Washington State Institute for Public Policy
110 Fifth Avenue SE, Suite 214 ● PO Box 40999 ● Olympia, WA 98501 ● 360.664.9800 ● www.wsipp.wa.gov

September 2023

Student Achievement and the Pandemic: *Analysis of Test Scores, Earnings, and Recovery Interventions*

The Washington State Institute for Public Policy (WSIPP) receives funding from the legislature to conduct research on K-12 education topics. In this report, we examine academic achievement among public school students in Washington during the COVID-19 pandemic.

First, we examine changes to student achievement before and after the pandemic and explore effects by student, school, and regional characteristics. Next, we estimate how changes to achievement (measured using test scores) influence future earnings. Finally, we identify interventions that may help support academic recovery.

Section I provides an overview of the report and main research questions. Section II offers background information on the COVID-19 pandemic, including existing research and the landscape of K-12 education in Washington. Section III describes the methodological approach we used to analyze test scores during the pandemic. Section IV summarizes the results of this analysis. Section V describes how we predict changes to future earnings, results, and limitations. Section VI describes interventions that may help recover student academic achievement in the coming years. Finally, Section VII details key takeaways.

Summary

In this report, we estimate how student math and English Language Arts (ELA) achievement changed during the COVID-19 pandemic, predict long-term effects on student earnings, and identify several interventions that may support academic recovery in the future.

Overall, we found that average math and ELA test scores were lower in 2022 than average scores before the pandemic, and math scores fell more than ELA scores. We observed the largest test score declines in middle school grades.

Further, we found larger test score declines among female students, students of color, and low-income students.

On average, we estimate that test scores fell 0.20 standard deviations (SD) in 2022, which is associated with a $32,000 decrease in future earnings per student. Due to variation in effects across grades, students in middle school during the pandemic may experience a greater loss in future earnings than elementary and high school students.

Interventions like tutoring, summer school programs, and double-dose classes may offset the decline in test scores we observe and help students recover academically after the pandemic.

Cramer, J., & Krnacik, K. (2023). *Student achievement and the pandemic: Analysis of test scores, earnings, and recovery interventions.* (Document Number 23-09-2201). Olympia: Washington State Institute for Public Policy.

I. Introduction

WSIPP receives funding from the legislature to conduct research on K-12 education topics that are relevant in Washington.[1] Based on research and conversations with stakeholders, including nonpartisan legislative staff and Office of Superintendent of Public Instruction (OSPI) staff, we decided to examine how student achievement changed during the COVID-19 pandemic and identify potential long-term consequences.

Overall, this report answers the following questions:

1) How have math and English Language Arts (ELA) test scores changed during the pandemic?
2) Are there different effects by student, school, and regional characteristics?
3) How do test score changes influence future earnings?
4) What interventions may help students recover academically, and by how much?

To answer the first two questions, we use data provided by OSPI to analyze how student achievement in mathematics and ELA changed during the pandemic and explore differences by student characteristics like gender, race, and income status, as well as by school and regional characteristics.

To answer the third question, we use WSIPP's benefit-cost model to estimate how changes to student achievement may influence earnings in the future.

Finally, to address the last question, we review previous meta-analyses to explore how the impact of several interventions may help students recover academically.

[1] Engrossed Substitute Senate Bill 6386, Chapter 372, Laws of 2006.

II. Background

This section reviews research on the pandemic's influence on the public education system and students in the United States and Washington.

The sudden outbreak of the coronavirus resulted in widespread economic, health, and social impacts. Government shutdowns to curb the spread of the virus led to job losses for millions of U.S. employees, economic uncertainty for families, shifts to online work, and changes to cultural ideas about work.[2] The virus' global spread affected individuals' physical and mental health and profoundly shaped the health care infrastructure and workforce.[3] Spillover effects from the pandemic are difficult to quantify but continue to shape daily lives.

The pandemic also resulted in an unprecedented shock to the education system. In March 2020, states implemented stay-at-home orders and closures of non-essential businesses and public and private schools to limit the spread of the virus.[4] During this period, seven out of ten schools in the U.S. moved some or all instruction online.[5] Over the next year and a half, students and educators were required to adapt to remote and hybrid instruction environments.[6]

Research Review

Though research focusing on student learning over the pandemic is still developing and will continue for years to come, in this section, we discuss early evidence on academic and non-academic outcomes.

Researchers and media often frame the disruption to student learning during the pandemic as "learning loss." This refers to the difference between the knowledge and skills students gained during the pandemic and the expected learning they would have gained in the absence of the pandemic, given historic learning trends.[7] In this report, we explore average math and ELA test scores in school year 2022, compared to historic averages before the pandemic.[8]

[2] Pew Research Center. *Fewer jobs have been lost in E.U. than in the U.S. during the COVID-19 downturn;* U.S. Bureau of Labor Statistics. *Labor Force Statistics from the Current Population Survey.*
[3] Lee, W., Park, S.W., & Viboud, C. (2023). *Direct and indirect mortality impacts of the COVID-19 pandemic in the U.S.* Epidemiology and Global Health; Deng, J., Zhou, F., Hou, W., Heybati, K., Lohit, S., Abbas, U., Heybati, S.(2022). *Prevalence of mental health symptoms in children and adolescents during the COVID-19 pandemic: a meta-analysis*. *Annals of the New York Academy of Sciences, 1512*(1); Ettman, C.K., Fan, A.Y., Subramanian, M., Adam, G.P., Goicoechea, E.B., Abdalla, S.M. . . . Galea, S.(2023). *Prevalence of depressive symptoms in U.S. adults during the COVID-10 pandemic: a systematic review.* SSM-Population Health, 21; and ASPE (2022). *Impact of the COVID-19 pandemic on hospital and outpatient clinician workforce.*
[4] Center for Disease Control and Prevention. *CDC Museum COVID-19 Timeline.*
[5] National Center for Education Statistics. *U.S. education in the time of COVID.*
[6] Hybrid instruction combines remote-and in-person instruction.
[7] Donnelly, R., & Patrinos, H.A. (2022). *Learning loss during COVID-19: an early systematic review*. *Prospects, 51*(4).
[8] School year defined using last year in academic calendar (e.g., 2022 refers to 2021-2022 school year).

To date, studies across the United States have found that, regardless of grade level, students experienced decreases in academic growth during the 2019-2020 and 2020-2021 school years. Some research shows that learning was impacted more in the first year of the pandemic than the second year; other studies report similar effects in both years.[9] While there is some evidence that test scores are rebounding, scores have not recovered to pre-pandemic levels.[10]

Consistently, studies show that students experienced greater decreases in math achievement than reading, though overall, test scores in both subjects fell.[11]

Research also indicates that the pandemic contributed to educational inequalities because some student populations were affected more than others. Low-income students, students of color, and students with disabilities experienced larger decreases in test scores compared to their economically advantaged, non-disabled, and White peers. Further, pre-existing achievement gaps between low- and high-income students, as well as gaps between students of color and White students have widened over the course of the pandemic.[12]

Research has also found that, on average, declines in achievement were larger in school districts that implemented remote instruction for longer periods.[13] However, it is important to note that because the pandemic had large societal effects, it is difficult to attribute what portion of the effect is due to remote instruction or other economic, health, and social factors that also influence student performance.[14]

There is also growing research on the pandemic's influence on non-academic outcomes. So far, studies have reported that during the pandemic, students experienced increased levels of anxiety, depression, and stress compared to before the pandemic.[15] Alternatively, some studies found that reported substance use and rates of suicide decreased during this period.[16]

[9] Cohodes, S., Goldhaber, D., Hill, P., Ho, A., Kogan, V., Polikoff, M., West, M. (2022). *Student achievement gaps and the pandemic: a new review of evidence from 2021-2022.* Center on Reinventing Public Education and Kuhfeld, M., Soland, J., & Lewis, K. (2022). *Test score patterns across three COVID-19-impacted school years*. Annenberg Institute at Brown University.

[10] Kuhfeld, M., & Lewis, K. (2022). *Student achievement in 2021-2022: cause for hope and continued urgency*. Northwest Evaluation Association and Halloran, C., Hug, C.E., Jack, R., & Oster, E.(2023). *Post COVID-19 test score recovery: initial evidence from state testing data*. National Bureau of Economic Research.

[11] Cohodes et al. (2022); Betthauser, B.A., Back-Mortensen, A.M., & Engzell, P. (2022). *A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic*. Nature Human Behaviour.

[12] Goldhaber, D., Kane, T.J., McEachin, A., Morton, E., Patterson, T., & Staiger, D.O. (2022). *The consequences of remote and hybrid instruction during the pandemic*. National Bureau of Economic Research; Kilbride, T., Hopkins, B., Strunk, K.O., & Imberman, S. (2021). *K-8 student achievement and achievement gaps on Michigan's 2020-21 benchmark and summative assessments*. Education Policy Innovation Collaborative; Kuhfeld & Lewis (2022).

[13] Jack, R., Halloran, C., Okun, J., & Oster, R. (2023). Pandemic schooling mode and student test scores: evidence from U.S. school districts. *American Economic Review, 5*(2); Goldhaber et al. (2022).

[14] Ravens-Sieberer, U., Kaman, A., Erhart, M.m Devin, J., Schlack, R., & Otto, C.(2021*).* Impact of the COVID-19 pandemic on quality of life and mental health in children and adolescents in Germany. *European Child Adolescents Psychiatry, 31*(6); Elharake, J.A., Akbar, F., Malik, A.A., Gilliam, W., & Omer, S.B.(2023). Mental health impact of COVID-19 among children and college students: a systematic review. *Child Psychiatry Human Development, 54*(3); Newlove-Delgado, T., McManus, S., Sadler, K., Thandi, S., Vizard, T., Cartwright, C. . . . Ford, T.(2021). Child mental health in England before and during the COVID-19 lockdown. *The Lancet Psychiatry, 8*.

[15] Arenas, A., & Gortazar, L. (2022). Learning loss one year after school closures: evidence from the Basque country. *Esade, 277*.

[16] Hansen, B., Sabia, J.J., & Schaller, J. (2022). *In-person schooling and youth suicide: evidence from school calendars and pandemic school closures*. National Bureau of Economic Research; Joint Legislative Audit & Review Committee (2023). *Racial equity effects of restricting in-person education during the COVID-19 pandemic.*

Further, there is evidence that low-income students and those with pre-existing mental health conditions were more likely to be affected by the pandemic.[17]

The Pandemic and Washington Schools

In January 2020, the first case of coronavirus in the U.S. was reported in Washington.[18] In March, Governor Inslee announced an executive order to close all K-12 public and private schools, a decision later extended for the rest of the 2020 school year.[19] At this time, educators and students shifted to remote instruction.

In the 2021 school year, school districts coordinated with local health departments to determine how to return to in-person learning. Decisions were based on factors like local infection rates, student needs, access to internet, and the ability to offer safe learning environments.[20]

Students received remote or hybrid instruction for most of the 2021 school year. In the spring of 2021, Governor Inslee required all schools to begin offering in-person instruction at least two days a week, though the return to in-person instruction remained optional for families through the end of the year. By the start of the 2022 school year, most schools were operating fully in-person.

During the pandemic, enrollment in Washington public schools declined. Between 2019 and 2022 (the years immediately before and after the pandemic), enrollment decreased by about 45,000 students.[21]

Though enrollments fell across all grades, a large portion of this decline occurred in pre-kindergarten and kindergarten. Part of the enrollment decrease was due to decisions by families to delay enrolling their children into pre-kindergarten and kindergarten classes or transferring children to private school and home-school settings.[22]

In terms of student population over time, the proportion of White students in public schools slightly decreased during this period, and the proportion of Hispanic, Asian, Black, and Native Hawaiian or other Pacific Islander students slightly increased.

[17] Ravens-Sieberer et al. (2021*).*

[18] The Seattle Times. *Snohomish county man has the United States' first known case of the new coronavirus.*

[19] Washington Governor Jay Inslee website, March 13, 2020. *Inslee announces statewide school closures, expansion of limits on large gatherings.*

[20] Office of Superintendent of Public Instruction. *Reopening Washington Schools 2020: District Planning Guide.;* Washington Governor's Office. *Inslee announces education recommendations for 2020-2021 school year.*

[21] OSPI report card.

[22] State Board of Education (2022). *Education System Health Report.*

The Pandemic's Impact on the Administration of Learning Assessments

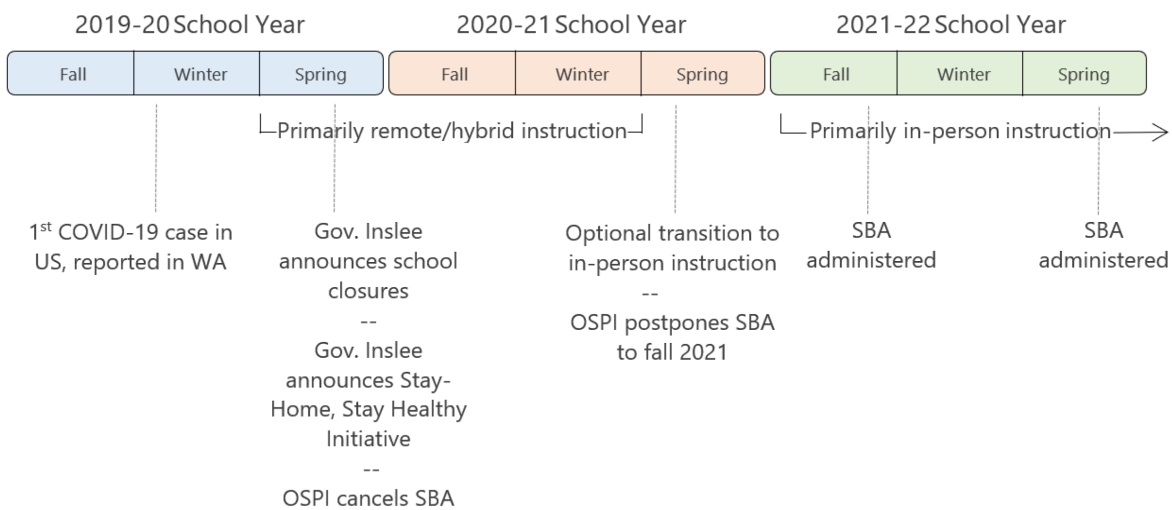
In Washington, students are assessed each spring to measure learning in various subjects and grades. Students in grades 3 through 8 and 10 receive the Smarter Balanced Assessment (SBA), a standardized assessment measuring proficiency in math and English Language Arts (ELA).[23]

After Governor Inslee announced school closures in March 2020, OSPI canceled spring 2020 assessments.[24] In the spring of 2021, OSPI further postponed the SBA. As a result, students were assessed twice in the 2022 school year, in fall and spring.[25]

Exhibit 1 provides a timeline of when schools implemented remote, hybrid, and in-person instruction and when SBA assessments were administered.[26]

Exhibit 1

Timeline of School Closures, Remote Instruction, and SBA Administration



[23] Students with cognitive disabilities and English language learners are also assessed in the spring with different assessments. Student's science proficiency is also assessed.
[24] U.S. Department of Education. *COVID-19 Waiver*
[25] Students were tested in fall of the 2022 school year based on their grade-level during the 2021 school year. In spring of the 2022 school year, students were assessed based on their current grade level. *OSPI Bulletin # 037-21.*
[26] Prior to the pandemic and school closures, the SBA had been administered in the spring of the 2018-19 school year.

Emergency Relief Funds & Recovery Response

In response to the pandemic, the federal government provided state education agencies with three rounds of Elementary and Secondary Schools Emergency Relief (ESSER) funds. In total, OSPI received almost $2.9 billion, 90% of which was directly allocated to school districts.[27] As of the publication of this report, districts have spent 70% of their ESSER funds and have until September 2024 to spend the remaining amount.[28]

Over the course of several funding rounds, school districts have used this support for a variety of activities. Early during school closures, districts used funds to provide students with basic needs like meals and internet access and to purchase personal protective equipment.[29] Later in the pandemic, districts used funds to bolster staff and student learning activities. For example, some districts hired counselors, some focused on student re-engagement, and others focused on after-school and summer programs.

In the final round of funding, school districts were required to spend at least 20% of their allocation (about $333 million) on "learning loss" activities. Districts were directed to implement "evidence-based interventions, ensure that those interventions respond to students' social, emotional, and academic needs, and address the disproportionate impact of COVID-19 on underrepresented students."[30]

To receive ESSER funds, school districts submitted Academic and Student Well-Being Recovery Plans to OSPI.[31] In their plans, districts were required to identify which student groups needed additional support, how they would provide that support, and how they would address learning recovery.

In their recovery plans, most districts reported they would use funds to provide social-emotional learning and mental health support, summer school programs, and additional institutional time before or after school. Some districts also indicated they would use ESSER funds to provide high-quality tutoring.[32]

Later in the report, we discuss several interventions that school districts in Washington and other states have considered implementing to help students recover academically after the pandemic. These interventions, which include various tutoring models, academically-focused summer school programs, and double-dose classes, may be options that school districts in Washington can implement using ESSER funds.

[27] *OSPI review of ESSER and state funds.*
[28] U.S. Department of Education. *Education stabilization fund.*
[29] Office of Elementary and Secondary Education. *Frequently asked questions about the elementary and secondary school emergency relief fund; Fact Sheet: Elementary and Secondary School Emergency Relief Fund II;* and OSPI. How Washington schools are using their emergency relief funding.
[30] Office of Elementary and Secondary Education. *American rescue plan act of 2021. Elementary and secondary school emergency relief fund.*
[31] OSPI. *Academic and student well-being recovery plan: planning guide 2021.*
[32] We cannot identify which activities districts have implemented or not. Joint Legislative Audit & Review Committee. (2023). *Racial equity effects of restricting in-person education during the COVID-19 pandemic.*

III. Methodology

In this section, we describe our methodological approach to analyzing math and ELA test scores. The next section reports our results.

In this analysis, we answer the following research questions:

- How have math and ELA test scores changed during the pandemic?
- Are there different effects by student, school, and regional characteristics?

Data and Outcomes

To conduct this analysis, we received administrative data from OSPI for school years 2015 through 2022. Data files included unidentifiable student-level demographics, school enrollment, program enrollment, and assessment information. We also supplemented this information using publicly available state and federal data.

Our main outcomes of interest are:

- *Change in the probability of meeting math and ELA standards during the pandemic and*
- *Change in math and ELA test scores during the pandemic.*

Sample

Our sample includes students who were enrolled in public schools between 2015-2019 and 2022 and completed math or ELA assessments in grades 3 through 8 and 10.[33] The sample includes 1.3 million students.

Exhibit 2 illustrates the characteristics of students in our sample. The student composition is not dramatically different from the statewide population of public school students.

Exhibit 2
Student Characteristics

Gender	
Female	49%
Male	51%
Race and ethnicity	
American Indian/Alaska Native	1%
Asian	8%
Black/African American	4%
Hispanic/Latino	24%
White	54%
Native Hawaiian/other Pacific Is.	1%
Two or more races	8%
Primary language	
English	77%
Spanish	14%
Other	9%
Program participation	
Free or reduced-priced meals	48%
Special education	13%
Migrant	1%
Limited English proficiency	10%
Gifted	8%

[33] Our sample does not include students in school years 2020 or 2021. We did not receive assessment data for these years because of test cancellations and postponement. Therefore, we could not examine outcomes in these years.

Research Design

From a statistical perspective, the ideal method for estimating the causal impact of the pandemic on student achievement is to randomly assign students to be affected by the pandemic or not. In this scenario, characteristics between students affected and unaffected would be the same, and we could attribute differences in outcomes to the pandemic.

Since everyone in society was affected by the pandemic, it is impossible to conduct this type of study. As a result, there are challenges to estimating a causal effect. For example, the transition to remote instruction may have changed the composition of students who were tested before and after the pandemic. This could lead to systematic differences between test takers, and these differences may explain changes in test scores.

To address these issues, we use statistical techniques that estimate the change in math and ELA test scores over time and account for time-varying student, school, and neighborhood factors correlated with achievement.[34] We also include school and year fixed effects, which control for time-invariant school factors (e.g., schoolwide culture or policies) and year differences (e.g., economic recessions, policy changes) that may impact test scores.

We also conduct subgroup analyses to see if test score results are different based on student characteristics like gender, race, ethnicity, and income status, as well as school and regional characteristics.

For more information about the data, sample, and research design, refer to Appendix I.

[34] Student controls include race, ethnicity, primary language, income status, special education enrollment, limited English proficiency, and migrant status. School controls include enrollment, high-poverty status, locale, and neighborhood factors like unemployment rate, median household income, and educational attainment.

IV. Academic Achievement During the Pandemic

In this section, we summarize the results and limitations of our test score analysis.

Probability of Meeting Math and ELA Standards

First, we examine the probability of meeting grade-level standards on the Smarter Balanced Assessment (SBA). Before the pandemic, students were more likely to meet ELA standards than math standards. As students progressed in grade, the likelihood of meeting ELA standards increased, and the likelihood of meeting math standards decreased.

Our analysis found a lower probability of meeting standards in the 2022 school year, compared to earlier years. We observed larger declines in math proficiency than ELA proficiency across grades.

Exhibit 3 shows the change in the probability of students meeting standards on math and ELA assessments before and after the pandemic by grade level.[35] The dots represent the average probability of meeting standards in pre-pandemic years, the arrows show the probability of meeting standards in 2022, and we also report the percent change over time. In all preceding graphs, math results are depicted in dark blue and ELA results in light blue.

In 2022, elementary students were 18% and 14% less likely to meet math and ELA standards, respectively, than students before the pandemic; middle school students were 28% and 13% less likely to meet math and ELA standards, respectively; and high school students were 18% and 5% less likely to meet math and ELA standards, respectively.

Math and ELA Test Scores During the Pandemic

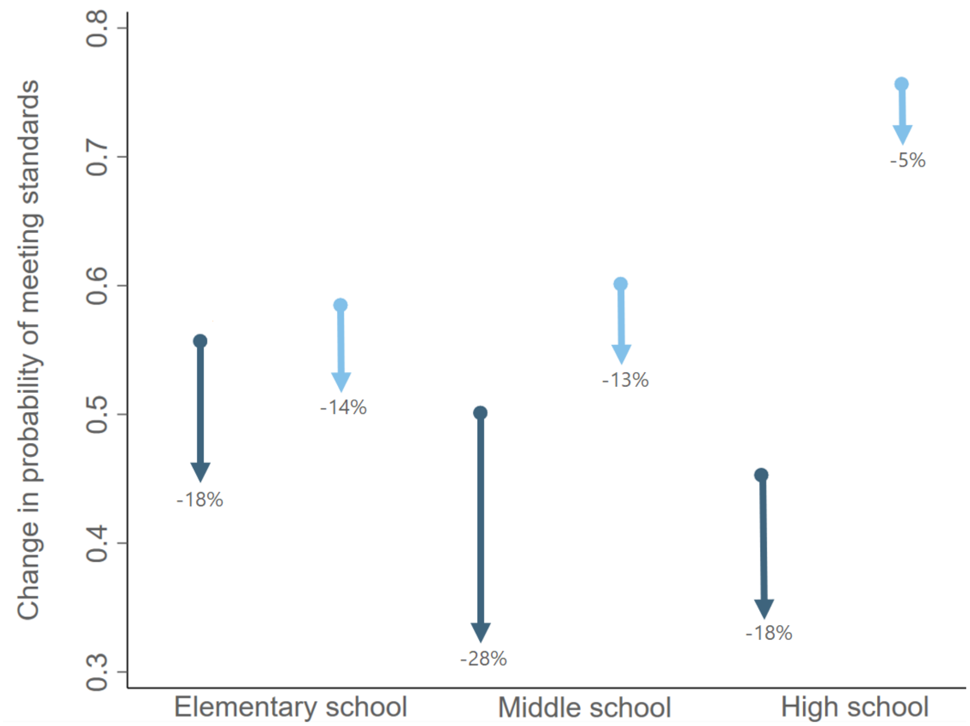
Next, we examine how math and ELA test scores changed during the pandemic. Before the pandemic, math and ELA scores were fairly stable but began to significantly decrease after 2019.

Through our analysis, we found that both math and ELA scores declined over the pandemic. There were greater decreases in math scores than ELA scores; the largest effects were in middle school.

[35] We define elementary grades as K-5, middle school grades as 6-8, and high school grades as 9-12.

Exhibit 3

Change in Probability of Meeting Math and ELA Standards in 2022
(Relative to Probability, Pre-Pandemic)



Notes:
Dark blue indicates math results, light blue indicates ELA results.
Results reflect predicted probabilities estimated from a linear regression with student and school controls. Standard errors are adjusted to account for clustering at the school-level.

In subsequent exhibits, we present results as standard deviation (SD) changes in test scores.[36] We use this standard metric to make comparisons across different grades, subjects, and student groups. We also use this metric later in the report to predict changes to future earnings.

For exemplary purposes, if we report that student test scores in 2022 were 0.15 SD lower than average test scores before the pandemic, this means that for a test with an average score of 2,500 and a standard deviation of 100, test scores decreased by 15 points.[37]

Note that SBA scores are on a continuous scale from approximately 2000 to 3000. Average scores vary depending on the subject (i.e., math or ELA) and grade level. We use a mean of 2500 and a SD of 100 in the example to approximate general score statistics across grades and test subjects.

Exhibit 4 shows our estimated change in math and ELA scores for elementary, middle, and high school students in 2022, compared to pre-pandemic scores.

[36] A standard deviation is a statistical measure that quantifies the amount of variation around a set of data points.

[37] 0.15 x 100 = 15-point decrease in test scores.

Among elementary students, in 2022, average math scores were 0.26 SD lower than math scores before the pandemic, and ELA scores were 0.16 SD lower than ELA scores before the pandemic.[38] Middle school students' average math scores in 2022 were 0.34 SD lower in 2022 than before the pandemic, and ELA scores were 0.19 SD lower than pre-pandemic scores.[39] Among high school students in 2022, math scores were 0.22 SD lower than pre-pandemic scores, and ELA scores were 0.05 SD lower than before the pandemic.[40]

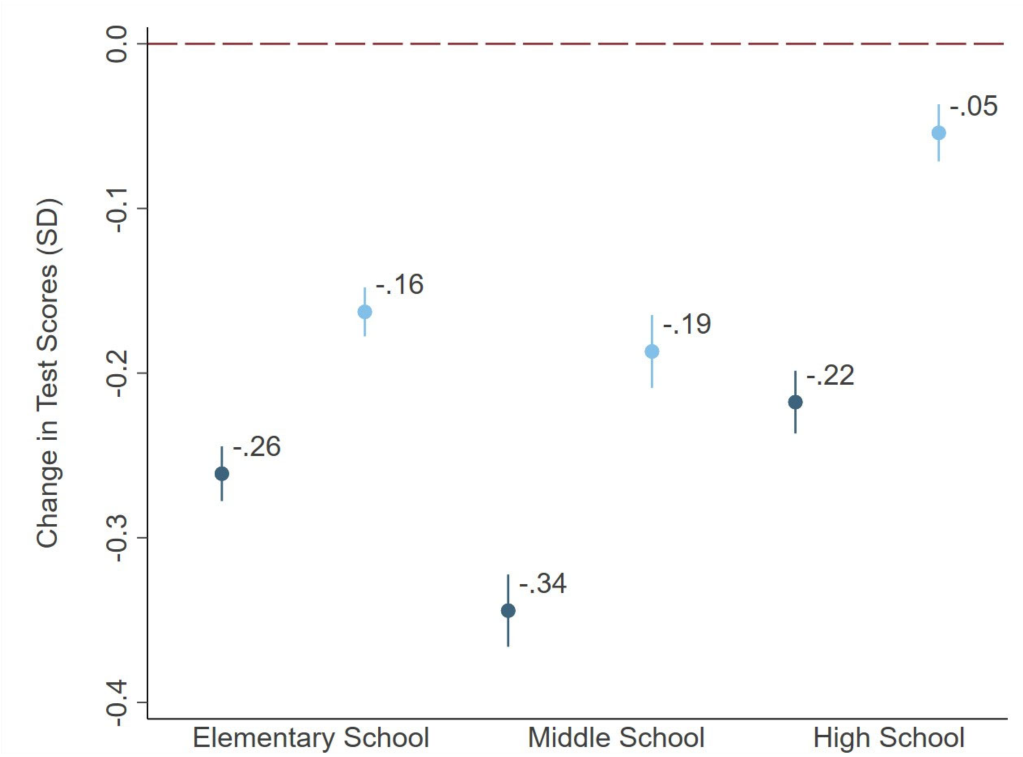
While SD differences between grades and test subjects are noticeable, it is important to note that these effects are relatively small. These effects represent only a 1% decrease in overall test scores (or less).

[38] SD changes are equivalent to a 24-point decrease in math scores and a 16-point decrease in ELA scores.
[39] SD changes are equivalent to a 40-point decrease in math scores and a 18-point decrease in ELA scores.

[40] SD changes are equivalent to a 29-point decrease in math scores and a 4-point decrease in ELA scores.

Exhibit 4

Average Change in Math and ELA Test Scores (SD) in 2022
(Relative to Average Pre-Pandemic Scores)



Notes:
Dark blue indicates math results; light blue indicates ELA results.
Results reflect predicted test score changes estimated from a linear regression with student and school controls.
Standard errors are adjusted to account for clustering at the school-level.

Interpreting Exhibits

The points in Exhibit 4 (and subsequent exhibits) reflect the estimated standard deviation (SD) change in math and ELA test scores in 2022, compared to average test scores before the pandemic (school years 2015 through 2019). For example, average math scores in elementary grades in 2022 were 0.26 standard deviations lower than average math scores pre-pandemic.

The vertical lines extending from each point represent 95% confidence intervals, a range that likely includes the true effect. Intervals that cross the red dashed line indicate the estimate is not statistically significantly different from zero.

Changes in Test Scores by Student, School, and Regional Characteristics

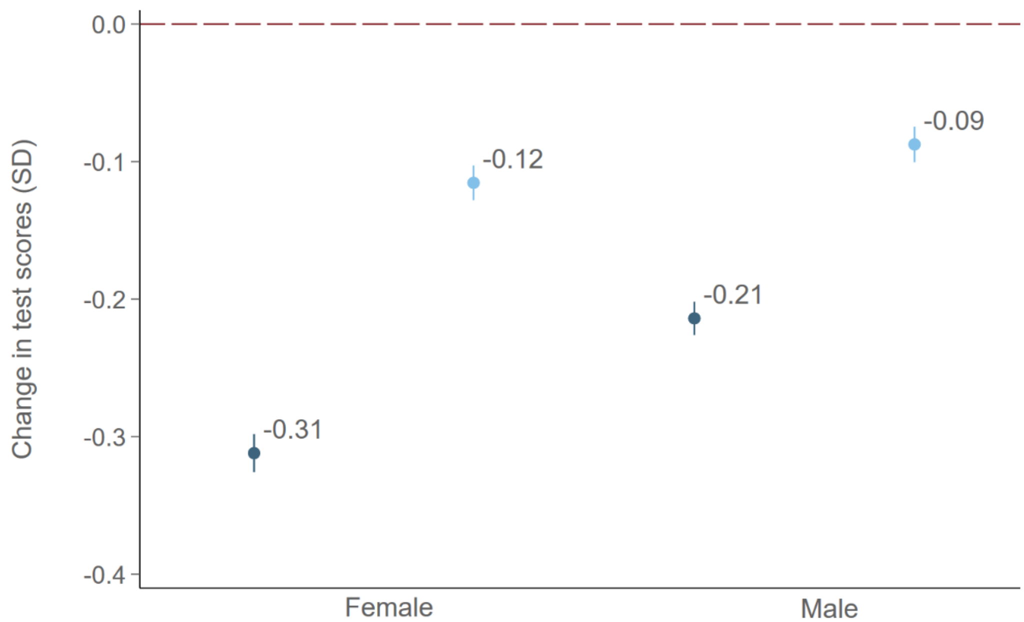
We also examined changes to test scores by student, school, and regional characteristics. In 2022, test scores among female students, students of color, Hispanic/Latino students, low-income students, and students in more populous areas fell more than their peers.

Effects by Gender

Math and ELA test scores decreased for both male and female students during the pandemic, but female students experienced larger declines in both subjects than male students. Average math scores for female students decreased by 0.31 SD compared to 0.20 SD for male students. ELA scores fell 0.12 SD compared to 0.09 SD for male students (Exhibit 5).[41]

Exhibit 5

Average Change in Math and ELA Test Scores (SD) in 2022, by Gender
(Relative to Average Pre-Pandemic Scores)



Notes:
Dark blue indicates math results; light blue indicates ELA results.
Results reflect predicted test score changes estimated from a linear regression with student and school controls. Standard errors are adjusted to account for clustering at the school-level.

[41] When examining variation across grades, we found the largest effect on math scores in middle school and similar effects on ELA scores in elementary and middle school. Before the pandemic, female students' average math and ELA test scores were slightly higher than male students' scores. After the pandemic, male students' average math scores slightly surpassed female students' scores.

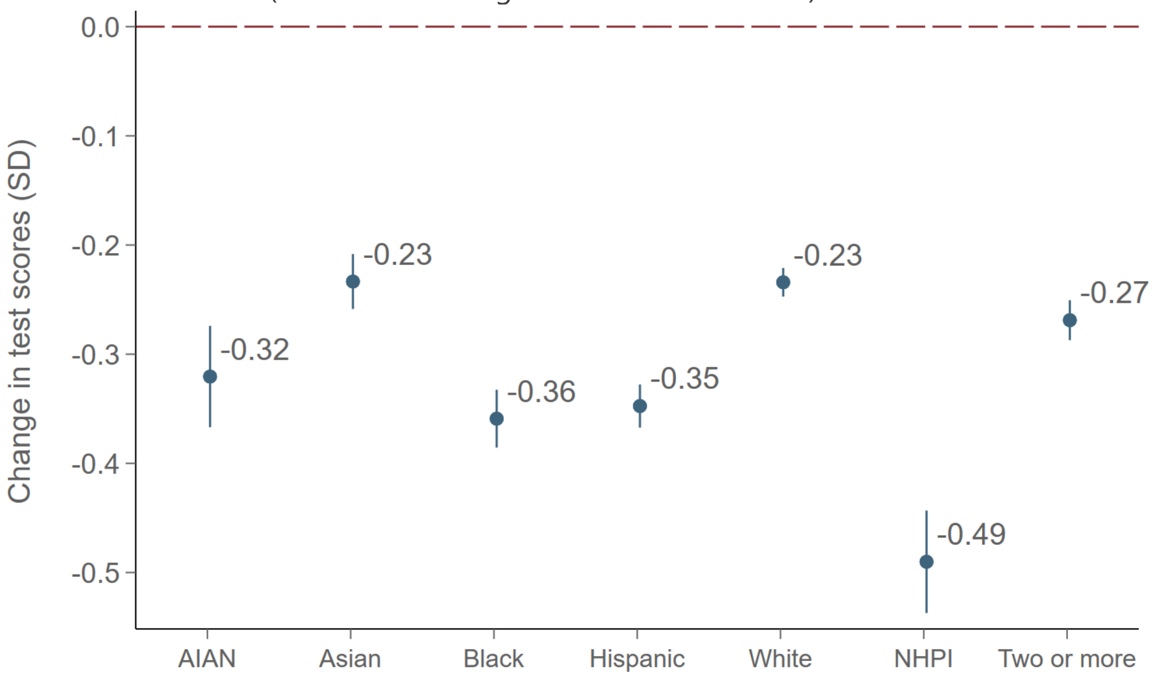
Effects by Race & Ethnicity

We examined the change in test scores by students' race and ethnicity. All student populations experienced declines in test scores, and math scores fell more than ELA scores (Exhibits 6 and 7). However, some populations were negatively affected more than others.[42]

Among students who identified as Native Hawaiian or other Pacific Islander (NHPI) in 2022, math and ELA scores were 0.49 SD and 0.24 SD lower than average pre-pandemic scores, respectively.

Exhibit 6

Average Change in Math Test Scores (SD) in 2022, by Race/Ethnicity
(Relative to Average Pre-Pandemic Scores)

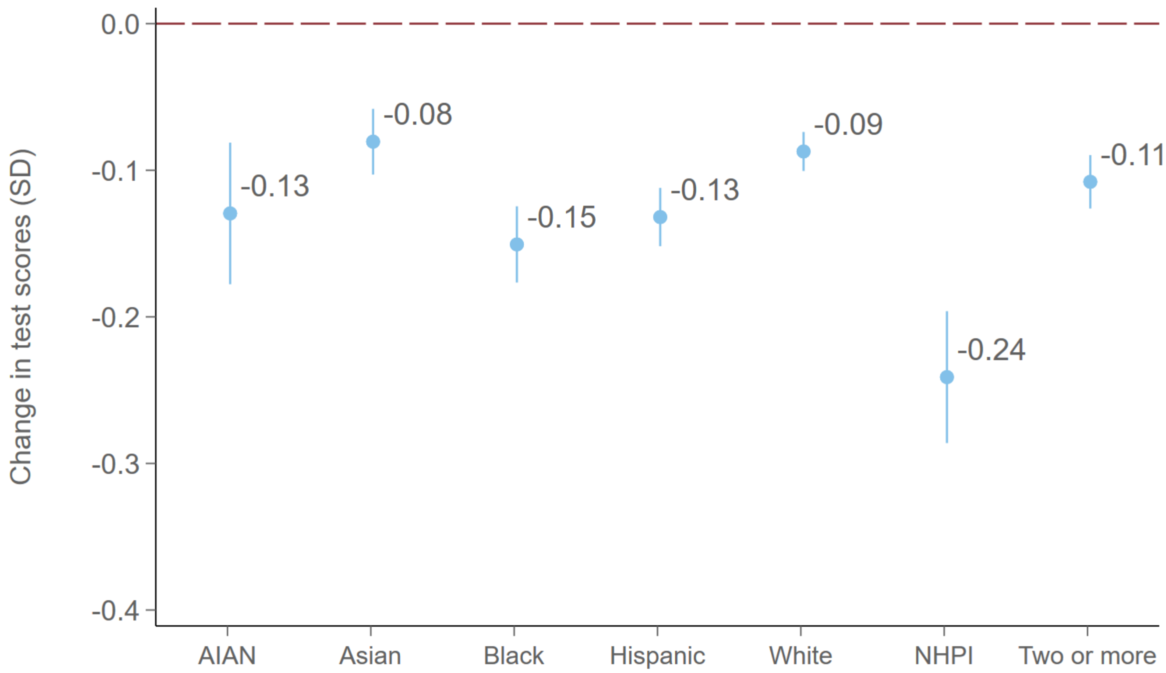


Notes:
Dark blue indicates math results.
Results reflect predicted test score changes estimated from a linear regression with student and school controls. Standard errors are adjusted to account for clustering at the school-level.
AIAN – American Indian/Alaska Native.
NHPI – Native Hawaiian or other Pacific Islander.

[42] Generally, we observed the largest effects in middle school with one exception. Among Native Hawaiian or Other Pacific Islander students, ELA scores fell the most in elementary grades.

Exhibit 7

Average Change in ELA Test Scores (SD) in 2022, by Race/Ethnicity (Relative to Average Pre-Pandemic Scores)



Notes:
Dark blue indicates math results.
Results reflect predicted test score changes estimated from a linear regression with student and school controls. Standard errors are adjusted to account for clustering at the school-level.
AIAN – American Indian/Alaska Native.
NHPI – Native Hawaiian or other Pacific Islander.

Among students who identified as American Indian or Alaska Native (AIAN), Black or African American, Hispanic/Latino, or two or more races, average math and ELA scores were about 0.36 SD and 0.15 SD lower than average pre-pandemic scores, respectively.[43]

Among students who identified as Asian or White, average math and ELA scores were 0.23 SD and about 0.08 SD lower in 2022 (respectively) than average scores before the pandemic.

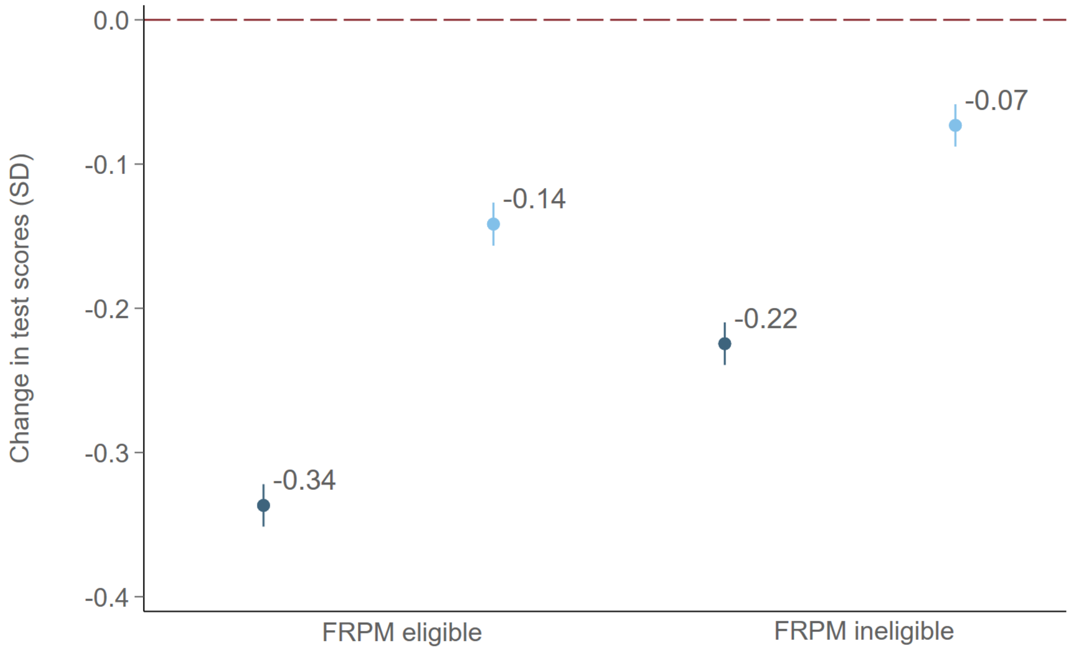
Effects by Income Status

Next, we examine test score changes by income status. We do not have a direct measure of family income in the data. We approximate whether a student is low-income or not based on their eligibility to receive free or reduced-priced meals (FRPM).

[43] ELA effects between Black, NHPI, and AIAN students are not statistically significantly different.

Exhibit 8

Average Change in Math and ELA Test Scores (SD) in 2022, by Income Status (Relative to Average Pre-Pandemic Scores)



Notes:
Dark blue indicates math results; light blue indicates ELA results.
Results reflect predicted test score changes estimated from a linear regression with student and school controls.
Standard errors are adjusted to account for clustering at the school-level.

Overall, students eligible for FRPM experienced larger declines in test scores in 2022 than ineligible students.[44] Eligible students experienced 0.34 SD and 0.14 SD decreases in math and ELA scores, respectively, compared to ineligible students who experienced 0.22 and 0.07 SD declines in scores (Exhibit 8).[45]

We also examined test scores by school poverty status and found similar effects as the ones based on student income status.[46]

[44] Before and after the pandemic, average math and ELA test scores among FRPM eligible students were lower than average FRPM ineligible students.

[45] When examining variation across grade levels, we found the largest effects in elementary grades.

[46] We consider high-poverty schools as defined in RCW 28A.150.260.

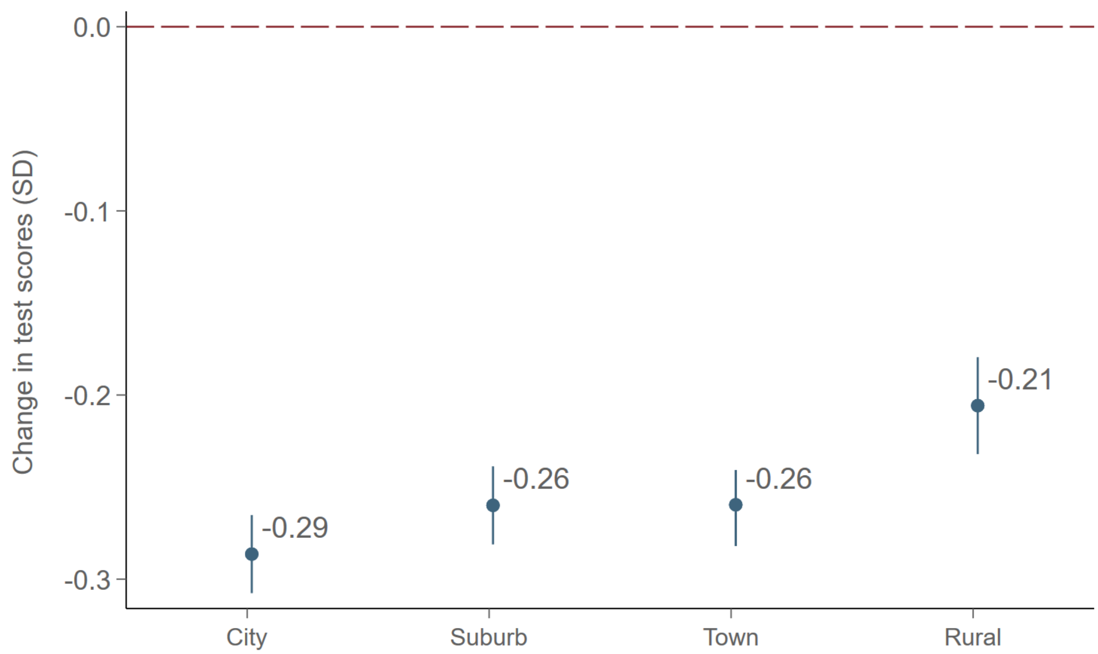
Effects by School Locale

Finally, we examine test scores by the geographic location of schools.[47] Similar to earlier findings, math scores decreased more than ELA scores across all locales.

Students in schools in more populous areas, including cities and suburbs, experienced slightly greater declines in math scores than students in rural areas and towns (Exhibit 9). We observed similar patterns among ELA scores (Exhibit 10).[48] We also illustrate test score effects by school district in Appendix I.

Exhibit 9

Average Change in Math Test Scores (SD) in 2022, by School Locale
(Relative to Average Pre-Pandemic Scores)



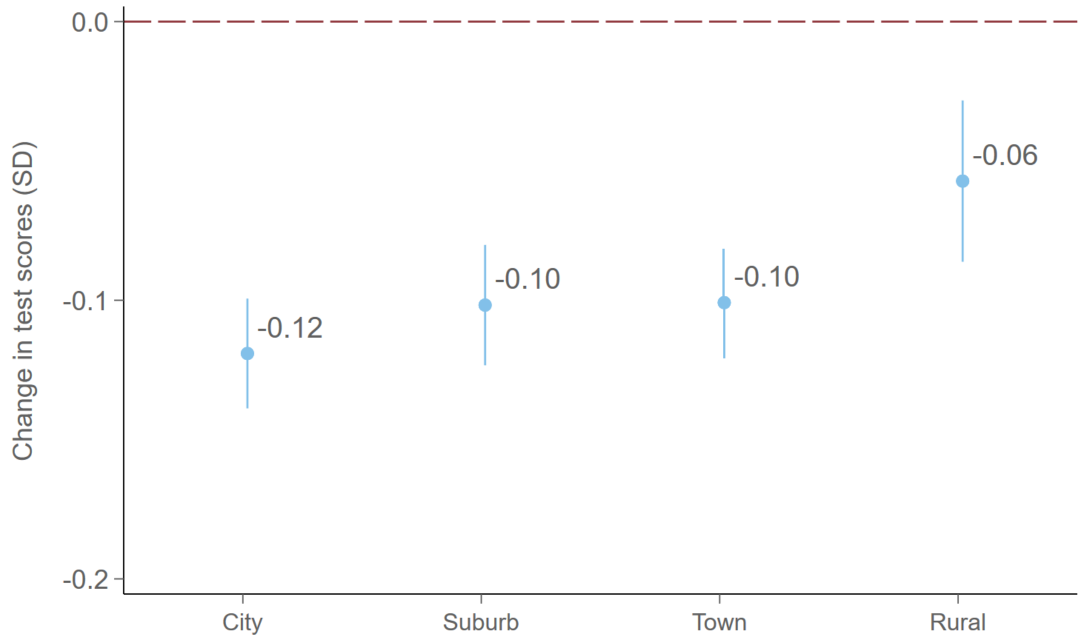
Notes:
Dark blue indicates math results.
Results reflect predicted test score changes estimated from a linear regression with student and school controls. Standard errors are adjusted to account for clustering at the school-level.

[47] We use NCES's locale indicators to define geographic areas.

[48] Math and ELA test score effects are not statistically significantly different between cities and suburbs. They are not significantly different between towns and rural areas.

Exhibit 10

Average Change in Math Test Scores (SD) in 2022, by School Locale
(Relative to Average Pre-Pandemic Scores)



Notes:
Light blue indicates ELA results.
Results reflect predicted test score changes estimated from a linear regression with student and school controls. Standard errors are adjusted to account for clustering at the school-level.

Limitations

We control for student and school characteristics that correlate with academic achievement but cannot fully account for all observed and unobserved factors. For example, there may be differences between students tested before and after the pandemic that explain changes in test scores other than the pandemic. We run additional analyses to explore these issues and include results in Appendix I.

Also, we cannot isolate the mechanisms that influence test scores in our analysis. The pandemic had widespread societal impacts, and we are unable to determine what proportion of our effect is due to remote instruction, economic impacts on families, or health and mental health effects, all of which influence student achievement to varying degrees. Our estimates likely include a combination of these factors.

Finally, standardized test scores are only one measure of academic achievement and do not capture the comprehensive learning and development that students experience in a school year. We focus on test scores because they are consistently collected by OSPI and offer a standard metric to compare over time. Standardized test scores are also a parameter in WSIPP's benefit-cost model that we use to estimate future earnings.

V. Predicting Changes to Students' Future Earnings

Next, we describe how we use estimates from our test score analysis to predict changes to future earnings. We also report the results and limitations of this analysis.

After estimating changes to math and ELA test scores during the pandemic, we input these results in WSIPP's benefit-cost model to answer the question:

- How do test scores changes influence future earnings?

WSIPP's benefit-cost model (BCM) estimates the long-term benefits associated with interventions, programs, or policies and compares them to program costs. Typically, we use the model to determine if the benefits of a program outweigh its costs. For this project, we use the model to estimate how changes to test scores during the pandemic may influence changes to students' future earnings.

We also use parameters in the BCM to analyze how different test score recovery scenarios may influence earning trajectories.

Finally, we use measures of uncertainty in our estimates to run sensitivity tests. We do this to determine the range of earnings loss among students who were directly affected by the pandemic compared to students who were not in school during the pandemic.

See Appendix II for more information about WSIPP's benefit-cost model, how we used it for this analysis, sensitivity results, and limitations.

Results

First, before including estimates from our test score analysis into WSIPP's BCM, we combine math and ELA results into a single "test score effect." We do this because the BCM does not estimate different earnings trajectories based on test subjects.

Looking across all grades, we estimate that 2022 average test scores were 0.20 SD lower than average pre-pandemic scores. We estimate this decline is associated with a $32,000 decrease in lifetime earnings per student on average (the equivalent of a 1.3% decrease in labor market earnings). In other words, students directly affected by the pandemic and tested in 2022 may earn $32,000 less than peers who were tested between 2015 and 2019 and unaffected by the pandemic.

Since we observe variation in test score effects across grades, students in middle school during the pandemic may experience a greater earnings decrease than those in elementary or high school grades during the pandemic.

Exhibit 11 shows the decrease in test scores by grade level and the predicted decrease in earnings per student.

Among elementary students tested in 2022, a 0.18 SD decline in test scores is associated with a $23,000 decrease in earnings per student.

Among middle school students tested in 2022, a 0.25 SD decline in scores is associated with a predicted $42,700 decrease in future earnings per student.

A 0.12 SD decline in test scores is associated with a $26,700 decrease in projected earnings among high school students.

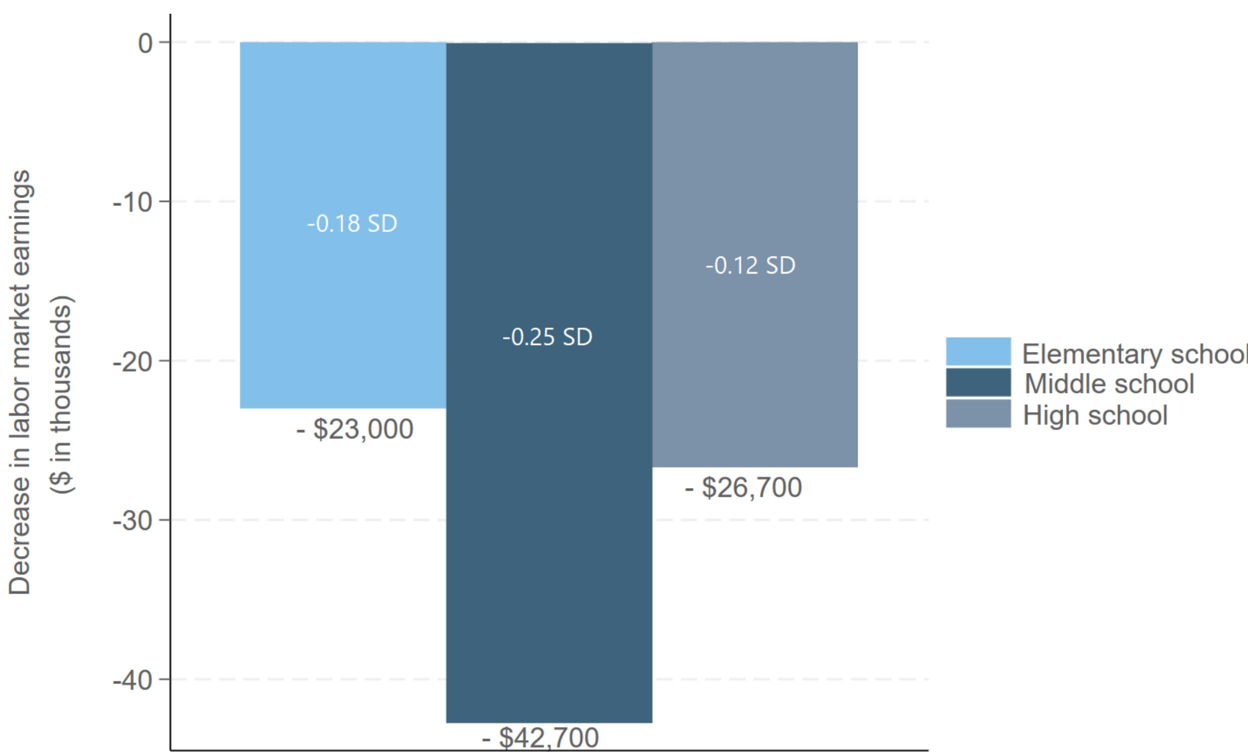
Note that while there is a larger decline in test scores among elementary students than high school students, the predicted impact on earnings is similar. This is because we assume test scores will rebound to some extent in the years after the pandemic.

As a result, students in elementary grades during the pandemic have more years to recover academically before they graduate than those in high school. Because of this, we expect test score effects between elementary and high school students to be similar by the time they enter the workforce, the point at which we link test scores to future earnings.[49]

The figures in Exhibit 11 reflect average estimates. Through sensitivity tests, we found a large range in expected earnings loss. See Appendix II for more information.

Exhibit 11

Predicted Change in Future Labor Market Earnings



[49] See Appendix II for more information.

Limitations

Our estimates of future earnings represent preliminary predictions, given patterns we have observed so far. There is currently no definitive research on the long-term economic effects of the pandemic. We need more years of data to understand what the pandemic's long-term impact will be on outcomes like employment and earnings.

Further, to conduct this analysis, we use WSIPP's BCM model differently than our normal approach. Because of this, the underlying assumptions of the BCM may or may not apply to the current analysis.

One difference is that we typically use the model to compare benefits and costs between program participants and non-participants. However, everyone in the world was affected by the pandemic. In other words, we cannot compare test scores between students affected by the pandemic and those unaffected during the same period.

The inputs used to create this piece of the benefit-cost model were derived from studies analyzing interventions, not system-wide shocks. This current modeling may, therefore, not apply to the current situation.[50]

Further, the BCM incorporates assumptions based on pre-pandemic factors that may not apply in a post-pandemic world. For example, based on research conducted before the pandemic, the model includes a parameter that estimates how a change in test scores predicts a change in labor market earnings. If test scores have a smaller effect on earnings after the pandemic, we may be overestimating the decrease in earnings.

Also, our predicted changes in earnings assume that student academic achievement will recover in the coming years. Research on natural disasters and early COVID-19 studies indicate that test scores rebound following large shocks to the education system.[51] We use existing parameters in the BCM to model test score recovery. However, these parameters are based on research on the fade-out of changes in test scores over time, which may not accurately reflect how student achievement will rebound post pandemic.

See Appendix II for more information about limitations and sensitivity analysis results.

[50] For example, if what matters for future earnings is your ranking relative to your peers, earnings will decrease through participating in an intervention that decreases your relative ranking. However, in this scenario, earnings would be unchanged in a system-wide decrease in test scores where no one's relative ranking changes.

[51] Sacerdote, B. (2012). When the Saints Go Marching Out: Long-Term Outcomes for Student Evacuees from Hurricanes Katrina and Rita. *American Economic Journal: Applied Economics, 4*(1); Kuhfeld & Lewis (2022) ; Halloran et al. (2023).

VI. Recovery Interventions

In this section, we describe the use of WSIPP's meta-analytic findings to explore how some interventions may help students recover academically. We also describe several key limitations.

We answer the following research question:

- What interventions may help students recover academically, and by how much?

WSIPP has developed a standard approach for conducting meta-analysis, a statistical technique used to estimate a program's average effect on outcomes.[52]

To date, WSIPP researchers have analyzed nearly 80 early education and K-12 programs and practices.[53] Some topics we have examined are interventions that school districts may use to help students recover after the pandemic. These interventions include:

- Tutoring
- Academically-focused summer school programs, and
- Double-dose classes

We compare program effects to estimates from our test score analysis and describe how these interventions may bolster test scores in the coming years.

See Appendix III for more information about WSIPP's approach to meta-analysis.

Results

In previous meta-analyses, WSIPP researchers found that tutoring models, summer school, and double-dose classes increase test scores between 0.03 and 0.39 SD, on average.

Though not an exhaustive list of recovery interventions, these are common approaches that school districts in Washington and other states have considered implementing to support learning post pandemic.[54] These programs may help offset the negative effects on test scores described earlier in the report.

Tutoring Models

WSIPP researchers have found that tutoring in elementary schools can increase test scores between 0.03 and 0.39 SD.[55] The wide range of effects is due to the various tutoring models we examined. In general, we found that models that employ teachers, paraeducators, or trained adult volunteers as tutors, use structured curricula, and have tutors and students meet one-on-one or in small groups multiple times per week yield the largest positive effects on test scores. These models are sometimes referred to as high-dosage tutoring.

[52] WSIPP. *Estimating program effects using effect sizes.* Olympia, WA: Author.

[53] Washington State Institute for Public Policy. (2023, August). *Pre-K to 12 education benefit-cost results.* Olympia, WA: Author.

[54] Carbonari, M.V., Davison, M., Dewey, D., Dizon-Ross, E., Goldhaber, D. Hashim, A. . . . Staiger, D.O. (2022). *The challenges of implementing academic COVID recovery interventions: evidence from the road to recovery project*. National Center for Analysis of Longitudinal Data in Education Research. ; JLARC (2023) ; FutureEd (2021). *With an influx of covid relieve funds, states spend on schools.*

[55] Washington State Institute for Public Policy. (2023, August). *Pre-K to 12 education benefit-cost results.* Olympia, WA: Author.

We found that tutoring models that employ untrained volunteers use unstructured curricula and meet with students irregularly have smaller effects.

The populations included in our tutoring program analyses include elementary students struggling to meet grade-level math or reading standards.

Summer Learning Programs

WSIPP researchers also examined the impact of academically-focused summer school programs and found that, on average, students who participated in these programs experienced an increase in test scores by 0.06 SD, compared to students who did not participate.[56]

This analysis includes school- and community-provided programs that serve elementary and middle school students who struggle to meet math or reading standards.

Double-Dose Classes

WSIPP researchers also reviewed the impact of double-dose classes and found that they also increase student test scores. Double-dosing is a practice in which middle or high-school students who struggle in math or reading enroll in multiple math or reading classes to increase their instructional time with the subject. We have estimated that students in double-dose classes experience a 0.09 SD increase in test scores compared to students not enrolled in extra classes.[57]

Exhibit 12 compares these intervention effects to the average 0.20 SD test score decline we estimated earlier.

If the intervention effects we estimated in meta-analyses remain the same when implemented in a larger student population, high-dosage and peer tutoring models could fully offset the decline in test scores during the pandemic. Parent tutoring could offset most of the decline in test scores. And double-dose classes, academically-focused summer school models, and unstructured tutoring by adults could help recover some achievement, though they would have a smaller impact.

As mentioned earlier, students of color, Hispanic and Latino students, and low-income students were disproportionately affected during the pandemic. Though achievement among White, Asian, and economically advantaged students was negatively affected too, generally, test scores among these groups continue to meet grade-level standards post-pandemic, while scores among historically disadvantaged students do not.[58]

It is unlikely that school districts will be able to administer recovery efforts to every student affected by the pandemic. From a resource allocation and equity perspective, districts may target interventions for students who have been disproportionately affected by the pandemic. If they do, we estimate that interventions would need to increase test scores among these students by about 0.25 SD to recover to pre-pandemic levels.[59]

[56] Washington State Institute for Public Policy. (2023, August). *Summer learning programs: academically focused benefit-cost results*. Olympia, WA: Author.

[57] Washington State Institute for Public Policy. (2023, August). *Double-dose classes benefit-cost results*. Olympia, WA: Author.

[58] Math and ELA scores among Asian students exceeded proficiency expectations before and after the pandemic in all grades. Among White students, average scores met standards before and after the pandemic, except students tested in math in grades 5 through 8 in 2022.

[59] For comparison, interventions would need to increase White and Asian students' test scores 0.16 SD to recover achievement to pre-pandemic levels.

Further, interventions would need to increase scores among historically disadvantaged student groups by about 0.50 SD to close achievement gaps between these groups and historically advantaged students.

Limitations

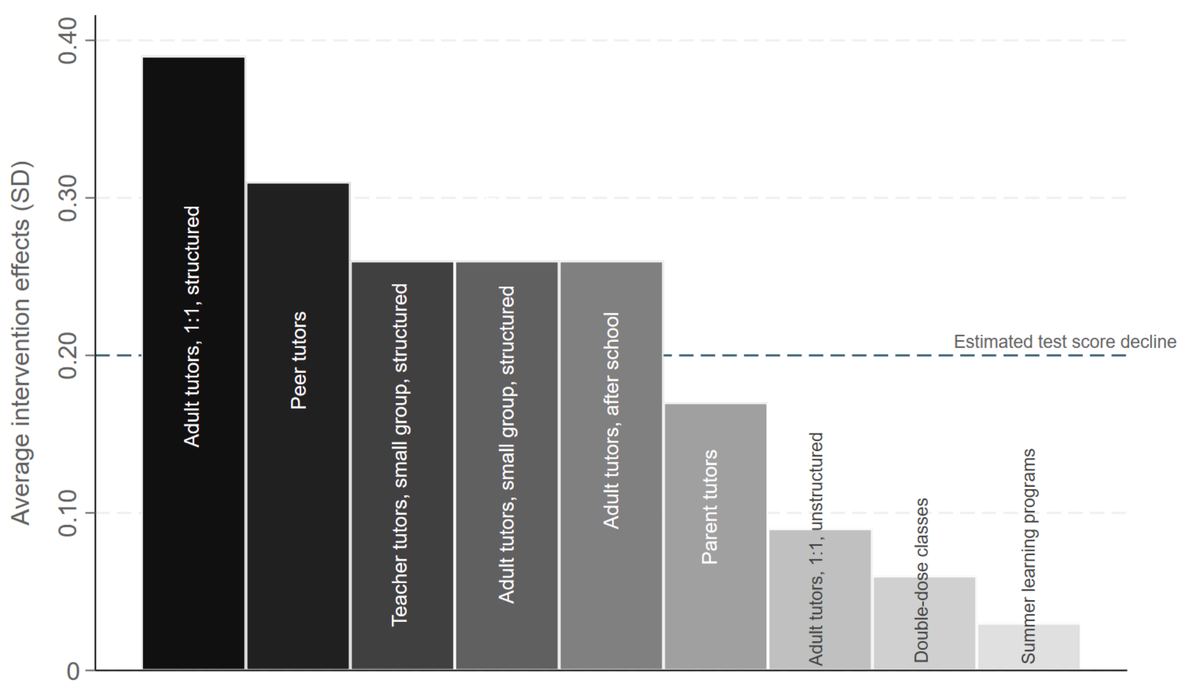
We cannot definitively say whether tutoring, summer school, and double-dose class interventions will recover the test score effects we've observed. When conducting meta-analyses on these interventions, the study populations were smaller than the sample of students in our test score analysis. Therefore, these interventions may not have the same effects if scaled to a larger and more diverse student population.

Further, the meta-analyses we conducted include program effects that were estimated before the pandemic. As a result, tutoring, summer school, and double-dose interventions may not have the same impact on student achievement if implemented in a post-pandemic period.

Finally, we acknowledge there are many other interventions besides the few listed here that school districts will consider when focusing on academic and non-academic recovery efforts post-pandemic.

Exhibit 12

Average Intervention Effects (SD) Compared to Estimated Decline (SD) in Test Scores



VII. Conclusion

In this section, we summarize the key takeaways and limitations of our analyses.

Key Takeaways

The goal of this report was to examine how student achievement was affected during the pandemic, explore long-term effects on earnings, and examine several recovery interventions.

We first examined how student achievement changed during the pandemic. We found that student test scores decreased across all grades, content areas, and student groups. Overall, we estimated that average test scores in 2022 were 0.20 SD lower than average test scores before the pandemic.

We also observed variations in test score effects. Specifically, we found the largest decline in test scores in middle school grades, though scores in elementary and high school grades fell noticeably, too. We also found larger decreases in math scores than ELA scores. Finally, we estimated greater declines among female students than male students, among low-income students compared to economically advantaged students, among students of color and Hispanic/Latino students compared to White and Asian students, and among students in populous areas of the state compared to students in rural areas.

Next, we used WSIPP's benefit-cost model to predict how test score changes influence future earnings. On average, we predict that students tested in 2022, and therefore directly impacted by the pandemic, may earn about $32,000 less over their lifetimes than students tested before the pandemic (about a 1.3% difference).

Finally, we reviewed past meta-analyses on tutoring, academically-focused summer school, and double-dose programs and compared their effects to our estimated test score effects. These interventions increase test scores between 0.03 and 0.39 SD on average.

We observe that high-dosage tutoring models could mostly offset the decrease in test scores that occurred during the pandemic. Other interventions like academically-focused summer school and double-dose classes could also recover some of the test score declines.

While all student populations were negatively affected over the course of the pandemic, average test scores among White, Asian, and economically advantaged students continue to meet grade-level standards, while scores among students of color, Hispanic/Latino students, and low-income students do not meet proficiency levels, on average. Students who have been historically disadvantaged may need targeted recovery support to help make up for learning reductions; tutoring, academically-focused summer school, and double-dose classes could be effective strategies.

Limitations

It is important to consider several key limitations when interpreting our results.

First, we cannot rule out the possibility that other factors, besides the pandemic, may be driving our estimated test score effects. However, we did run additional analyses and found results to be robust across different statistical models and specifications.

We also cannot conclude that test score effects are influenced by a specific factor or set of factors. Our estimates likely encompass widespread pandemic effects, including economic, health, and social impacts, as well as educational system impacts that influence academic performance.

We further acknowledge that test scores are a limited measure of achievement and that the pandemic had effects on non-academic student outcomes, too. We focus on test scores because they are a consistent metric to analyze over time and one we can use to predict earnings.

In terms of our earnings estimates, the results are speculative. We need years of data to fully understand the pandemic's long-term effects on employment and earning outcomes. Further, our estimates rely on a set of assumptions in WSIPP's benefit-cost model developed using information before the pandemic and for scenarios where we compare effects between program participants and non-participants, not society-wide effects. As a result, these assumptions may or may not apply to our current analysis.

Finally, our meta-analytic results were conducted before the pandemic and illustrate the effect that certain tutoring models, summer school programs, and double-dose classes have on student test scores, on average. These interventions may not have the same impact on test scores if applied in a post-pandemic setting or if scaled to a larger student population.

Acknowledgments

The authors would like to thank Student Information staff at the Office of Superintendent of Public Instruction for providing us with the data necessary to complete this report. We would also like to thank WSIPP staff Rebecca Goodvin, Chasya Hoagland, Amani Rashid, and Nathan Adams for their feedback on analyses and earlier drafts of this report. Finally, we thank Heather Grob, WSIPP's Benefit-Cost Lead, for her assistance with the benefit-cost model and estimating future earnings for this report.

Appendices

Student Achievement and the Pandemic: Analysis of Test Scores, Earnings, and Recovery Interventions

I. Test Score Analysis

Data Sources

We received administrative data from OSPI's longitudinal data system (i.e., the Comprehensive Education Data and Research System). Most of the data described below covered school years 2015-2022:

- **Student demographics and enrollments**—files include student-level information about race, ethnicity, gender, and primary language, as well as school and district enrollments over time.
- **Program participation**—files include program participation or eligibility information, including eligibility for Free or Reduced-Price meals, enrollments in special education, the gifted program, the Learning Assistance Program, and English language development services.
- **Absence and discipline information**—files contain information about absences, disciplinary events, and the type of disciplinary action a student received (e.g., in-school suspension, out-of-school suspension, expulsion, no intervention).
- **Spring summative assessments**—files contain assessment information like the Smarter Balanced Assessment (SBA), the Access to Instruction & Measurement (AIM), and the Washington Comprehensive Assessment of Science (WCAS). Data includes assessment scores, grades tested, subjects tested, and accommodations. Due to cancellations and postponements during the pandemic, we did not receive assessment data for the 2020 or 2021 school years.

OSPI sent us unidentifiable data, and we linked files together using anonymous student IDs, school year, and school information. We supplemented OSPI data with the following publicly available information:

- **School and district information**—we included information about school poverty status, enrollments, and geographic location using data from the Department of Education's Common Core of Data (CCD) warehouse and OSPI's data portal for public use.[60]
- **School Neighborhood information**—Using the American Community Survey, we linked census-tract level data to school buildings to include neighborhood information like education attainment levels, the proportion of individuals unemployed, and median household income.[61]

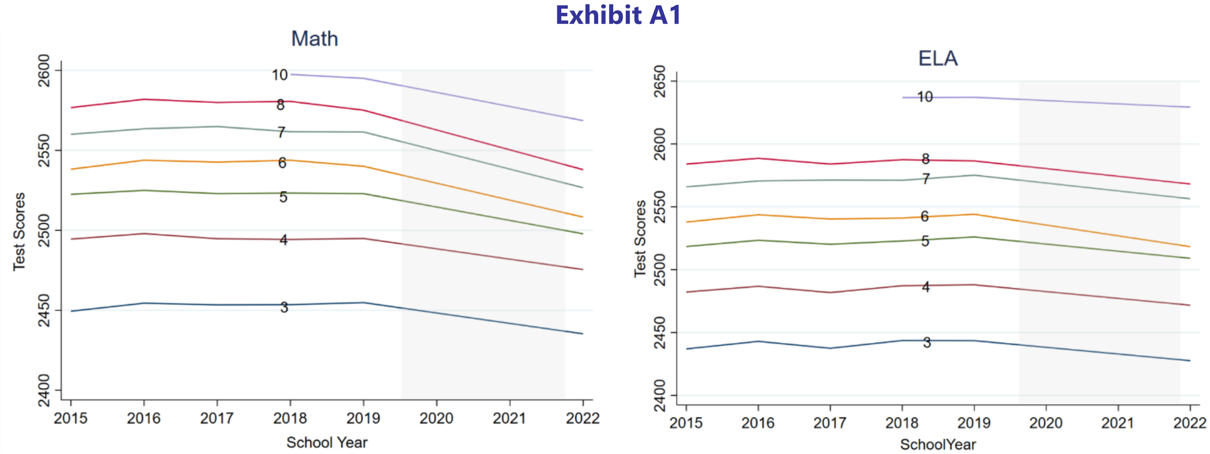
[60] National Center for Education Statistics. *Common Core of Data.*; OSPI. *Data for Public Use.*
[61] U.S. Census Bureau. *American Community Survey Data.*

Outcomes

In Washington, the SBA is used to assess students' skills and knowledge in mathematics and ELA in grades 3 through 8 and 10.[62] Test scores fall on a continuous scale ranging from approximately 2000 to 3000, depending on grade level and subject.[63] Cut scores identify proficiency standards, which also vary depending on subject and grade.[64]

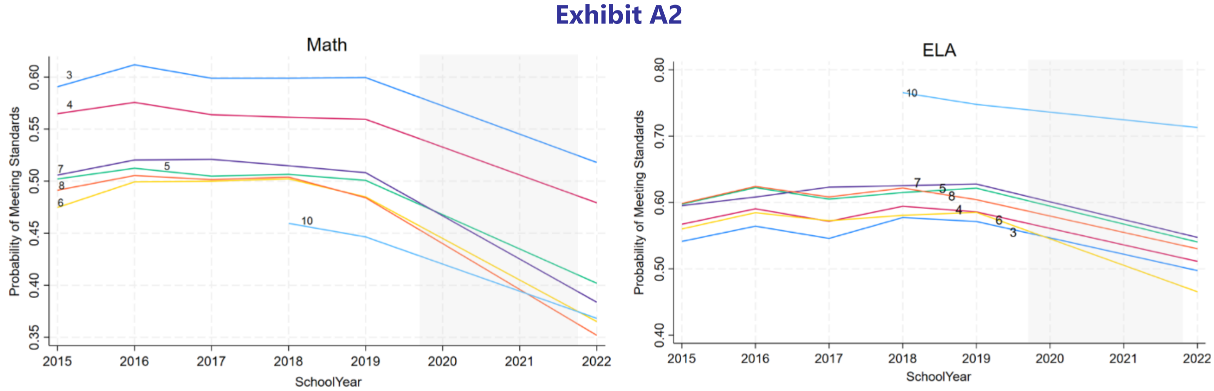
Our main outcome of interest is the change in average math or ELA scores in 2022 relative to pre-pandemic scores. Exhibit A1 shows the trend in average math and ELA scores over time. The grades students are tested in are shown on trend lines. The shaded area shows the pandemic period. We did not receive test data for school years 2020 or 2021.

Exhibit A1



We also examine the change in the probability of students meeting standards on math or ELA tests in 2022 compared to the probability of meeting standards before the pandemic. Exhibit A2 shows the probability of students meeting math and ELA standards over time.

Exhibit A2



[62] Prior to 2017, students were tested in 11th grade instead of 10th grade.
[63] OSPI. *Scale Scores State Assessments.*
[64] Smarter Balanced*.* *Reporting scores.*

Analytic Sample Construction

We processed data files separately and made restrictions to create a longitudinal student-by-school-by-year dataset. Key restrictions are noted below:

- In OSPI enrollment files, about 30% of students moved between schools in a year. For students who switched schools in a year, we kept records associated with their first school enrollment. We also kept records associated with a student's primary school. Finally, we excluded all non-public school enrollment records (see results from sensitivity analyses on page 41. These restrictions do not impact our overall results)
- In the assessment data files, we kept records associated with the SBA and excluded all other test types. We only kept records for students who completed tests and excluded students from the data if they did not complete math or ELA tests. We also excluded records with missing scores.

For all other files, we cleaned data and removed duplicate records. After processing the data, we merged files to create a longitudinal dataset. Our final analytic sample included 3.29 million observations, with 1.3 million unique students in school years 2015-2019 and 2022.

Exhibit A3 shows descriptive statistics, comparing student, school, and school neighborhood characteristics between the overall data sample we received from OSPI and our final analytic sample. Student and school compositions between the two samples are similar.

Exhibit A3

Student and School Characteristics: Student Population vs. Analytic Sample

Student characteristics	Population	Analytic sample
% Female	0.48 (0.50)	0.49 (0.50)
% American Indian/Alaska Native	0.01 (0.12)	0.01 (0.11)
% Asian	0.08 (0.27)	0.08 (0.27)
% Black/African American	0.05 (0.21)	0.04 (0.20)
% Hispanic/Latinx	0.23 (0.42)	0.24 (0.43)
% White	0.54 (0.50)	0.54 (0.50)
% Native Hawaiian/other Pacific Islander	0.01 (0.11)	0.01 (0.10)
% Two or more races	0.08 (0.27)	0.08 (0.27)
% Primary language is English	0.78 (0.41)	0.77 (0.42)
% Primary language is Spanish	0.13 (0.34)	0.14 (0.35)
% Primary language is other	0.09 (0.28)	0.09 (0.28)
% Free or reduced-priced meals	0.47 (0.50)	0.48 (0.50)
% Special education	0.15 (0.36)	0.13 (0.33)
% Limited English proficiency	0.11 (0.32)	0.10 (0.30)
% Migrant	0.01 (0.11)	0.01 (0.11)
Total observations	10,401,017	3,296,009

Note:
Standard deviations reported in parentheses.

Exhibit A3 Continued

Student and School Characteristics: Student Population vs. Analytic Sample

School characteristics	Population	Analytic sample
Average school enrollment	790.08	755.88
	(538.13)	(454.48)
% High poverty schools	0.45	0.46
	(0.50)	(0.50)
% Schools in cities	0.34	0.32
	(0.47)	(0.47)
% Schools in suburbs	0.41	0.43
	(0.49)	(0.50)
% School in towns	0.12	0.12
	(0.33)	(0.33)
% Schools in rural areas	0.12	0.13
	(0.33)	(0.33)
School neighborhood characteristics		
% Pop. unemployed	5.89	6.13
	(3.27)	(3.25)
Median household income	$ 85,618	$ 84,523
	(35,610)	(34,512)
% Pop. with high school diploma	24.31	24.17
	(9.19)	(8.96)
% Pop. with associate degree	10.35	10.34
	(3.70)	(3.56)
% Pop. with bachelor's degree	19.70	19.60
	(10.56)	(10.50)
Total observations	10,401,017	3,296,009

Notes:
Standard deviations reported in parentheses.

Methodology of Test Score Analysis

We use statistical techniques to estimate the change in math and ELA test scores in pre- and post-pandemic periods, accounting for student and school controls and fixed effects. We model the relationship between the pandemic and test scores with the following model:

$$Score_{ij} = \beta_0 + \beta_1(Post_i) + \beta_2(X_{ij}) + \beta_3(S_{js}) + \alpha_s + \gamma_j + \varepsilon_{ij}$$

$Score_{ij}$ Math or ELA scale score for student (i) in school year (j). We estimate separate models for math and ELA subjects and grades 3 through 8 and 10.

$Post_i$ Indicates if the student is enrolled in a school year that is pre- or post-pandemic. The only school year that is post-pandemic is 2022. Our parameter of interest is β_1, and is interpreted as the change in math/ELA test score in 2022 relative to average scores between school years 2015-2019.

X_{ij}	Vector of student-level controls, including gender, race, ethnicity, primary language spoken, free-or reduced-priced meals eligibility, migrant status, limited English proficiency status, enrollment in special education, and enrollment in the gifted program
S_{js}	Vector of time-varying and invariant school-level and school neighborhood controls, including school enrollment and geographic location. School neighborhood controls include the percentage of the population unemployed, the percentage of the population with a high school diploma, an associate degree, a bachelor's degree, or a graduate degree, and the median household income.
α_s	School fixed effects control for time-invariant school factors.
γ_j	Year-fixed effects control for differences across years that impact scores.
ε_{ij}	Random error term clustered at the school level.

Limitations

The main limitation of our test score analysis is that we cannot estimate a causal relationship between the pandemic and changes to test scores. We attempt to control for as many observable student and school characteristics that may be correlated with academic achievement as possible, but we may not capture all factors that explain changes in test scores during the pandemic. We run additional models using different statistical techniques that account for changes between student populations tested before and after the pandemic and factors that occur during the same period as the pandemic and may be confounded with our estimate. We also test different specifications with varying control variables to test the robustness of our main results. Mostly, results are consistent across models and specifications (see pages 40-43).

Exhibit A4 compares student and school characteristics between pre- and post-pandemic periods. There are minimal differences in composition between the periods. There is a slightly larger proportion of Hispanic/Latino students in the post-period than in the pre-period and a smaller proportion of White students in the post-period than in the pre-period. The average school enrollment is slightly lower in the post-period than the pre-period, there is a smaller proportion of high-poverty schools in the post-period than the pre-period, and a slightly larger proportion of schools are located in rural areas in the post-period than the pre-period. Because of the large sample size, all differences are statistically significant.

Exhibit A4

Student and School Characteristics: Pre-Pandemic vs Post Pandemic

Student characteristics	Pre	Post
% Female	0.49	0.49
	(0.50)	(0.50)
% American Indian/Alaska Native	0.01	0.01
	(0.11)	(0.11)
% Asian	0.08	0.09
	(0.27)	(0.28)
% Black/African American	0.04	0.04
	(0.20)	(0.21)
% Hispanic/Latino	0.24	0.26
	(0.42)	(0.44)
% White	0.54	0.49
	(0.50)	(0.50)
% Native Hawaiian/Other Pacific Islander	0.01	0.01
	(0.10)	(0.11)
% Two or more races	0.08	0.09
	(0.27)	(0.28)
% Primary language is English	0.77	0.76
	(0.42)	(0.43)
% Primary language is Spanish	0.14	0.15
	(0.35)	(0.35)
% Primary language is other	0.09	0.10
	(0.28)	(0.29)
% Free or reduced-priced meals	0.48	0.48
	(0.50)	(0.50)
% Special education	0.13	0.13
	(0.33)	(0.33)
% Limited English proficiency	0.10	0.11
	(0.30)	(0.32)
% Migrant	0.01	0.01
	(0.11)	(0.11)
Total observations	2,778,903	517,106

Notes:
Standard deviations reported in parentheses.

Exhibit A4 Continued

Student and School Characteristics: Pre-Pandemic vs Post Pandemic

School characteristics	Pre	Post
Average school enrollment	761	723
	(452.46)	(463.85)
% High poverty schools	0.47	0.45
	(0.50)	(0.50)
% Schools in cities	0.32	0.32
	(0.47)	(0.47)
% Schools in suburbs	0.43	0.42
	(0.50)	(0.49)
% School in towns	0.12	0.13
	(0.33)	(0.33)
% Schools in rural areas	0.12	0.14
	(0.33)	(0.34)
School neighborhood characteristics		
% Pop. unemployed	6.32	5.13
	(3.24)	(3.10)
Median household income	$ 83,333	$ 90,921
	(33,788)	(37,527)
% Pop. with high school diploma	24.19	24.06
	(8.80)	(9.78)
% Pop. with associate degree	10.34	10.30
	(3.46)	(4.09)
% Pop. with bachelor's degree	19.38	20.81
	(10.43)	(10.81)
Total observations	2,778,903	517,106

Notes:
Standard deviations reported in parentheses.

Primary Analysis Results

Exhibit A5 reports the predicted probability of meeting grade-level standards on math and ELA assessments in 2022, compared to the average probability of meeting standards before the pandemic. For example, 3rd graders had a nine-percentage point lower probability of meeting math standards in 2022 than 3rd graders in pre-pandemic periods (a 15% decrease). 3rd graders had a six-percentage point lower probability of meeting ELA standards in 2022 than 3rd graders in pre-pandemic periods (a 12% decrease).

Exhibit A5

Predicted Change in the Probability of Meeting Grade-Level Standards

	Math 3rd Grade	4th Grade	5th Grade	6th Grade	7th Grade	8th Grade	10th Grade
Post	-0.089*** (-16.84)	-0.096*** (-18.13)	-0.10*** (-18.60)	-0.11*** (-17.85)	-0.13*** (-18.92)	-0.14*** (-20.18)	-0.084*** (-14.74)
Total observations	457,217	449,719	448,884	438,489	432,335	429,523	287,511
% Change	-15%	-17%	-22%	-23%	-26%	-29%	-18%
	ELA **3rd Grade**	**4th Grade**	**5th Grade**	**6th Grade**	**7th Grade**	**8th Grade**	**10th Grade**
Post	-0.059*** (-12.06)	-0.071*** (-14.37)	-0.065*** (-12.78)	-0.093*** (-16.30)	-0.054*** (-8.83)	-0.069*** (-11.38)	-0.029*** (-6.44)
Total observations	456,765	449,646	449,253	439,082	432,980	430,670	236,521
% Change	-12%	-11%	-16%	-10%	-11%	-22%	-5%

Notes:
Standard errors in parentheses. Standard errors clustered at the school level.
*** Significant at the 0.001 level, ** significant at the 0.05 level, * significant at the 0.10 level.

Exhibit A6 shows the change in average math and ELA test scores in 2022, compared to average scores in pre-pandemic years. For example, the average math score among 3rd graders in 2022 was 17.7 points lower than average math scores before the pandemic (a 0.20 SD decrease).

Exhibit A6

Predicted Change in Test Scores

	Math 3rd Grade	4th Grade	5th Grade	6th Grade	7th Grade	8th Grade	10th Grade
Post	-17.79*** (-16.21)	-21.81*** (-20.48)	-26.38*** (-20.81)	-31.23*** (-19.13)	-37.16*** (-21.35)	-41.37*** (-22.33)	-25.49*** (-18.20)
Total observations	447,515	446,885	446,547	435,498	429,804	426,922	201,100
Effect size (change in SD)	-0.20	-0.24	-0.27	-0.28	-0.32	-0.33	-0.22
	ELA **3rd Grade**	**4th Grade**	**5th Grade**	**6th Grade**	**7th Grade**	**8th Grade**	**10th Grade**
Post	-13.40*** (-12.56)	-14.59*** (-13.95)	-11.76*** (-10.40)	-20.04*** (-15.76)	-11.81*** (-8.01)	-16.98*** (-11.27)	-4.39*** (-3.83)
Total observations	447,515	446,885	446,547	435,498	429,804	426,922	201,100
Effect size (change in SD)	-0.14	-0.15	-0.12	-0.21	-0.11	-0.16	-0.05

Notes:
Standard errors in parentheses. Standard errors clustered at the school level.
*** Significant at the 0.001 level, ** significant at the 0.05 level, * significant at the 0.10 level.

Subgroup Results

In Exhibit A7, we report results from our subgroup analyses. The exhibit includes the number of observations in each estimate, the average change in test scores in the post period and standard error, and the associated effect size (change in SD). For example, average math scores among female students in 2022 were 34 points lower than average scores among female students pre-pandemic.

Exhibit A7

Predicted Change in Math and ELA Scores by Gender, Income Status, Race & Ethnicity, and School Locale

		Math			ELA		
	N	Post	SE	Effect size (SD)	Post	SE	Effect size (SD)
Gender							
Female	144,0878	-34.58***	(-37.08)	-0.30	-12.74***	(-15.04)	-0.12
Male	151,8271	-25.00***	(-29.15)	-0.21	-9.75***	(-11.13)	-0.08
Income status							
FRPM eligible	142,2775	-34.98***	(-37.64)	-0.34	-14.97***	(-15.59)	-0.13
FRPM ineligible	153,6374	-24.34***	(-25.02)	-0.22	-7.56***	(-8.20)	-0.07
Race & Ethnicity							
American Indian/Alaska Native	35,516	-33.86***	(-11.36)	-0.31	-13.98***	(-4.41)	-0.12
Asian	238,085	-27.96***	(-15.27)	-0.23	-8.82***	(-5.91)	-0.07
Black/African American	124,606	-37.77***	(-22.28)	-0.35	-16.12***	(-9.53)	-0.14
Hispanic/Latino	713,396	-35.78***	(-28.94)	-0.35	-14.01***	(-10.91)	-0.12
White	1,579,428	-25.55***	(-29.42)	-0.23	-9.35***	(-10.78)	-0.08
Native Hawaiian/other Pacific Islander	32,390	-50.23***	(-17.18)	-0.49	-25.00***	(-8.82)	-0.23
Two or more races	235,490	-30.20***	(-24.20)	-0.26	-11.89***	(-9.74)	-0.10
Geographic locale							
City	949,545	-33.57***	(-22.21)	-0.29	-13.56***	(-9.95)	-0.11
Suburb	1,276,116	-29.60***	(-20.18)	-0.25	-11.25***	(-7.75)	-0.10
Town	364,249	-28.27***	(-14.19)	-0.25	-11.10***	(-5.22)	-0.10
Rural area	369,239	-22.07***	(-12.86)	-0.21	-6.232**	(-3.25)	-0.05

Notes:
Standard errors in parentheses. Standard errors clustered at the school level.
*** Significant at the 0.001 level, ** significant at the 0.05 level, * significant at the 0.10 level.

Effects by School District

Exhibits A8 and A9 show the change in math and ELA test scores in 2022 across school districts.

Exhibit A8

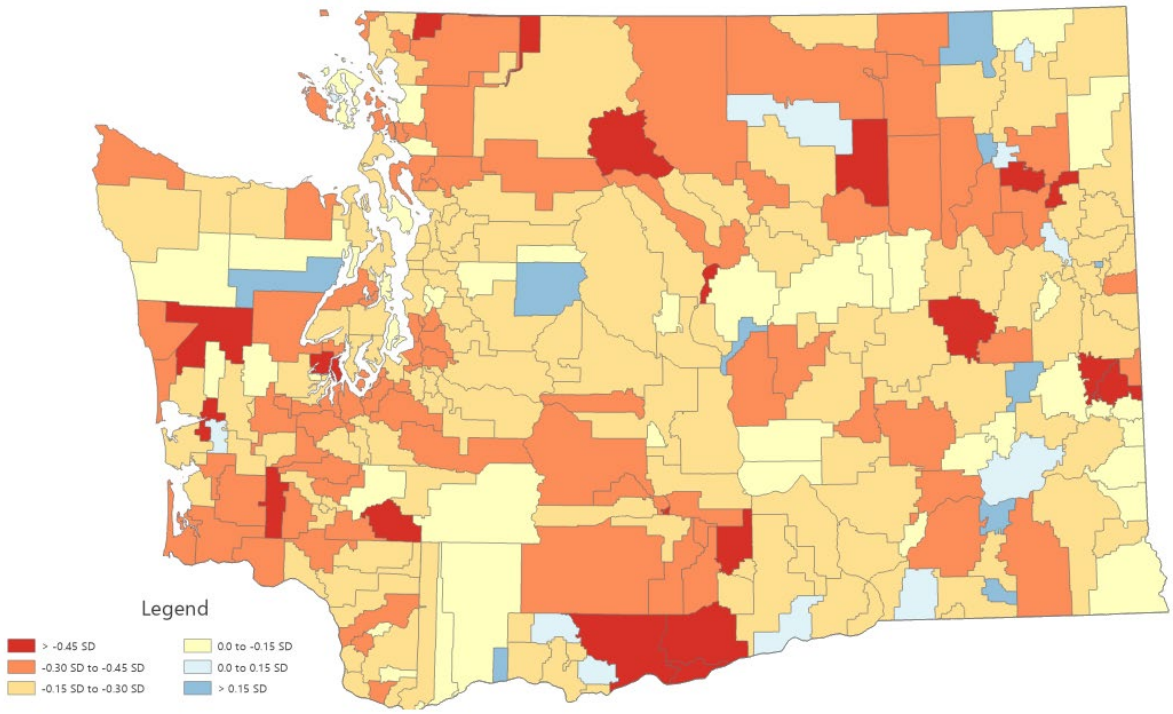
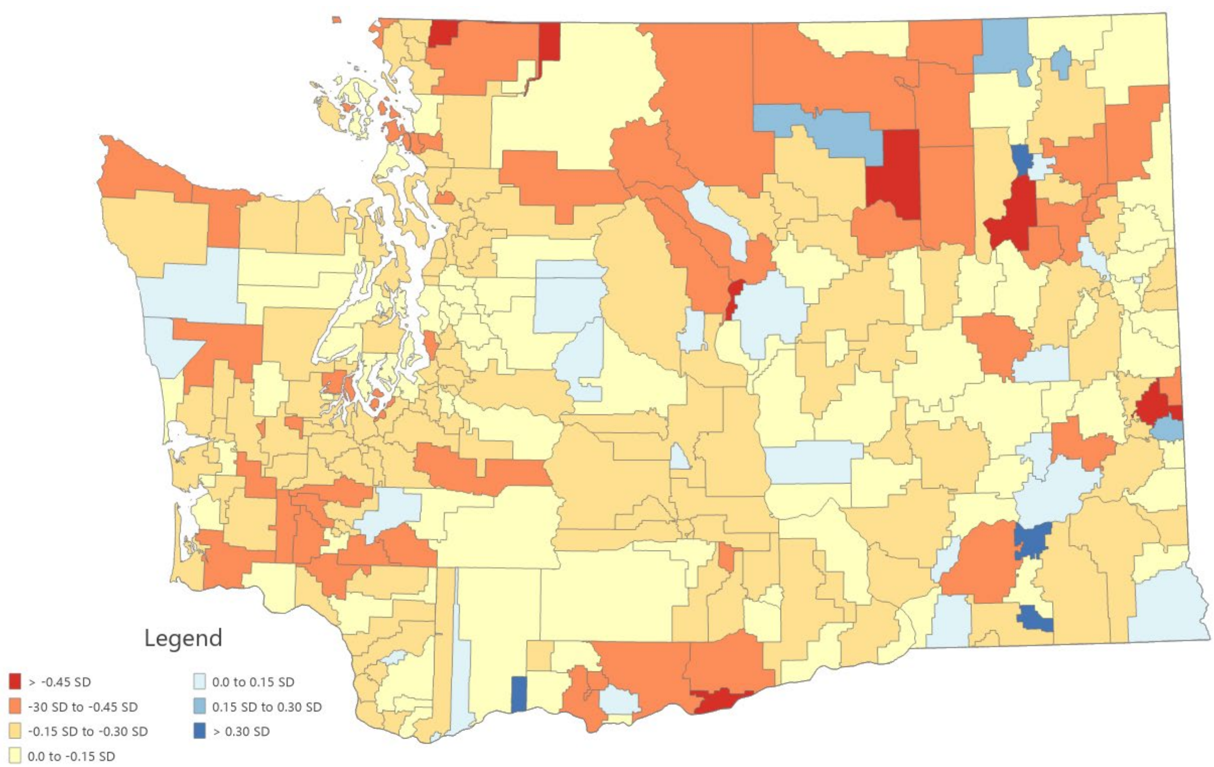


Exhibit A9



Sensitivity Analysis and Results

While we control for a robust set of student and school characteristics in our main model, we cannot capture all observed and unobserved factors that explain changes in test scores. Therefore, our estimated effect may be biased. We run additional analyses to test the sensitivity of our results across various specifications, to control for differences between student populations before and after the pandemic, and to control for events that may have occurred at the same time as the pandemic.

Testing Various Specifications

We run specifications incrementally, adding controls and observing the robustness of results. Exhibit A10 shows math and ELA test score results for each specification.

Exhibit A10

Sensitivity Results Based on Specifications

	Specifications				
	1	2	3	4	5
			Math		
Post	-24.73*** (-37.77)	-27.79*** (-37.27)	-28.85*** (-39.89)	-29.03*** (-39.31)	-29.68*** (-34.51)
			ELA		
Post	-7.86*** (-44.50)	-10.43*** (-14.90)	-11.03*** (-15.95)	-11.05*** (-15.44)	-11.26*** (-13.61)
Included controls					
School/year fixed effects		X	X	X	X
Student controls			X	X	X
School controls				X	X
Neighborhood controls					X
Total observations	3,296,009	3,296,009	3,296,009	3,296,009	3,296,009

Notes:
Specification 5 illustrates our preferred model.
Standard errors in parentheses. Standard errors clustered at the school level.
*** Significant at the 0.001 level, ** significant at the 0.05 level, * significant at the 0.10 level.

Specifications Comparing Sample Restrictions

Exhibit A11 shows the effects as we make or do not make certain data cleaning restrictions. Column 1 shows results if we include all school types (e.g., alternative schools, vocational schools, Tribal compact schools, etc.). Columns 2 and 3 show results if we exclude students in special education or the Gifted program. Column 4 shows results if we restrict the sample so that only students assessed in the same grade they are enrolled in are included (sometimes students are administered math and ELA tests that assess grade-level standards of a grade other than the one they are enrolled in). Column 5 shows our primary model results. The estimated test score effects remain similar across sample restrictions.

Exhibit A11

Sensitivity Results Based on Restriction Decisions

	Specifications				
	1	2	3	4	5
			Math		
Post	-29.14*** (-34.85)	-30.59*** (-33.42)	-30.52*** (-35.88)	-30.13*** (-33.80)	-29.68*** (-34.51)
			ELA		
Post	-11.00*** (-13.83)	-11.34*** (-13.04)	-11.75*** (-14.29)	-13.88*** (-18.17)	-11.26*** (-13.61)
Total observations	3,039,890	2,580,787	2,733,141	2,834,271	3,296,009

Notes:
Standard errors in parentheses. Standard errors clustered at the school level.
*** Significant at the 0.001 level, ** significant at the 0.05 level, * significant at the 0.10 level.

Sample with Entropy Weights

Entropy weighting is a data processing method that balances covariates between treatment and comparison groups in observational studies.[65] This method estimates weights for a comparison group such that the reweighted comparison group and treatment group balance on covariates that include information about known sample moments (e.g., mean, variance, skewness) and minimize entropy distance (i.e., uncertainty).[66] Entropy weights allow us to adjust for inequalities in observed predictors across treatment and comparison groups. Since the population of students tested before and after the pandemic has changed in ways that predict changes to test scores, we estimated entropy weights to balance characteristics between students tested in 2022 and students tested in the years 2015-2019. Exhibit A12 compares the effect on test scores estimated using our main analytic sample and a sample with entropy weights. Results are substantively equivalent whether or not we use entropy weights.

Exhibit A12

Main Results With and Without Entropy Weights

	Sample with entropy weights	Analytic sample
	Math	
Post	-30.05*** (-31.54)	-29.68*** (-34.51)
	ELA	
Post	-11.99*** (-13.68)	-11.26*** (-13.61)
Total observations	3,296,009	3,296,009

Notes:
Standard errors in parentheses. Standard errors clustered at the school level.
*** Significant at the 0.001 level, ** significant at the 0.05 level, * significant at the 0.10 level.

[65] Hainmueller, J. (2012). Entropy balancing for casual effects: a multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis, 20,* 25-46.
[66] Ibid.

Difference-in-Differences Model

Our primary model may not account for differences between students tested before and after the pandemic, which may drive changes in performance. To control for this, we run a difference-in-differences (DiD) model. This approach compares the change in outcomes over time for a treated group to a comparison group. If assumptions hold, DiD will control for observed and unobserved factors that occur during the same time as the pandemic and affect test scores.

In the DiD model, we compare changes in test scores between two grades (e.g., 3rd and 6th grade) for cohorts of students tested before and after the pandemic. This allows us to control for cohort effects by examining the same students before and after the pandemic (e.g., we observe Student A's 3rd-grade test score in 2019 and their 6th-grade test score in 2022). We observe the change in 3rd- and 6th-grade test scores for students in a post-pandemic cohort (treatment group) relative to the change in 3rd- and 6th-grade test scores for two pre-pandemic cohorts (comparison groups). Exhibit A13 illustrates an example of these comparison and treatment cohorts.

Exhibit A13

School year	Comparison cohort 1	Comparison cohort 2	Treatment cohort
2015	3rd grade		
2016		3rd grade	
2017			
2018	6th grade		
2019		6th grade	3rd grade
2020			
2021			
2022			6th grade

We also examine changes to math and ELA scores between grades 4 and 7, 5 and 8, and 7 and 10. We focus on these specific grades in order to compare changes over the same lengths of time, before and after the pandemic (we cannot analyze test scores in 2020 or 2021 due to the cancelation and postponement of spring assessments). We use the following DiD model to estimate this change:

$$\Delta Score_i = \beta_0 + \beta_1(Post_i) + \beta_2(X_{ij}) + \beta_3(S_{js}) + \alpha_s + \varepsilon_{ij}$$

$\Delta Score_i$ — Difference between a student's math/ELA test scores in two grades. We run separate models to estimate the difference between 3rd- and 6th-grade scores, 4th- and 7th-grade scores, 5th- and 8th-grade scores, and 7th- and 10th-grade scores and estimate separate models for math and ELA subjects.

$Post_i$ — Indicates if a student's second test occurred before or after the pandemic. In all cohorts, the first test will be pre-pandemic, but in the 2022 cohort, the second test occurred after. β_1 is the parameter of interest and can be interpreted as the change in test scores resulting from the pandemic.

X_{ij} — Vector of student-level controls (same as the primary model)

S_{js} — Vector of time-varying and invariant school-level and neighborhood controls (same as primary model)

α_s	School fixed effects
ε_{ij}	Random error term clustered at the school level.

Exhibits A14 and A15 report math and ELA results for the DiD model and our primary model. DiD estimates indicate that the cohorts in school during the pandemic experienced less growth than pre-pandemic cohorts. In other words, there was a larger increase in scores between grades (e.g., 3rd and 6th) before the pandemic than after the pandemic. We compare our estimates between our primary model and the DiD model. In terms of magnitude, the DiD and primary model results for math scores are similar across grades. However, the DiD estimates for ELA scores are larger across all grades than our primary model estimates.

Exhibit A14

DiD Model Results: Math

	Change in 6th-grade scores	Change in 7th-grade scores	Change in 8th-grade scores	Change in 10th-grade scores
Estimated effect from DiD model	-35.67***	-35.78***	-42.18***	-31.89***
	(-32.14)	(-34.64)	(-35.27)	(-33.09)
Total observations	183,253	181,014	181,451	165,225
Effect size (change in SD)	-0.32	-0.31	-0.33	-0.25
Estimated effect from primary model	-31.23***	-37.16***	-41.37***	-25.49***
	(-19.13)	(-21.35)	(-22.33)	(-18.20)
Total observations	435,498	429,804	426,922	201,100
Effect size (change in SD)	-0.28	-0.32	-0.33	-0.22

Notes:
Standard errors in parentheses. Standard errors clustered at the school level.
*** Significant at the 0.001 level, ** significant at the 0.05 level, * significant at the 0.10 level.

Exhibit A15

DiD Model Results: ELA

	Change in 6th-grade scores	Change in 7th-grade scores	Change in 8th-grade scores	Change in 10th-grade scores
Estimated effect from DiD model	-27.43***	-21.29***	-25.92***	-17.26***
	(-32.03)	(-24.41)	(-28.79)	(-17.02)
Total observations	183,253	181,014	181,451	165,225
Effect size (change in SD)	-0.28	-0.21	-0.25	-0.17
Estimated effect from primary model	-20.04***	-11.81***	-16.98***	-4.394***
	(-15.76)	(-8.01)	(-11.27)	(-3.83)
Total observations	435,498	429,804	426,922	201,100
Effect size (change in SD)	-0.21	-0.11	-0.16	-0.05

Notes:
Standard errors in parentheses. Standard errors clustered at the school level.
*** Significant at the 0.001 level, ** significant at the 0.05 level, * significant at the 0.10 level.

II. Predicted Future Earnings

WSIPP has developed a benefit-cost model (BCM) to produce consistent estimates of the benefits and costs of programs in policy areas like adult criminal and juvenile justice, K-12 education, child welfare, and substance use. In a typical application, we compare the benefits of a program or policy to the costs associated with it. The model does this by valuing changes in outcomes (e.g., test scores) produced by programs and comparing them to the costs of the program. For our purposes, we use the BCM to estimate how changes in test scores predict changes to future earnings. See WSIPP's Benefit-Cost Technical Document for more information.[67]

Valuation of Test Scores in the BCM

The BCM monetizes outcomes across policy areas. For K-12 programs and policies, the model includes inputs that monetize standardized test scores. For example, the model includes a parameter that links changes in test scores (measured in standard deviations) to changes in future labor market earnings at a per-student level.[68] For this report, we include estimates from our test score analysis to determine how a change in test scores (post-pandemic) predicts changes to future earnings. We input the following effects from our analyses into the BCM:

- -0.20 effect size (0.002).[69] The SD change in test scores among all students tested in 2022, compared to average scores before the pandemic.
- -0.18 effect size (0.002). The SD change in test scores among elementary students tested in 2022, compared to average scores before the pandemic.
- -0.25 effect size (0.002). The SD change in test scores among middle school students tested in 2022, compared to average scores before the pandemic.
- -0.12 effect size (0.005). The SD change in test scores among high school students tested in 2022, compared to average scores before the pandemic.

Note that we have combined math and ELA results into a single effect because the BCM does not value math or ELA tests differently or estimate different earnings trajectories based on test subjects.

Using the BCM's Fade-out Estimates to Model Test Score Recovery

Research has found that test score gains from program participation, particularly in earlier grades, fade over time. The BCM accounts for these fade-out effects.[70] For example, suppose a 3rd grader participates in a program that increases test scores by 0.15 SD. In that case, this effect will fade to about 0.05 SD by the time the student is 17, when they may enter the labor market, and the period at which we estimate a program's impact on earnings.

[67] Washington State Institute for Public Policy. (December 2019). *Benefit-cost technical documentation.* Olympia, WA: Author.
[68] Ibid. Section 4.8.
[69] Standard error in parentheses.
[70] WSIPP (2019). Section 4.8g.

While it is still too early to determine how student academic achievement will progress post-pandemic, some research indicates that test scores will recover to some extent in the coming years.[71] We use the BCM's existing fadeout parameters to estimate how test scores may recover over time. Exhibit A16 shows our estimated test score effects for all elementary, middle, and high school students in 2022. The table also shows how these effects may recover by age 17. We also estimate a less likely scenario in which test scores do not recover in the coming years.

Exhibit A16

Predicted Test Score Change (SD) by Grade Level and Estimated Recovery by Age 17

Population	Estimated test score effect (change in SD)	Average age in 2022	Estimated test score effect at age 17 (change in SD) Gradual recovery
All students	-0.20	11	-0.15
Elementary	-0.18	9	-0.11
Middle	-0.25	12	-0.19
High School	-0.12	16	-0.11

Test Score Link to Future Earnings

Exhibit A17 shows predicted change in future earnings, assuming test scores recover gradually or do not recover. For example, assuming a gradual recovery, a 0.20 SD decrease in test scores for all students in 2022 is associated with a $32,000 decrease in lifetime earnings per student on average. If test scores do not recover, we estimate a $44,500 decrease in earnings.

Exhibit A17

Predicted Change to Future Earnings: Gradual Test Score Recovery vs. No Test Score Recovery

	Estimated test score effect (change in SD)	Decrease in lifetime earnings (present value) Gradual Recovery	Decrease in lifetime earnings (present value) No Recovery
All students	-0.20	$32,086	$44,479
Elementary	-0.18	$22,940	$38,234
Middle school	-0.25	$42,687	$55,336
High school	-0.12	$26,715	$29,412

Note that elementary and high school students experience a similar decrease in earnings, even though we found different test score effects between the groups. This is because elementary students have a longer time to recover than high school students, who only have a few years before graduation. Due to fadeout estimates in the BCM, test score effects among elementary and high school students recover to a similar level by the time students are 17 years old.

[71] Sacerdote (2012); Kuhfeld & Lewis (2022); and Halloran et al. (2023).

Sensitivity Analysis

We run a Monte Carlo simulation to account for the uncertainty around inputs and assumptions in the BCM. In this analysis, the BCM runs results 10,000 times, each time varying inputs after sampling from estimated ranges of uncertainty around key inputs. We find that the estimated average decrease in per-student earnings (as shown in Exhibit A17) remains the same, but we observe a large range in potential earnings loss.

- All students - 90% of the 10,000 estimates fall between $13,200 and $53,400.
- Elementary students - 90% of the 10,000 estimates fall between $9,400 and $38,000.
- Middle school students - 90% of the 10,000 estimates fall between $18,000 and $69,800.
- High school students- 90% of the 10,000 estimates fall between $11,400 and $43,000.

Assumptions and Limitations

To conduct this analysis, we use WSIPP's BCM model differently than our normal approach. Typically, we compare benefits and costs between program participants and non-participants. However, in our analysis, everyone in Washington was affected by the pandemic, and we cannot compare test scores between students affected and unaffected by the pandemic. We assume that our predicted earnings, which reflect a society-wide impact, would be the same if the pandemic impacted some students and not others. However, it is possible that in a scenario where the pandemic affects some students and not others, earnings estimates would be different.

Further, the BCM has been developed with underlying assumptions based on pre-pandemic relationships. We assume that these pre-pandemic parameters will hold in a post-pandemic environment.

For example, the model includes a parameter that links changes in test scores to changes in future earnings based on research conducted before the pandemic. If test scores after the pandemic have a smaller effect on earnings than before the pandemic, we may be overestimating the decrease in earnings. The model also includes estimates of average earnings over an individual's working life (from age 17 to 65). We assume these earnings trajectories are the same before and after the pandemic. So far, WSIPP has not found any indication that earnings patterns have changed due to the pandemic.

We also assume that the fadeout estimates in the BCM accurately model test score recovery in Washington in the post-pandemic period. Some research has found that test scores recover in the following years after large natural disasters.[72] Also, initial COVID research has found that test scores in other states are recovering.[73] We assume that some recovery will occur in the coming years and use the BCM's existing fadeout model to estimate how a gradual recovery will influence test score effects. It is possible that the BCM parameters do not accurately predict academic recovery. Because of this, we also estimated results assuming test scores do not recover post-pandemic.

[72] Sacerdote (2012).
[73] Kuhfeld & Lewis (2022) and Halloran et al. (2023).

III. Recovery Interventions Analysis

WSIPP has developed a standardized approach to conduct literature reviews and meta-analyses to estimate average program effects. In the first phase, a literature review, we systematically review the national and international research literature on a given topic and review all available studies, regardless of findings. Next, we screen studies to identify rigorous evaluations. Before conducting a meta-analysis, we ensure that a study reasonably attempts to demonstrate a causal relationship between the program and targeted outcomes. For example, studies must include treatment and comparison groups with an intent-to-treat analysis. Studies that do not meet our minimum standards are excluded from the analysis.

Once we identify all rigorous program evaluations on a given program or intervention, we conduct a meta-analysis. In this stage, we use a set of statistical procedures to calculate an average weighted effect size for each outcome, which indicates the expected magnitude of change caused by a program (e.g., tutoring by adults) for each outcome (e.g., standardized test scores). We can also incorporate the estimated effect size from the meta-analysis into WSIPP's benefit-cost model to estimate future benefits to program participants.

To date, WSIPP researchers have conducted meta-analytic reviews for almost 80 early education and K-12 programs and practices.[74] For this report, we pulled from previous analyses on tutoring programs, academically-focused summer school programs, and double-dose classes. We considered how these programs may offset the decline in test scores we estimated in earlier analyses. We chose these specific interventions because they are commonly reported as strategies that school districts use to help students recover academically after the pandemic.[75]

Exhibit A18 reports the tutoring models, summer school programs, and double-dose classes we have analyzed and estimated effects on test scores (reported as effect sizes). Each program is linked to WSIPP's website with additional information.

[74] Washington State Institute for Public Policy. (2014, July). *Pre-K to 12 education benefit-cost results*. Olympia, WA: Author.
[75] Carbonari et al. (2022); JLARC (2023); and FutureEd (2021).

Exhibit A18

Interventions and Estimated Impact on Test Scores

Program name	Average effect on test scores (effect size)
Tutoring: by adults, 1:1, structured	0.39
Tutoring: by peers	0.31
Tutoring: by certificated teachers, small groups, structured	0.26
Tutoring: by non-certificated adults, small groups, structured	0.26
Tutoring: by adults, after school	0.26
Tutoring: by parents with teacher oversight	0.17
Tutoring: by adults, 1:1, unstructured	0.03
Double-dose classes	0.09
Academically focused summer learning programs	0.06

Assumptions and Limitations

We conducted our meta-analyses on tutoring, summer school, and double-dose programs before the pandemic. As a result, the average effects we estimate may not be the same if applied to students post-pandemic. For example, because of achievement declines during the pandemic, peer tutoring strategies may impact student test scores less than we estimate in our meta-analysis.

Further, our meta-analyses include studies with far fewer participants and a more homogenous population of students than the students included in our test score analysis. Because of this, it's difficult to determine if tutoring models, summer learning programs, or double-dose classes will have the same effects if scaled to a larger and more diverse student population.

For further information, contact: Julia Cramer
Name at 360.664.9073, julia.cramer@wsipp.wa.gov

Document No. 23-09-2201