona	-0.299	-0.658	-0.175	0.483
Arkansas	-0.238	-0.482	0.095	0.577
District of Columbia	-0.375	-0.657	-0.198	0.459
Illinois	-0.402	-0.720	-0.212	0.507
Indiana	-0.360	-0.700	-0.193	0.507
Kentucky	-0.297	-0.749	-0.024	0.725
Louisiana	-0.347	-0.631	-0.175	0.455
Michigan	-0.336	-0.772	-0.154	0.618
Missouri	-0.283	-0.783	-0.110	0.674
New Jersey	-0.343	-0.794	-0.022	0.771
New Mexico	-0.359	-0.657	-0.175	0.482
New York - Upstate	-0.386	-0.804	-0.175	0.629
New York City	-0.365	-0.657	-0.175	0.482
North Carolina	-0.335	-0.788	0.035	0.823
Rhode Island	-0.355	-0.657	-0.175	0.482
South Carolina	-0.391	-0.961	-0.116	0.845
Tennessee	-0.273	-0.580	0.028	0.608
Utah	-0.307	-0.657	-0.130	0.527
Wisconsin	-0.235	-0.597	0.124	0.722
Minimum	-0.402	-0.961	-0.212	0.455
Maximum	-0.235	-0.482	0.124	0.845

Notes:

(1) Columns titled *Mean, Largest,* and *Smallest* show the average, minimum value, and maximum value, respectively, of the estimated total learning loss in 2019-20 student achievement in standard deviation units in each State.

(2) Column titled *Range* shows the difference between the smallest and largest (smallest–largest). estimated total learning loss in 2019-20 student achievement in each State.

(3) Row titled *Minimum (Maximum)* shows the minimum (maximum) of the total learning loss statistic represented in each column.

(4) Estimated learning loss at the end of the school year includes both the effect of school building closures and the learning slide.

Insights for Policy

The aim of this meta-analysis was to make technical and policy contributions to the efforts to mitigate the impacts of the coronavirus pandemic on public education in the United States. The findings in no way denigrate the heroic efforts of policy makers and educators to find alternative ways to educate students in safe environments. Those efforts continue and cannot be fairly valued.

At the same time, these simulations offer important insights into the contemporary landscape of US public K-12 education. They demonstrate the challenges that arise when critical measures of student performance are interrupted. They also tell a stark story of harm that our students have experienced. The discussions going forward need to focus on their needs. Those discussion can be informed by several implications that arise from the work presented here.

Implications from Simulations

There are three chief insights from these simulations that might guide future efforts to fill in when states or districts encounter periods of missing data. The first is that immediate history is not a good "fix" for missing performance data. Achievement growth was shown to be highly variable from year to year, even after historical trends were taken into account.

Second, the insight that the errors for a given simulation produced more consistent errors across states than did different simulations within states points to the possibility that the states are more alike than they are different in their production of education. This bodes well in the coming years when all states will need to plot extensive recovery plans. State recovery strategies will differ, providing important opportunities to identify policy and program elements that contribute positively to supporting students and teachers as they move forward.

Third, information about student achievement serves a wide range of purposes in state and local education agencies, as presented in Table 5. Knowing that the best performing simulations required multiple years of data to estimate the missing achievement values for Spring 2020, there is a clear hazard to education agencies if more than a year of assessments is deferred. This has immediate and urgent implications for state education leaders.

Table 5: How Assessment Data Are Used	States	Districts
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

Implications from Learning Loss Estimates

The findings on learning losses support four general inferences. First, the findings are chilling – if .31 std equals a full year of learning, then recovery of the 2019-2020 losses in our nation could take years. These findings are corroborated by independent "top-down" estimates of the long-term macroeconomic effects of the losses: a three percent decline in annual earnings lifetime for the affected students, summing up to a 1.5% reduction in Gross Domestic Product until the end of the century.[3] The estimates presented here are larger, in part because these use a "bottom-up" approach based on individual students. Any further losses incurred in the current year will create additional impacts and extend the recovery timeline. In addition, the underlying variations in 2019-2020 learning losses highlight the fact that school closures had highly differentiated impacts, with disadvantaged students generally suffering much more than students from advantaged families.

Second, the wide variation within states (and often within schools) means that conventional models of classroom based instruction – a one-to-many, fixed pace approach -- will not meet the needs of students in the 2020-2021 school year. New approaches must be allowed to ensure high quality instruction is available in different settings, recognizing that different skills may be needed for the different channels.

Third, the need for rigorous student-level learning assessments has never been higher. In particular, this crisis needs strong diagnostic assessments and frequent progress checks, both of which must align with historical assessment trends to plot a recovery course. The losses presented here implicitly endorse a return to student achievement testing *with the same assessment tools* for the foreseeable future. At the same time, preserving and expanding the existing series is the only way to reliably track how well states and districts are moving their schools through recovery and into the future.

Fourth, the measures of average loss and the range around it immediately call into question the existing practice of letting communities plot their own path forward. The communities whose schools have the largest estimated loss of learning are far less likely to have the means and capacity to create and implement recovery plans on their own. Insistence on local autonomy in this case will not yield equitable responses.

[3] Hanushek, E. and L. Woessmann (2020), "The economic impacts of learning losses", *OECD Education Working Papers*, No. 225, OECD Publishing, Paris, https://doi.org/10.1787/21908d74-en.