A Broader View of Recovery: Exploring Non-Test Student Outcomes in the Wake of the COVID-19 Pandemic

Ian Callen
Elise Dizon-Ross
Dan Goldhaber
Thomas J. Kane
Andrew McEachin
Emily Morton

August 2026

WORKING PAPER No. 344-0826

CALDER
National Center for Analysis of Longitudinal Data in Education Research

AIR

A Broader View of Recovery: Exploring Non-Test Student Outcomes in the Wake of the COVID-19 Pandemic

Ian Callen
University of Virginia

Elise Dizon-Ross
American Institutes for Research / CALDER

Dan Goldhaber
American Institutes for Research / CALDER
University of Washington / CEDR

Thomas J. Kane
Harvard University

Andrew McEachin
ETS

Emily Morton
NWEA

Contents

Acknowledgments

This research was supported by the AIR Opportunity Fund. CALDER is funded by a consortium of foundations (for more information about CALDER funders, see www.caldercenter.org/about-calder). We are grateful to Guilford County Schools, without whose support and ongoing partnership this research would not have been possible. We would also like to thank Michael DeArmond and Ben Backes for helpful feedback.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any errors are attributable to the authors.

CALDER • American Institutes for Research

1400 Crystal Drive 10th Floor, Arlington, VA 22202

202-403-5796 • www.caldercenter.org

A Broader View of Recovery: Exploring Non-Test Student Outcomes in the Wake of the COVID-19 Pandemic

Ian Callen, Elise Dizon-Ross, Dan Goldhaber, Thomas J. Kane, Andrew McEachin, & Emily Morton

CALDER Working Paper No. 344-0826
August 2026

Abstract

The COVID-19 pandemic sharply reduced student test scores, and ESSER-funded academic interventions largely failed to recover those losses. Yet few researchers have examined whether these interventions affected students in ways beyond test scores. We use rich longitudinal data from Guilford County Schools (GCS) to document trends before and after the pandemic in attendance, suspensions, disciplinary referrals, and course grades from 2016-17 through 2023-24. All non-test outcomes worsened in GCS after the pandemic, widening pre-existing gaps by race and school poverty level. We also estimate how participation in GCS's tutoring and summer school in 2021–22 and 2022–23 affect these non-test outcomes. Students who participated in recovery interventions showed improved attendance and discipline in the following year. Effects varied across interventions: summer school robustly improved attendance the following school year, while tutoring improved discipline in one year (2021–22) and attendance in the other (2022–23). These results demonstrate that academic interventions support students on dimensions beyond traditionally measured academic performance and contribute to recovering student engagement and well-being.

1. Introduction

The disruptions caused by the COVID-19 pandemic had wide-ranging consequences for students, not only for their academic achievement but also their behavior and well-being. A consistent body of research has documented the pandemic's large negative impacts on student test scores (e.g. Fahle et al. 2023; Halloran et al., 2023; Lewis and Kuhfeld, 2023; U.S. Department of Education, 2022). But studies across multiple states have also found pandemic-driven declines to student outcomes beyond test scores (Fuller et al., 2023), including attendance (Dee, 2024; Malkus, 2025) and measures of college readiness (Polikoff et al., 2023).

Studies of COVID-19 recovery in schools find that schools have struggled to reverse the pandemic's academic damage. Students have regained test performance only gradually, and recovery interventions have produced gains too small to offset the full extent of pandemic-era setbacks (Callen et al., 2025; Carbonari et al., 2024).[1]

Recovery on *non-test* outcomes has received less systematic attention (Polikoff et al., 2023). Outside of a few exceptions (e.g., Bastian and Fuller, 2024; Lee et al, 2025), we know little about what happened across the era of the pandemic on non-test student outcomes and the degree to which different types of interventions succeeded in academic recovery efforts.

We track non-test outcomes, such as attendance, discipline, and grades, in one school district and examine whether academic recovery interventions there improved these outcomes. Our data come from Guilford County Schools (GCS) in North Carolina, spanning the years from 2016-17 through 2023-24, and include student demographics, enrollment, four non-test outcomes (attendance, disciplinary referrals, suspensions, and course grades), and performance on state standardized tests and other interim assessments. Importantly, we observe student participation

[1] Estimates from state assessments, for example, suggest that student achievement in the 2023-24 school year lagged 2018-19 by nearly half a grade (Dewey et al. 2025).

in school-year tutoring and summer recovery interventions that targeted elementary and middle school students from fall 2021 through summer 2023. We address three questions about non-test outcomes during and after the pandemic:

1. To what extent did attendance, disciplinary referrals, suspensions, and course performance change from pre-pandemic levels?
2. What is the relationship between these non-test outcomes and test-score outcomes and how did the relationship change over the course of the pandemic?
3. How did non-test outcomes differ between students who participated in tutoring or summer school and non-participants?

Our data extend prior work by covering more outcomes, a longer time period, multiple interventions, and variation across student and school characteristics. We begin by tracking outcome trends, with a particular focus on different student subgroup trends in attendance, behavioral measures, and grades across the pandemic period. We then examine the association between students' non-test measures and their end-of-grade (EOG) performance on state standardized tests to understand how this relationship may have shifted before and after the pandemic. Finally, we use value-added models (VAMs) to compare the non-test outcomes of students participating in academic interventions with the outcomes of non-participants. In this VAM approach, we control for observable pre-treatment characteristics and baseline measures of non-test outcomes.

As we discuss in more detail in the Methods section, the VAM methodology we employ accounts for differences in observable characteristics of students but cannot rule out the possibility of selection into treatment based on unobserved characteristics; hence we do not interpret our estimates as necessarily causal. However, for brevity, we use the terms “impacts” or “effects” to describe the estimated relationships between intervention participation and outcome.

Consistent with national evidence on absenteeism (Malkus, 2025), we find that absence rates of students in GCS increased sharply following the pandemic, peaking in 2021-22 and improving slightly the following years. Differences in absenteeism by race/ethnicity and school poverty also increased, raising concerns about lasting disparities. Students experienced notable increases in the likelihood of suspensions and disciplinary referrals following the pandemic, particularly among Black students and those attending higher poverty schools. Lastly, for nearly all students, course grades decreased slightly and as of 2022-23, had not fully recovered.

Additionally, consistent with evidence on the pandemic-era comorbidity between students’ disengagement in school and their outcomes (Malkus, 2025; Xu et al., 2025), we find stronger associations post-pandemic between absenteeism and suspensions and students’ end-of-grade state standardized test scores. We also find stronger associations between course grades and test scores in math, contrary to results found in Washington state (Goldhaber and Goodman Young, 2024).

Both summer school and tutoring improved non-test outcomes in the following school year. Summer school most clearly boosted attendance and reduced chronic absenteeism. Tutoring effects were more uneven: in 2021–22, tutored students received fewer disciplinary actions, while in 2022–23, they showed better attendance. We generally do not find any significant impacts of recovery interventions on students’ future grade point averages.

This paper contributes to a growing literature that applies education production function methods, traditionally used to examine test score outcomes (e.g. Hanushek, 1971, 1992), to the study of non-test outcomes (see for example: Backes et al., 2024; Jackson, 2018; Liu and Loeb, 2019; McEachin, Lauen, Fuller, Perera, 2020; McEachin, Welsh, & Brewer, 2016). Research in this area highlights the importance of measuring the effect of educational policies, programs, and inputs on outcomes beyond student test scores. We complement this literature by examining whether academic interventions implemented during the pandemic recovery period improve non-test outcomes. We build on a nascent body of research documenting the non-academic benefits of pandemic-era interventions, such as increased concurrent attendance from high-dose tutoring (Lee et al., 2025). To our knowledge, ours is the first study to examine non-test outcomes in relation to multiple types of ESSER-funded academic recovery interventions across sustained periods of time.

2. Background and Literature

Students' academic outcomes declined sharply in the wake of the COVID pandemic outbreak (e.g. Lewis and Kuhfeld, 2023; U.S. Department of Education, 2022). By spring 2022, 3rd through 8th graders had lost roughly half a grade level in math and a third of a grade level in reading (Fahle et al., 2023). Students of color and students in high-poverty schools lost the most ground, widening disparities that predated the pandemic (e.g., Goldhaber et al., 2022). Evidence on recovery to pre-pandemic achievement levels is mixed. Some studies document initial gains that then stalled (Kuhfeld & Lewis, 2022; Dewey et al., 2024; Lewis & Kuhfeld, 2023), though others find more substantial progress (Fahle et al., 2023). By any estimate, students have not yet fully regained lost ground.

COVID disruptions affected students well beyond their performance on academic tests (Polikoff et al., 2023). Chronic absence rates nationwide among public school students nearly doubled between 2018-19 to 2021-22 and, by 2024, still remained far above pre-pandemic averages (Dee, 2023; Malkus, 2025). Across North Carolina, Guilford County's home state, students experienced sharp increases in absences, as well as increases in suspensions and decreases in grade point averages[2], during the early phases of the pandemic (Fuller et al., 2023). In particular, the state struggled to support students with prior difficulties, middle school students, and students from marginalized backgrounds.

Suspension trends shifted during the pandemic. In one anonymous school district, suspensions dropped temporarily during virtual learning in 2020-21, then rose the following year. Even amid the one-year decline, Black students still received exclusionary discipline at disproportionate rates (Welsh, 2022). Our study is the first, to our knowledge, to document how disciplinary referrals, a less severe and more common behavioral outcome, changed over the course of the pandemic.

Because we rely on disciplinary referrals and suspensions as key outcomes, it is important to recognize what these records capture. Disciplinary records reflect adults' decisions more than students' behavior (e.g. Barrett et al., 2021). Black students receive suspensions and office referrals at rates far exceeding those of white students, and these gaps persist even after accounting for differences in observed behavior and school context (Skiba et al., 2002; Barrett et al., 2021). Experimental evidence documents teachers judge identical behaviors as more severe when they believe the student is Black, and they escalate discipline more quickly for Black students after repeated incidents (Okonofua & Eberhardt, 2015). Further, school leaders suspend Black students involved in interracial incidents at higher rates and for longer periods of time than their white peers (Barrett et al. 2021; Liu, Hayes, & Gershenson, 2024; Shi & Zhu, 2022). This differential treatment has a lasting negative effect for Black students (Davison et al., 2021). Changes in referral rates over time could reflect genuine shifts in student behavior, shifts in how

[2] This trend contrasts with evidence from Washington state, where Goldhaber and Goodman Young (2023) find that both math and English GPAs actually rose in 2020-21 and 2021-22 following state guidance to districts to ease grading standards.

adults monitor and sanction that behavior, or both. We cannot disentangle these mechanisms with our data, and readers should interpret changes in discipline outcomes with this limitation in mind.

Trends in non-test outcomes vary by region but paint a consistently negative picture, forcing schools to confront a daunting problem. Schools must find ways to improve not only students' academic performance, but also their attendance, course grades, and the school's own use of disciplinary practices. The stakes for improving non-test outcomes are arguably just as high as for test scores. Research has long established a clear relationship between absenteeism and academic achievement (e.g. Chang and Romero, 2008). Yet district leaders report that remote learning accelerated a cultural shift among students and families, one in which school attendance feels optional rather than compulsory (Myung and Hough, 2023). This weakened sense of obligation has eroded engagement with school as an institution, and schools themselves have faced challenges implementing the range of initiatives needed to re-engage students (Diliberti, 2025). Schools cannot recover academically without rebuilding student engagement.

Whether ESSER-funded interventions improved outcomes beyond test scores remains an open question. A nascent body of pre-pandemic evidence suggests it is possible for these types of interventions to affect these outcomes (e.g. Fryer and Howard-Noveck, 2020; Pyne et al., 2023). The pandemic-era interventions we study, however, differed sharply from those that came before. Districts designed and implemented them at scale, under considerable urgency, and in response to unprecedented disruptions to schooling.

Few studies have examined whether COVID-era interventions improved non-test outcomes. In North Carolina, elementary and middle school students who attended the 2021 summer program at high rates experienced less chronic absenteeism and earned higher course grades the following year than low attenders (Bastian and Fuller, 2024). High school summer attenders repeated courses less often than observably similar non-attenders and showed modest improvements in attendance and GPA, though they received suspensions at higher rates (Fuller et al., 2024). In Washington D.C., students in a high-impact tutoring program during 2022-23 were less likely to miss school on days they received tutoring (Lee et al., 2025). Our study builds on this small body of work in two ways. First, we span both school-year and summer interventions and focus on sustained rather than contemporaneous impacts. Second, our examined outcomes include disciplinary referrals, a more sensitive measure of school disciplinary practices, offering more nuanced insight into how schools' behavioral decision-making changed following the interventions.

3. COVID-Era Academic Recovery Programs in GCS

For RQ3, we focus on school-year math and English Language Arts (ELA) tutoring and summer school[3] interventions serving K-8 students[4] that took place during the 2021-22 and 2022-23 school years and the following summers. These programs targeted a limited subset of students based on district-determined eligibility criteria.

GCS launched its school-year tutoring program, designed to help students recover from COVID-related academic setbacks, in 2021-22. The district designed the program for students in grades K-8 who scored below the 20th percentile on the MAP Growth assessment the previous

[3] GCS additionally offered tutoring in the summer of 2022 to a small number of students. Supplemental analyses of the impacts of summer tutoring find similar results to those of summer school in 2022. Results available from authors upon request.

[4] We focus on K-8 students because of the structure/targeting of academic recovery programs in those grades and the related ability to obtain credible estimates of program effects. GCS did offer a recovery intervention program, known as Learning Hubs, for high school students during this period. However, this program was less targeted than the interventions we study here and was available to all interested students. The implementation of Learning Hubs also differed in critical ways across school sites, for example, in how they were described to students and their usage (e.g. as an opportunity for extended instructional time, or as a consequence for misbehavior). Given concerns about selection bias and variation in implementation, we do not include this program in our study.

spring. Teachers and schools had discretion to identify additional students falling behind in courses and/or with poor attendance.

GCS drew tutors from a range of sources: high school students, college students, graduate assistants, community tutors, North Carolina Education Corps members, and GCS teachers. When scheduling allowed, tutors pushed into students' classrooms or pulled students out during non-core instruction blocks. Schools offered after-school sessions when tutoring could not fit within the school day. Most sessions took place in person in small groups of up to five students, though some students received one-on-one virtual tutoring.

GCS designed the tutoring program with the goal of students receiving two to three sessions per week, each lasting thirty to sixty minutes. Actual dosage, however, fell short of program goals and varied considerably across participants, consistent with patterns in other districts' COVID recovery tutoring programs (Carbonari et al., 2022, 2024). In 2021-22, tutored students averaged 20.8 sessions, or 13.0 hours, and most (1,706 of 2,832) received tutoring only in the spring semester. Dosage increased modestly the following year, with participating students averaging 27.0 sessions, or 14.8 hours, in 2022-23.[5]

Following the pandemic outbreak, GCS expanded summer school in 2021 to include all grades (rising K-12). In prior years, the program had served only rising grades 1-4 and high school students. The district maintained the expanded model in 2022 but limited the program to students at Title 1 schools in rising grades 3-4 in 2023. Both years, the district explicitly encouraged low-performing students to enroll in the program.[6] District administrators invited students who scored: at or below the 40th percentile on their winter MAP Growth math or

[5] Averages based on students included in our analytic samples.
[6] GCS' website for the middle and high school programs said, "The program is specifically designed for students with persistent learning gaps or demonstrated to have been disproportionately impacted by the COVID-19 pandemic and students transitioning to middle school."

reading test; scored below grade-level on their winter Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test or prior spring standardized state test; or were otherwise identified as "at-risk" based on absenteeism, disability status, English proficiency, and/or homelessness.[7] Students who did not meet these criteria but still wanted to participate were redirected to district-sponsored enrichment camps (e.g., LEGO robotics, Native/Indigenous Summer Culture Camp), with the goal of reserving the primary academic program for those students with the highest need.

A subset of schools (i.e., "hub sites") hosted summer school programs and served students from their school and up to two additional neighboring schools. The programs generally ran for 18 days in 2022 and 19 days in 2023.[8] The program operated a full day (7:45am to 2:30pm) for elementary students and a half-day (9:00 am to 1:30 pm or 10:15 am to 3:30 pm) for middle and high school students. GCS designed summer school to combine academic and enrichment content. Elementary students spent two hours each on literacy and math each morning, then moved to social studies or science and enrichment activities like art, music, or physical education in the afternoon. Middle school students spent an hour and a half on literacy or social studies, an hour on math, and an hour on science each day. High school students took a course for an hour and fifteen minutes in the morning and spent the afternoon on credit recovery or electives, ranging from an hour and a half to three hours.

[7] Despite the 40th percentile cutoff for summer school targeting, we were unable to conduct a regression discontinuity (RD) analysis due to a weak first stage, driven in part by the multiple criteria for targeting and limited summer school take-up among invited families. We were similarly unable to conduct an RD analysis of school-year tutoring, despite guidance from the district around targeting the lowest 20th percentile of students, due to the significant role that teacher and school discretion played in assignment, also resulting in a weak first stage.
[8] A small number of schools operating on alternative extended school day calendars offered 12-day programs.

4. Data and Measures

We obtained administrative data on all enrolled students in grades K-12 from 2016-17 through 2023-24.[9] These data contain information on students' demographic characteristics, school enrollment and grade level, course grades[10], monthly attendance, office referrals for behavioral incidents, and suspensions and expulsions. They also include students' participation in school-year tutoring and summer school from 2021-22 through summer 2023, including session dates and duration in hours. Our data also include interim exam scores from Kindergarten through 8th grade on the NWEA MAP Growth test and DIBELS. We supplement these with scale scores on North Carolina's end-of-grade (EOG) tests from 2016-17 through 2022-23, excluding 2019-20.

4.1 Outcome Measures and Non-Test Factor Index

We examine three categories of non-test outcomes: attendance, discipline, and course grades. For attendance, we measure the number of days absent (days enrolled minus days present) and an indicator for chronic absenteeism, defined as missing at least 10 percent of instructional days. For discipline, we use the number of disciplinary referrals, the number of suspension days resulting from a referral, and indicators for receiving at least one referral or at least one suspension. Because absences, referrals, and suspension days all have right-skewed distributions with a high frequency of zeros, we log transform these continuous measures (adding one) in our regression models (Backes et al., 2024). The district awards course grades to students

[9] This work evolved from a research practice partnership (RPP) between GCS and researchers from American Institutes for Research (AIR), Harvard University, and NWEA. This partnership, which launched in 2021-22 with the help of funding from AIR, is part of the larger Road to COVID Recovery (R2R) project, a multi-year partnership between the same group of researchers and a coalition of eleven school districts around the country focused on exploring the impact of COVID academic recovery interventions on students' interim test scores.
[10] Complete course grade data is not available for 2016-17 and 2017-18.

in grades 3 to 12. We calculate an unweighted average of each student's math and English Language Arts (ELA) grades.

For RQs 1 and 2, we examine trends in each outcome separately and explore their relationship to test scores. All measures span seven years, from 2016-17 through 2023-24, except for course grades, which we observe starting in 2018-19. However, two measurement breaks limit year-over-year comparisons. First, GCS changed its disciplinary referral recording system at the start of 2019-20. District officials report that the implementation ramp up distorted how teachers and administrators recorded referral counts. We therefore treat referral rates as incomparable across this transition, though more serious outcomes like suspensions, which the recording change did not affect, remain comparable throughout. Second, the pandemic onset in 2019-20 compromised the continuity of grading practices across the district. We therefore exclude 2019-20 grades from our analysis of course performance trends.

For RQ3, in which we estimate the impact of recovery interventions, we additionally construct a behavioral index measure of multiple outcomes using exploratory factor analysis, following the lead of McEachin, Welsh, and Brewer (2016), Jackson (2018) and Backes et al. (2023), which we refer to as the Non-Test Factor index (NTF). These earlier papers used non-test measures of student outcomes (e.g. student absences, suspensions, course grades, and grade retention) as components of their index. Like these earlier papers, we use the natural log (+1) of the number of days absent, the natural log (+1) of the number of days suspended, math and ELA GPA, and an indicator for a student repeating the same grade the next year. We additionally include the natural log (+1) of the number of behavioral referrals received, though the inclusion of this added variable does not meaningfully change our results.[11]

[11] We use all students enrolled in grades 3-9 to estimate the factor model using the Bartlett scoring method. The factor coefficients are days absent (-0.30), days suspended (-0.51), number of behavioral referrals (-0.56), GPA

4.2 Analytic Sample and Descriptive Statistics

For RQs 1 and 2, we include all GCS students in grades 3-8 from 2016-17 through 2023-24. For RQ3, we include GCS students enrolled in grades 3-9 in the year following each intervention (grades 2-8 during the intervention itself). Although interventions were available for kindergarten and first grade, we do not estimate the program impact for these younger students because course grades are assigned only in grade 3 and above.

We further restrict our RQ3 sample to include only those students with non-missing pretreatment NTF measures. For example, in models where we estimate the effects of 2021-22 school-year interventions, we require non-missing NTF measures in 2020-21; in Summer 2022 intervention models, we require non-missing NTF measures in 2021-22.

Table 1 reports descriptive statistics for the analytic sample used for models estimating SY 2021-22 tutoring impacts, as well as for the full GCS student population from grades 3-9 in 2022-23.[12] The full student population (n=38,864) and the analytic sample (n=29,817) look observably similar across rates of participation in recovery interventions, demographics, prior test scores, non-test measures in the outcome period (SY 2022-23), and non-test measures in the baseline period (SY 2020-21). In our analytic sample, 9% of students received at least some tutoring in 2021-22, 10% attended summer school in 2022, 17% received at least some tutoring in 2022-23, and 5% attended summer school in 2023.

In 2022-23, students in our analytic sample missed an average of 15.4 school days, or approximately 8.5 percent of all instructional days, and 27 percent were chronically absent,

(0.34), and grade retention (-0.10). For grades 1-2, we estimate a factor model that excludes GPA given the lack of course grades in these younger grades; here, the factor coefficients are days absent (-0.07), days suspended (-0.78), number of behavioral referrals (-0.78), and grade retention (-0.03). We standardize the index measure at the year level.

[12] See Appendix Table A1 for descriptive statistics of analytic samples for the other intervention periods: summer 2022, SY 2022-23, and summer 2023.

missing 10 percent or more of instructional days. They received an average of 0.66 disciplinary referrals per student and 21% of students received at least one referral over the course of the year. Common types of behavioral referrals included fighting, disruptive behavior, cutting school/class, and non-violent aggressive behavior. GCS suspended nine percent of students in the analytic sample at least once. Students averaged a 2.52 GPA across math and reading courses at all grade levels, and approximately 1% repeated a grade. Students in 2023-24 had similar characteristics (see Appendix Table A1).

Table 2 compares the subsets of students who participated in each of the academic recovery programs to the full analytic sample. For this table and throughout all impact estimates, we define students' grade level as their grade during the year following the intervention.

Most participants in school-year tutoring were in elementary school. Summer school participation was more evenly split across elementary and middle school. In summer 2023, high school participation (i.e. rising 9th grade students) was very low and virtually zero in our analytic sample. Generally, students participating in school-year tutoring were demographically similar to the overall analytic sample, whereas summer programming participants were more likely to be ELL, special education, Black, or Hispanic. When comparing non-test measures at baseline (i.e., in the year prior), intervention participants averaged more days absent than the overall sample. Students who received tutoring had similar GPAs to the overall sample, while students who received summer programming had considerably lower GPAs. The district issued a similar or lower number of disciplinary referrals and days suspended at baseline to tutoring participants compared to the sample average, but higher numbers for summer school participants. Baseline discipline-related measures were higher for the 2022-23 school year and 2023 summer participants than for the 2021-22 school year and 2022 summer participants, reflecting the overall decrease in school discipline incidents during the partially remote 2020-21 school year.

5. Methods

5.1 Trends in Non-Test Outcomes

Research on trends in academic achievement before and after the pandemic has found that the most negatively impacted students were Black and Hispanic students, as well as students attending high poverty schools (Goldhaber et al., 2022). To set the context and understand whether similar patterns hold among non-test outcomes in GCS, for RQ1 we document descriptive trends in measures of attendance, disciplinary referrals, and suspensions by estimating the following regression model:

$$Y_{it} = \alpha Grade_{it} + \beta StudChars_{it} + \gamma Race_i + \tau_t + \delta Race_i * \tau_t + \epsilon \qquad (1)$$

Where Y_{it} is a non-test outcome for student *i* in year *t* (e.g. number of absences); $StudChars_{it}$ is a vector of indicators for English language learner, student with disability, and gifted status; $Race_i$ is a categorical variable indicating students' race; τ_t are year fixed effects; and $Race_i * \tau_t$ is a race-by-year interaction.[13] We then plot predicted values of the non-test outcome for a student in the average grade and with average characteristics, allowing for different trends by race and year. This approach allows us to isolate average trends by racial group from any potential changes driven by shifts in the basic composition of enrolled students over time (for example, shifts in the proportion of elementary school students or special education students). We use the same approach to estimate annual trends by school poverty level, defined as the school's percent of students eligible for free or reduced-price lunch.

[13] The results of this model are robust to including students' math and reading test scores from the prior spring (in year *t*-1) as an additional control. We have elected to use the more parsimonious model as our preferred model. The results that include prior test scores as a control are available by request.

5.2 Relationships Between Non-Test Outcomes and Test Scores

For RQ2, we examine the association, by year, between students' non-test measures and their end-of-grade (EOG) performance on state standardized tests to understand how this relationship may have shifted before and after the pandemic. For attendance and disciplinary measures, we plot this relationship for all years from 2016-17 through 2022-23; for course grades, our data begin in 2018-19. In addition to graphically plotting unadjusted means, we estimate regression models of students' non-test outcomes on their same-year EOG score percentile, including year fixed effects and interactions of the percentile score with year, to more precisely quantify the relationships and their changes over time.

5.3 Estimating Non-Test Impacts of Academic Recovery Interventions

For RQ3, we use value-added models to estimate the impact of participating in each COVID recovery intervention on students' non-test outcomes, using outcomes during the school year prior to the intervention as a baseline and outcomes during the school year following the intervention as the dependent variable. We focus on estimating the effect of participating in any amount of each intervention (i.e. in at least one session), rather than dosage effects, out of concern that the number of sessions a student attends is possibly correlated with unobserved differences in student motivation or ability (see Carbonari et al., 2024 for further discussion of this concern). For evaluating school-year tutoring, our models take the following general form:

$$Y_{ijt+1} = \beta_0 + \beta_1 Intervention_{it} + \beta_2 OthInterv_{it} + \sum_{k=1}^{3} \lambda_k baselineNTF_{it-1}^{k} * Grade_{it} + \tau X_{ijt} * Grade_{it} + \psi SchlGrade_{ijt} + \epsilon_{ijt+1} \quad (2)$$

Y_{ijt+1} is a non-test outcome measure for student *i* in school *j* in school year *t*+1. The main treatment variable is $Intervention_{it}$, an indicator for student *i* participating *in at least one session* of a given intervention during school year *t*, where *t* is inclusive of the following summer. $\widehat{\beta_1}$, our main coefficient of interest, is therefore our estimate of the effect of an intervention on a non-test outcome in the *following* school year. $OthInterv_{it}$ is a vector of indicators equal to one if student *i* participated in any other recovery interventions or related programming (e.g. voluntary online tutoring, summer school) that occurred between the baseline and outcome time periods, where the baseline is the school year *t-1*. We control for $baselineNTF_{it-1}$, or student *i*'s non-test factor index from the year prior to the intervention year, as a cubic polynomial function. X_{ijt} is a vector of student characteristics[14] at the start of school year *t* and prior test scores, including indicators for student *i*'s gender, race, English learner status, special education status, gifted status, cubic functions of the two most recent prior MAP Growth scores in math and reading, and cubic functions of the most recent prior DIBELS and EOG math and reading scores.[15] We interact both $baselineNTF_{it-1}$ and X_{ijt} with student grade. $SchlGrade_{ijt}$ are school-by-grade fixed effects, and ϵ_{ijt+1} is an idiosyncratic error term. Lastly, we cluster standard errors at the school-by-grade level.

Our models evaluating summer school differ from what is described above only in that we use the school year immediately preceding the relevant summer (i.e. school year *t*) as our baseline time period. For example, for summer school 2022, we control for: non-test measures from the 2021-22 school year; participation in any other interventions that occurred in summer 2022; student characteristics and test scores from 2021-22; and school-by-grade fixed effects.

We estimate several versions of these models. We begin by estimating the effects of interventions on the factor-based measure, NTF. We then estimate the model using individual

[14] We are unable to control for students' eligibility for free or reduced-price lunch given limitations in data access to that variable.
[15] We condition our analytic samples on having non-missing NTF index measures in the baseline time period, but included students may have missing MAP Growth, DIBELS, and/or EOG scores. To address this, we impute missing values and create missing indicators for each type of test score, which we interact with the functions of prior test scores in the regression model.

measures of attendance, discipline, and course grades[16] as the dependent variable to tease out which outcomes might be particularly sensitive to the interventions. While our primary specification includes students' NTF at baseline as a control, we also estimate auxiliary regressions that control for a baseline measure of the specific outcome being predicted instead. For example, in models where the outcome is a binary indicator of chronic absence, or a continuous measure of the number of absences, we control for a cubic polynomial function of attendance rate in the baseline year.

As detailed in Equation #2, the terms that we consider to be part of the baseline period and outcome period depend on the particular intervention we are evaluating. **Figure 1** illustrates these relevant time periods. For each intervention, we estimate effects on outcomes *during the following school year*, where the non-test outcomes and NTF index are measured over the course of a school year from fall through spring. For interventions occurring during SY 2021-22 and Summer 2022, the dependent variables are outcomes from 2022-23. For SY 2022-23 and Summer 2023 interventions, the dependent variables are outcomes from 2023-24.

One potential concern with our models is that the temporal proximity to the COVID pandemic itself and the disruptions it caused to students' non-test outcomes may limit our ability to isolate changes in these outcomes related to intervention participation. While our empirical approach of comparing treated and non-treated students in the same time periods alleviates that concern to some degree, it is possible that non-test outcomes in the baseline periods we use—particularly the 2020-21 school year—may simply be too abnormal to provide useful baseline

[16] Because course grades are issued starting in 3rd grade, for models estimating effects on course grades (i.e. GPA), we limit the analytic sample to exclude grades 3 and 4 (in the outcome year). Since these students would have been too young to receive course grades in the baseline year for school-year tutoring models (i.e. two school years prior), they would be missing the most relevant baseline control variable. For consistency, we implement this rule across models for school-year tutoring and summer school.

controls. To some extent, proximity to the highly disrupted years of pandemic era education is unavoidable. However, we do attempt to address the issue of 2020-21 acting as a questionable baseline by additionally estimating semester-level models for tutoring that took place in spring 2022, allowing us to use fall 2021 as the baseline (and fall 2022 as the outcome period). Results from these semester-level models are consistent in general magnitude and statistical significance with the school-year level models that we report in this paper.[17]

The choice of relevant time periods to be included in baseline and outcome measures is not a trivial one. Examining an intervention's impact on non-test outcomes during the period of the intervention is also a valuable research question, as demonstrated by Lee et al.'s (2024) study of the contemporaneous relationship between absences and tutoring in which they find that students are less likely to be absent on the days they have tutoring. We structure our analysis to focus on effects in the period *following* the intervention, however, for multiple reasons. First, Lee and co-authors' findings suggest students feel motivation or pressure to show up for their assigned tutoring sessions, but it is not clear whether this subsequently leads to a sustained improvement in attendance. A change in behavior that is evident beyond the period of time in which an intervention takes place could speak more to concerns about the lasting impacts of COVID on students' engagement and connection with school and could arguably be a more meaningful intervention outcome.[18] This would be analogous to research on the "long-run"

[17] Results from semester-based models are available upon request.

[18] One potential concern with this distinction between contemporaneous and sustained effects of an intervention is the possibility that students who participate in an intervention one year may be more likely to participate in it a second year as well. If this were the case, what we interpret to be sustained effects of an intervention could in fact be students' response to contemporaneous treatment in the outcome school year. To assess the likelihood of this scenario, we run regression models of a similar form to our main value-added models, where the outcome is an indicator for participating in school-year tutoring in the year following the intervention of interest, allowing us to estimate the predictiveness of treatment in time t on treatment in time $t+1$. Because we do not have data on tutoring participation in 2023-24 at the time of analysis, we are only able to run these models where 2021-22 tutoring and 2022 summer school are the time t predictors. Our results, shown in Appendix Table A2, indicate that treated students are more likely to be treated in the outcome year, but this only appears to be the case for elementary school

value-added of teachers, which finds that measures of teachers' contributions to student test scores that persist to the following year are more predictive of student long-run outcomes than standard (i.e. contemporaneous) value-added measures (Gilraine and Pope, 2021).

Secondly, according to our interviews with administrators in GCS (and confirmed empirically), students in GCS are assigned to tutoring based largely on teacher discretion if they see that a student is struggling, suggesting that their struggles during a given school year could be the impetus for their enrollment in that same school year's intervention. To avoid confounding outcome measures with measures determining treatment enrollment, we estimate impacts to the school year following the intervention. Lastly, to evaluate summer programming, we assess outcomes in the following school year as most of our non-test outcome measures do not apply (or would be notably different) during the summer.

A central concern with using value-added models to estimate the impact of recovery interventions is the potential for bias associated with selection into treatment. Our models control for observable characteristics and measures of academic and non-test performance in an attempt to limit selection bias; however, the possibility remains that characteristics or situations unobservable to us—for example, a student struggling academically or socio-emotionally for a short period of time, or within the time period of the intervention—play a role in assigning a student to treatment, resulting in negative selection bias. This is particularly true given the role that teacher and school discretion play in enrolling students in tutoring. There is also the possibility of positive selection into treatment—for example, students who are more motivated or whose families are more engaged opting into tutoring or summer programming. Because of the

students. Given that our significant findings are not limited to elementary students, this scenario seems unlikely to be a main driver of our results.

difficulty in fully eliminating the possibility of bias, our work here should be interpreted as suggestive of the effects of academic interventions, but not necessarily causal. We do, however, use the terms "impacts" and "effects" in this report for simplicity of wording.

A different concern arises from the fact that we are focused on non-test outcomes, which are likely to be more subjective measures (than tests). For example, evidence shows that there is significant variation across teachers and principals in their disciplinary practices and the number of behavioral referrals and suspensions they issue to students (Liu et al., 2023; Sorenson et al., 2022). Perhaps the outcome most influenced by teachers are course grades, which are themselves determined by teachers and can vary based on teacher experience and expectations (Link, 2018; Gershenson, 2020). To try to account for the variation in non-test outcome measures that are likely associated with these factors, we include school-by-grade fixed effects (using the school and grade in the outcome period following the intervention) in our models so that we are comparing the outcomes of treated and non-treated students within each grade in each school.[19] Similarly, we use outcome measures that are in a different school year than students' treatment assignment to ensure that in most cases, the teachers assigning grades are not the same teachers who helped determine that a student should be prioritized for intervention.

5.4 Subgroup Models and Additional Robustness Checks

In addition to the estimation models for RQ3 described above, we explore whether there is evidence of heterogeneity by estimating models separately by gender, race, English learner status, disability status, and gifted status. We also conduct robustness checks to help test the validity of our empirical strategy. We run placebo tests in which we assign students to treatment

[19] To more fully account for teacher-specific influences on non-test outcomes, we could theoretically control for school-by-classroom fixed effects. We chose not to specify our models in this way because of concerns about reductions in statistical power given the small number of students who may be assigned to an intervention in any given classroom.

via random assignment and estimate treatment effects based on this "fake" treatment status rather than students' true treatment status. Reassuringly, these models arrive at null results, supporting the assertion that our models are not simply picking up spurious trends. We also vary our model specifications as further robustness checks. Specifically, for all models where our dependent variable is a binary measure (e.g. chronically absent), we run probit models in addition to linear probability models and arrive at consistent results. We also estimate versions of our models that include multiple lagged NTF measures as a control (i.e. NTF_{it-1} and NTF_{it-2}) for all interventions except for SY 2021-22 tutoring, for which NTF_{it-2} would correspond to 2019-20 and the onset of COVID, and again arrive at consistent results.[20]

6. Results

6.1 Trends in Non-Test Outcomes

Figure 2 shows trends in average measures of absenteeism (panel A) and percent of students with a disciplinary referral or suspension (panel B) in GCS from 2016-17 through 2023-24. These estimates are from regressions described in Equation 1 (excluding the interaction term that would result in separate trends by race or poverty level). Because GCS changed their system of reporting disciplinary referrals in 2019-20, measures before and after the system change may not be comparable. Although the new reporting system theoretically did not affect data on severe disciplinary outcomes such as suspensions, we indicate the timing of the system change in the figures for both outcomes as a visual reminder.

As the 95% confidence intervals indicate, there is little ambiguity that overall measures of absenteeism have increased significantly in GCS since before the pandemic. The chronic absence rate increases from 16% in 2019 to 25% in 2021. And while it peaked in 2022 at 33%, it

[20] Results from all models with alternative specifications are available upon request.

remains significantly higher than pre-pandemic years. The continuous measure of days absent follows a similar pattern.

Though less dramatic, disciplinary measures are also on an upward trajectory; referral rates go from 17% in 2022 to 21% in 2024. Notably, the pandemic caused the greatest disruption to absences and disciplinary measures at different times: absence measures plummeted in 2019-20 when school was abruptly moved online, which generally impacted daily data collection and the strictness of attendance policy (Schwartz et al. 2021), whereas disciplinary measures plummeted the following year when schooling continued to be remote, consistent with national trends (Belsha, 2023).[21]

Figure 3 breaks out trends in absenteeism by race (panel A) and school poverty (panel B). The continuous and binary measures of absenteeism show similar trends, though the chronic absence rate reveals more separation across student subgroups. Prior to the pandemic, Asian students had the lowest chronic absence levels, while Black students and those in the other race category had the highest. The chronic absence rate plummeted for all subgroups in 2019-20 and grew sharply in the years following. Notably, the racial gaps in attendance grew considerably in the post-pandemic years, with chronic absence among Black, Hispanic, and other students at about 10-15 percentage points higher than White students in 2020-21, as compared to gaps in the 2-3 percentage point range in the pre-pandemic period. Though attendance improved for nearly all groups by 2022-23 and 2023-24, the gaps remained larger than in pre-pandemic years and all groups had approximately twice the rate of chronic absence that they had prior to COVID. This overall pattern is similar for trends broken out by school poverty quartile, with the highest rates of absenteeism among students attending higher poverty schools and increasing gaps across the

[21] General trends in absenteeism and disciplinary measures are robust to conditioning on student test scores, both for the overall student population and by student subgroup as shown in Figures 3 and 4. See footnote 13.

levels of disadvantage post-pandemic. In 2022-23 and 2023-24, even after some improvement in attendance among most quartiles, about a third of students in higher poverty schools (3rd and 4th quartiles) were chronically absent compared to about a fifth of students in low poverty schools (1st quartile).

Figure 4 shows trends in disciplinary referrals and suspensions, again broken out by race (panel A) and school poverty (panel B). Following the sharp decline in 2020-21 (when they were close to zero), average suspension rates were only slightly elevated for most racial groups post-pandemic compared to pre-pandemic rates; a notable exception, however, is Black students, whose average suspension rates grew from about 9 percent in 2019-20 to about 13 percent in 2021-22 and remained roughly constant in the years following. Racial gaps in disciplinary referral rates also appear to be growing as of 2022-23. When broken out by school poverty level, the lowest poverty schools have suspension and discipline rates far below their more disadvantaged counterparts in the post pandemic years. Interestingly, in 2021-22, suspension and disciplinary referral rates are highest among third quartile schools (FRPL between about 70% and 78%). By 2023-24, rates for both outcomes are roughly similar amongst the 2nd, 3rd, and 4th quartiles of schools, all of which still far exceed those for the lowest quartile.

In summary, absenteeism, disciplinary issues, and course grades in GCS all worsened during the pandemic, with chronic absence showing the largest and most persistent increases. These challenges disproportionately affected students of color, students in higher-poverty schools, and lower-achieving students, widening pre-existing gaps.

6.2 Relationships Between Non-Test Outcomes and Test Scores

Figure 5 plots the average percent of students who are chronically absent and the average percent who have at least one suspension by their EOG performance at the end of that school year, by school year. Pre-pandemic data points are shown in grey; data points from 2020-21 onward are shown in color. For simplicity, only EOG scores in math are shown, though patterns are similar across the distribution of ELA scores. Consistent with Figures 2 and 3, chronic absence and suspension rates increased following the pandemic (apart from suspension rates in 2020-21, which were virtually zero). Moreover, they increased for all deciles in the EOG score distribution, though the increase in suspension rates among high-performing students is minimal. Also notable is that the slope of the association between non-test outcomes and test scores became steeper in post-pandemic years, suggesting a stronger relationship between the two and concerningly high levels of chronic absence (and to a lesser extent, suspensions) among students who go on to have low achievement on end-of-year exams.

Figure 6 is comparable to Figure 5 but shows the relationship between GPA of math and ELA courses and test scores in both math and reading, by school year. The GPA patterns in 2020-21 appear to be anomalous, as grades were abnormally lower in that year and the association between grades and reading scores in the low percentiles deviates from the rest of the distribution. But when comparing associations in the other years, namely 2018-19 with 2021-22 and 2022-23, two key findings emerge. The first is that GPAs decreased overall in GCS since prior to the pandemic. The second is that GPA decreased at nearly all deciles of the achievement distribution, not just among low-achieving students, though declines appear to be the largest among students who score between the 35^{th}-50^{th} percentiles on EOG exams. As noted previously, this is one non-test outcome for which findings on post-pandemic trends have differed across states. Fuller et al. (2023) find that GPAs dropped statewide in North Carolina in 2020-21 and rebounded only partially the following year, consistent with GCS; meanwhile, Goldhaber and Goodman Young (2024) find that grades increased post-pandemic in Washington state and remained elevated in math in 2021-22.

To quantify these associations, we estimate regressions of non-test outcomes on EOG test scores, year, and year-by-test score interactions. The results, shown in **Table 3,** corroborate the patterns shown in the plots. As expected, there is a negative association between both chronic absenteeism and receiving suspensions with test scores, and a positive association between GPA and test scores. The results also indicate that chronic absence was greater in post-pandemic years compared to pre-pandemic, as were suspensions (except for 2020-21), and that GPA was lower in all post-pandemic years compared to 2018-19. Finally, the results show that the associations between non-test and test outcomes grew stronger after the pandemic—for absences and suspensions, they became more negative, and for GPA they became more positive— though the coefficients on the year-by-EOG interaction terms are not always statistically significant.

6.3 Recovery Intervention Impact Estimates

We turn next to the estimated impacts of recovery interventions on non-test outcomes (RQ3). **Table 4** shows results for a range of model specifications where the intervention is school year tutoring in 2021-22 and the outcome is the non-test factor index (NTF). The specifications gradually include more controls up until the final specification (column 7), which is our preferred specification as described in Equation 2. Specifically, the first model controls for only student demographics, the second for only baseline NTF, the third for demographics and baseline NTF, then prior test scores are added in, then school fixed effects, then grade fixed effects, and then school-by-grade fixed effects. The results illustrate how the estimated treatment participation coefficients go from negative and statistically significant to positive (though not statistically significant) with the gradual inclusion of additional controls, demonstrating a reduction in the selection bias of the treatment effect estimates.

Estimates for the effects of school year tutoring in 2021-22 and 2022-23 on all non-test outcomes, using our preferred full specification, are reported in **Table 5**. For each intervention, we show results for all students in grades 3-9 in the outcome year, as well as broken out by elementary, middle school, and high school. We find that the SY 2021-22 tutoring had positive but not statistically significant effects on students' NTF measure in 2022-23 across all grades, but the effects were positive and significant (a gain of .073 SDs) for middle school students.

To better understand which outcomes drive the effect on the summative NTF measure, we examine the results for the individual non-test outcomes. We find that 2021-22 tutoring did not lead to changes in the likelihood of being chronically absent or in the number of absences in the following period. However, we do find that participating in 2021-22 tutoring led to statistically significant declines in the number of disciplinary referrals among middle school students and in the number of days suspended across all grades, though primarily among middle school students (the magnitude of the treatment coefficient for high school students is greater than that for middle school, but not significant). Among middle school students, participants saw a roughly 5% reduction in the number of referrals per student (going from an estimated 1.01 referrals per participating student to 0.96) and nearly a 6% reduction in the number of days suspended (going from an estimated 0.66 days suspended to 0.62). Finally, we do not find a detectable impact of 2021-22 tutoring on GPA in the following year.

Our results for SY 2022-23 tutoring are different. Interestingly, we find no significant impacts of 2022-23 tutoring on disciplinary referrals or suspensions for participants in the following year, but we do find positive and significant effects on participants' future attendance.

Again, impacts to middle school students are the strongest. We estimate that middle school tutoring participants had a 3.4 percentage point drop in the likelihood of being chronically absent and a nearly 5% reduction in the number of days absent in the year following tutoring. The estimated impacts correspond to a decrease in chronic absence rates among middle school participants from 0.32 to 0.29 and a decrease in days absent from 16.2 to 15.5. Though these impacts may appear small in magnitude, they are proportionally greater than the estimated impacts found in DC of the impact of a high dosage tutoring program on attendance in the *same year* as the program.[22]

Despite the estimated positive impacts to attendance, we find negative impacts to GPA that are statistically significant for all grades combined (-0.041) and negative but non-significant for each grade band of students.[23] In auxiliary models, we estimate effects to math and reading GPAs separately and find that this negative result is driven primarily by negative impacts to course grades in math. While these results are somewhat puzzling—given the explicit intent of the tutoring to improve students' academic performance—they are broadly consistent with earlier research findings from Carbonari et al. (2024) of the effects of GCS' (and other districts') 2022-23 tutoring programs on student test scores. That study found negative effects of GCS's intervention on end-of-year math test scores (and null effects for reading), suggesting either that unobservable negative selection biased the estimates downward, or that participating in the tutoring actually decreased academic performance relative to non-participating students. This

[22] Researchers studying the DC high dosage tutoring program estimated that each day of tutoring resulted in a 1.2 percentage point decrease in the likelihood of being absent that day (Lee et al. 2024). They reasoned that tutored students would attend, over the course of a year, an increased average of 1.2% multiplied by the total number of days tutored. In GCS, students attended an average of 27 sessions of tutoring a year. Assuming sessions each occurred on a different day, the results in DC would correspond to 0.32 (27*.012 = 0.32) fewer days missed during the year tutoring took place, or approximately half of the reduction in days absent that we estimate for the following year.

[23] As noted in footnote 16, we exclude grades 3 and 4 from models where GPA is the outcome; the elementary grade band therefore consists only of students in grade 5 in the outcome year.

could have been the case if, for example, tutoring pulled students out of core math courses or otherwise useful group instruction. Together, our results for GPA and the results from Carbonari et al. (2024) suggest that while the tutoring program may have benefitted students' behavior as measured by attendance, it was unable to measurably benefit their academic performance.

Table 6 is comparable to Table 5 but shows results from models estimating the effect of summer school in 2022 and 2023. In 2022, estimated impacts to students' NTF measure are positive and statistically significant overall (0.05 SD across grades 3-9) and for elementary students (0.06 SD). The primary drivers of the positive overall effects are improvements in attendance among summer program participants. Summer school 2022 attendees saw an overall 6 percentage point reduction in the likelihood of being chronically absent, with a 7 percentage point estimated impact for elementary school students and a 6 percentage point estimated impact for middle school students (the estimate for high school students is not significant). The overall 6 percentage point decrease suggests an approximately 17% decline in chronic absence rates for participants, going from an estimated 35% to an observed 29% in 2022-23. This effect size appears large in comparison to results from pre-pandemic research on summer school, which found an average decrease in chronic absenteeism of 1.4 percentage points; however, the average chronic absence rate of non-participants in that context was only 8.3%, implying a similar 17% reduction in chronic absence associated with the program (Pyne et al., 2021). For the number of days absent, we find that summer school led to a 13% decrease in days missed overall (suggesting 2.3 more days attended), with a 13% decrease for elementary students, a 14% decrease for middle school students, and 9% decrease for high school students.

For summer school 2022, we find a negative and statistically significant estimated impact on math and reading GPA (-0.06 points), driven largely by the negative impact estimate for high school students (i.e. 9th grade students, -0.15 points). In auxiliary models that look at math and reading GPAs separately, we find that the estimates are similar in magnitude for both subjects. Unlike GCS' school-year tutoring, which earlier research found to have a negative impact on math student test scores (Carbonari et al. 2024), the district's summer school was found to positively impact math test scores and have a null effect on reading scores (Bolyard et al. 2025), making this result more difficult to reconcile. We further interrogate which students are the primary drivers of the negative finding by estimating our value-added model separately by quartile of baseline GPA. 75 percent of 9th grade summer school 2022 participants fall in the first quartile and 18 percent fall in the second. We find that high school students in the second quartile (with baseline GPAs in the 1.78 to 2.73 range) are the main drivers of the overall finding, with an estimated impact estimate of -0.15 (not statistically significant) compared to -0.036 for the first quartile. Interestingly, these second quartile students also see a positive (not statistically significant) impact on the share of math or English courses they take that are advanced, accelerated or honors, suggesting that participating in summer school may also have led them to enroll in more difficult classes the next year. This difference in the course enrollment patterns of summer school participants and non-participants is also reflected in the estimated impact on second quartile students' *weighted* GPA (-0.078), which is considerably lower than the estimate for the standard GPA.

While this further exploration provides helpful context, it cannot fully explain our high school GPA results. Other potential factors include the fact that, according to district interviews, the high school program in particular was used for many students as an opportunity for credit recovery, suggesting that there may have been more negative selection among the older student sample. There is also the possibility that summer school, like many educational interventions, is more effective for younger students than for teenagers (see Guryan et al., 2023), or that something about the transition to high school reduces potential non-test benefits of summer programming for students.

Our findings for 2023 summer school are broadly similar to those for summer 2022. Because of the extremely low participation of rising 9th grade students in summer 2023 (n=1), we estimate effects only for elementary and middle school. Again, we find that summer school participation reduced students' chronic absence rates and number of days absent, though results are only statistically significant for elementary students and the all-grades sample.[24]

In addition to our preferred models that control for baseline measures of students' NTF measure, we also estimate models that control for baseline continuous measures of the individual outcome in question in lieu of the NTF. Results for these models are shown in **Appendix Tables A3 and A4**. Generally, the overall conclusions from these alternative models are similar. We find that participants in 2021-22 tutoring experienced improvements in the number of disciplinary referrals and suspensions received the following year, driven largely by middle schoolers. 2022-23 tutoring led to reductions in absenteeism. We again find that summer school in 2022 and 2023 also led to reductions in absenteeism, though we only detect significant impacts in the continuous measure of the number of absences for the first year and in the discrete

[24] For additional insight into whether impacts of tutoring or summer school are concentrated in the early part of the outcome year (i.e. soon after the intervention took place) or are sustained throughout the full school year, we estimate the effects of each of the four interventions on non-test outcomes in the following year, separately by fall and spring semester. Our findings suggest that the pattern of when impacts are detected varies. For 2021-22 tutoring, reductions in suspensions appear to be more concentrated in the spring of the following year. For 2022-23 tutoring, on the other hand, the reduction in days absent is only apparent in the fall. For 2022 summer school, reductions in chronic absenteeism and number of absences are significant and similar in magnitude across both semesters the following year; the same is generally true for summer 2023, though chronic absence impacts are only significant in the fall semester. At least for summer programming, these findings suggest a sustained change in behavior that lasts at least through the course of the year.

chronic absence measure for the second year. We again estimate a negative effect on GPA for high school students who attended summer school in 2022.

To investigate the potential for heterogeneity across student subgroups, we estimate our models for each subgroup separately; the results of these models are shown in **Appendix Tables A5 through A8**. Generally, we do not find much definitive evidence of heterogeneity, though there is suggestive evidence that some interventions had greater effects for some student subgroups. Male students, for instance, appear to have benefitted more than female students from tutoring in 2021-22. And we find evidence that Black and Hispanic students had greater reductions in absenteeism following summer school 2022 (14% and 15%, respectively, across all grade levels) than did Asian or White students. Additionally, Hispanic students had greater improvements in attendance, disciplinary referrals, and suspensions following 2022-23 tutoring compared to all other racial groups. We caveat these subgroup estimates by pointing out the large number of regressions run and the potential for arriving at statistically significant estimates due to chance.

7. Discussion

Presented with unprecedented challenges following the COVID pandemic outbreak, school districts around the country implemented recovery interventions aimed at getting students back on track academically. As other research has documented (Carbonari et al., 2022), districts had difficulty establishing and expanding these programs at a scale and pace necessary to meet student needs. Prior studies have shown that the interventions largely fell short of recovering test score declines (e.g. Carbonari et al., 2022; Callen et al., 2025; Carbonari et al., 2024), but a focus on test score recovery alone may limit our ability to fully evaluate the potential benefit of these programs. Research suggests that for many students and their families, their relationship with and connection to school fundamentally shifted after lengthy periods of remote instruction and disrupted learning (Myung and Hough, 2023). Although it may not always be the explicit end goal of academic recovery interventions, the potential for these programs to improve student engagement may be a critical component of their value.

The results of our study highlight this need for student engagement recovery in GCS, as we document substantial reductions across the pandemic in engagement as measured by attendance, disciplinary referrals, suspensions, and course grades. We also find that racial and socioeconomic disparities in attendance and disciplinary outcomes increased and, despite some improvement, remained elevated as of 2023-24. Additionally, we find that non-test outcomes worsened for students across the full distribution of test score achievement and that the association between test scores and non-test outcomes actually became stronger post-pandemic. Put differently, non-test outcomes declined while at the same time they became increasingly predictive of test scores.

Our study further finds that educational interventions such as tutoring and summer school may be able to help partially recover levels of student engagement lost during the pandemic. Participating in these interventions was associated with small but meaningful improvements in attendance and disciplinary outcomes in the year following, suggesting a sustained response to the intervention. The positive non-test impacts of summer school align to some extent with the findings of related work evaluating GCS' summer programs on test scores, which found positive impacts to participants' test scores, particularly in math (Callen et al., 2025; Bolyard et al., 2025). The findings for school-year tutoring, on the other hand, diverge more depending on the outcome; while we find positive effects on discipline following 2021-22 tutoring and on attendance following 2022-23 tutoring, the related research found null to negative effects on test scores (Carbonari et al. 2022; Carbonari et al. 2024). Interviews with district officials, tutoring directors, and principals in GCS revealed that school-year tutors often had to divert tutoring time away from academic topics to address their students' socioemotional needs, which can perhaps help explain the collective results.

As discussed earlier, one limitation of our study is its reliance on a selection on observables research design, which leads us to caution readers against interpreting our findings as necessarily causal. It is plausible that there is some degree of selection into treatment that we are unable to control for with our data. Our findings of a negative impact estimate of 2022-23 tutoring and 2022 summer school on participating students' GPA underscore this point, as one potential explanation for this result is that students were negatively selected into the interventions, biasing the treatment effect estimate downward. It is worth noting, however, that if this were the case, the positive effects we estimate on discipline and attendance would likely also be downward biased and could therefore be interpreted as conservative estimates.

On balance, our evidence suggests post-pandemic academic interventions in GCS were associated with some improvement in students' non-test outcomes. At the same time, these interventions do not appear to be a silver bullet capable of erasing serious declines in student engagement after the pandemic. As practitioners and researchers look forward beyond pandemic-era interventions to ongoing improvements in student outcomes, our findings point both to the potential of academic interventions to benefit students in multi-dimensional ways, and to the need for a nuanced view of their potential impacts. Recognizing the meaningful but limited non-test benefits of programs like tutoring and summer school may help demonstrate their broader value and inform program design going forward.

References

Adukia, A., Cong, B., Fan, J., Feigenberg, B., Kang, J., & Momeni, F. (2025, February 7). A restorative approach to student discipline shows promise in reducing suspensions and arrests. Brookings Institute. https://www.brookings.edu/articles/a-restorative-approach-to-student-discipline-shows-promise-in-reducing-suspensions-and-arrests/

Backes, B., Cowan, J., Goldhaber, D., & Theobald, R. (2024). How to measure a teacher: The influence of test and nontest value-added on long-run student outcomes. *Journal of Human Resources*, 1023-13180R2.

Barrett, N., McEachin, A., Mills, J., & Valant, J. (2021). "Disparities in student discipline by race and family income." Journal of Human Resources, 56(3), 711-748.

Bastian, K. C., and Fuller, S. C. (2024, February). Summer 2021 School Extension Programs – Elementary and Middle School Impacts Brief. Education Policy Initiative at Carolina.

Belsha, K. (2023, November 15). Suspensions and bullying plunged as many students learned remotely, national data shows. Chalkbeat. https://www.chalkbeat.org/2023/11/15/drops-in-suspensions-during-pandemic-federal-data-show/

Bolyard, A., DeArmond, M., Dizon-Ross, E., Goldhaber, D., Isaacs, J., Kane, T. J., Keating, M., Klinenberg, J., McDonald, A., McEachin, A., Morton, E., Muroga, A., Salazar, A., & Staiger, D. O. (2025). Summer school as an academic recovery strategy after COVID-19: Evidence from summer 2023. CALDER Working Paper No. 325-0825. https://doi.org/10.63302/XFRZ5691

Callen, I., Carbonari, M. V., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Isaacs, J., Kane, T. J., Kuhfeld, M., McDonald, A., McEachin, A., Morton, E., Muroga, A., and Staiger, D. O. (2025). Summer school as a learning loss recovery strategy After COVID-19: Evidence from summer 2022. *American Education Research Journal*, *62*(5), 982–1018.

Carbonari, M. V., Davison, M., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Hashim, A. K., Kane, T. J., McEachin, A., Morton, E., Muroga, A., Patterson, T., & Staiger, D. O. (2024). The impact and implementation of academic interventions during COVID-19: Evidence from the Road to Recovery project. *AERA Open*, *10*.

Carbonari, M. V., DeArmond, M., Dewey, D., Dizon-Ross, E., Goldhaber, D., Kane, T. J., McDonald, A., McEachin, A., Morton, E., Muroga, A., Salazar, A., & Staiger, D. O. (2026). Impacts of academic recovery interventions on student achievement in 2022 to 2023. *Educational Evaluation and Policy Analysis*, *0*(0).

Chang, H. and Romero, M. (2008). Present, engaged, and accounted for: The critical importance of addressing chronic absence in the early grades. National Center for Children in Poverty.

Davison, M., Penner, A. M., Penner, E. K., Pharris-Ciurej, N., Porter, S. R., Rose, E. K., Shem-Tov, Y., & Yoo, P. (2022). School discipline and racial disparities in early adulthood. *Educational Researcher*, 51(3), 231-234.

Dee, T. S. (2023, August 10). Higher chronic absenteeism threatens academic recovery from the COVID-19 pandemic. https://doi.org/10.31219/osf.io/bfg3p

Dewey, D., Fahle, E., Kane, T., Kuhfeld, M., & Reardon, S. F. (2024). Reconciling recent evidence on academic recovery in the US from 2022 to 2023. NWEA Research Report. https://www.nwea.org/uploads/Reconciling-Recent-Evidence-on-Academic-Recovery-in-the-US-from-2022-to-2023.pdf.

Dewey, D. C., Fahle, E., Kane, T. J., Reardon, S. F., Staiger, D. O. (2025, February). Pivoting from pandemic recovery to long-term reform: A district-level analysis. Education Recovery Scorecard. https://educationrecoveryscorecard.org/wp-content/uploads/2025/02/Pivoting-from-Pandemic-Recovery-to-Long-Term-Reform-A-District-Level-Analysis.pdf

Diliberti, M. K., Chu, L., Rainey, L. R., DiNicola, S. E., Lake, R., J., & Schwartz, H. L. (2025, Aug. 14). Chronic absenteeism still a struggle in 2024-2025. RAND. https://www.rand.org/pubs/research_reports/RRA956-34.html

Fahle, E., Kane, T. J., Patterson, T., Reardon, S. F., Staiger, D. O., and Stuart, E. A. (2023). School district and community factors associated with learning loss during the COVID-19 pandemic. Harvard University Center for Education Policy Research Working Paper. Retrieved from https://cepr.harvard.edu/sites/hwpi.harvard.edu/files/cepr/files/explaining_covid_losses_5.23.pdf

Fryer, R. G., and Howard-Noveck, M. (2020). High-dosage tutoring and reading achievement: Evidence from New York City. *Journal of Labor Economics*, 38(2), 421-452.

Fuller, S. C., Bastian, K. C., Swiderski, T., and Mikkelsen, C. N. (2023, June) Effects of the COVID-19 pandemic on student attendance, grades, grade retention, and suspensions in North Carolina through the 2021-22 school year. Education Policy Initiative at Carolina, EPIC Insights.

Fuller, S. C., McDonnell, J., Bastian, K. C., and Lauen, D. (2024, February). Summer 2021 School Extension Programs – High School Impacts Brief. Education Policy Initiative at Carolina.

Fuller, S. C., Swiderski, T., Mikkelsen, C., and Bastian, K. C. (2024). In school, engaged, on-track? The effect of the pandemic on student attendance, course grades, and grade retention in North Carolina. *Economics of Education Review*, *103*, 102621.

Gershenson, S. (2020). Great expectations: The impact of rigorous grading practices on student achievement. Washington, DC: Thomas B. Fordham Institute. Retrieved from https:// fordhaminstitute.org/national/research/great-expectations-impact-rigorous-grading-practicesstudent-achievement.

Gilraine, M., and Pope, N. G. (2024). Making teaching last: Long-run value-added. *Journal of Labor Economics*, *42*(3), 693-738.

Goldhaber, D., & Goodman Young, M. (2024). Course grades as a signal of student achievement: Evidence of grade inflation before and after COVID-19. *Journal of Policy Analysis and Management*, 43, 1270–1282.

Goldhaber, D., Kane, T. J., McEachin, A., and Morton, E. (2022). A comprehensive picture of achievement across the COVID-19 pandemic years: Examining variation in test levels and growth across districts, schools, grades, and students. CALDER Working Paper No. 266-0522.

Guryan, J., Ludwig, J., Bhatt, M.P., Cook, P. J., Davis, J. M. V., Dodge, K., Farkas, G., Fryer Jr., R. G., Mayer, S., Pollack, H., Steinberg, L., and Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review, 113*(3), 738–65.

Halloran, C., Hug, C. E., Jack, R., and Oster, E. (2023). Post COVID-19 Test Score Recovery: Initial Evidence from State Testing Data. National Bureau of Economic Research, WP No. w31113.

Hanushek, E.A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro-data. American Economic Review, 61(2), 280-288.

Hanushek, E.A. (1992). The trade-off between child quantity and quality. *The Journal of Political Economy* 100(1), 84-117.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072–2107.

Kuhfeld, M., & Lewis, K. (2022). Student achievement in 2021-22: Cause for hope and continued urgency. Center for School and Student Progress, NWEA Research. https://www.nwea. org/research/publication/student-achievement-in-2021-22-cause-for-hope-and-continued-urgency

Lee, M. G., Loeb, S., & Robinson, C. D. (2025). The impact of high-impact tutoring on student attendance: Evidence from a state initiative. (EdWorkingPaper: 24-1107). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/wghb-4864

Lewis, K. and Kuhfeld, M. (2023). Education's long COVID: 2022-23 achievement data reveal stalled progress toward pandemic recovery. Center for School and Student Progress, NWEA Research.

Link, L. (2018). Teachers' perceptions of grading practices: How pre-service training makes a difference. *Journal of Research in Education, 28*(1), 62-91.

Liu, J., Hayes, M. S., & Gershenson, S. (2024). JUE Insight: From referrals to suspensions: New evidence on racial disparities in exclusionary discipline. *Journal of Urban Economics, 141*.

Liu, J., and Loeb, S. (2019). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, *59*(1).

Liu, J., Penner, E. K., and Gao, W. (2023). Troublemakers? The role of frequent teacher referrers in expanding racial disciplinary disproportionalities. *Educational Researcher, 52*(8), 469-481.

Malkus, N. (2025). Lingering absence in public schools: Tracking post-pandemic chronic absenteeism into 2024. American Enterprise Institute. https://www.aei.org/wp-content/uploads/2025/06/Lingering-Absence-in-Public-Schools-Tracking-Post-Pandemic-Chronic-Absenteeism-into-2024.pdf?x97961.

McEachin, A., Lauen, D.L., Fuller, S.C., & Perera, R. (2020). "Social returns to private choice? Effects of charter schools on behavioral outcomes, arrests, and civic participation." *Economics of Education Review*, 76.

McEachin, A. J., Welsh, R. O., & Brewer, D. J. (2016). The variation in student achievement and behavior within a portfolio management model: Early results from New Orleans. *Educational Evaluation and Policy Analysis*, *38*(4), 669-691.

Myung, J. and Hough, H. J. (2023). Why aren't students showing up for school? Understanding the complexity behind rising rates of chronic absenteeism. [Commentary]. Policy Analysis for California Education. https://edpolicyinca.org/newsroom/why-arent-students-showing-school.

National Student Support Accelerator. (2024, March 1). Early findings show evidence that high-impact tutoring increases student attendance in D.C. schools. Retrieved from https://studentsupportaccelerator.org/news/early-findings-show-evidence-high-impact-tutoring-increases-student-attendance-dc-schools.

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, 26(5), 617-624.

Polikoff, M., Clay, I., and Silver, D. (2023). Beyond test scores: Broader academic consequences of the Covid-19 pandemic on American students. Center on Reinventing Public Education, Arizona State University.

Pyne, J., Messner, E., and Dee., T. S. (2022). The dynamic effects of a summer learning program on behavioral engagement in school. *Education Finance and Policy*, 18(1), 125-155.

Schwartz, H. L., Diliberti, M. K., Berdie, L., Grant, D., Hunter, G. P., and Setodji, C. M. (2021). Urban and rural districts showed a strong divide during the COVID-19 pandemic: Results from the second American School District Panel survey. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RRA956-2.html.

Shi, Y., and Zhu, M. (2022). Equal time for equal crime? Racial bias in school discipline. *Economics of Education Review, 88*.

Skiba, R.J., Michael, R.S., & Nardo, A.C. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review,* 34, 317–342. https://doi.org/10.1023/A:1021320817372

Sorenson, L. C., Bushway, S., D., and Gifford, E. J. (2022). Getting tough? The effects of discretionary principal discipline on student outcomes. *Education Finance and Policy, 17*(2), 255-284.

Thum, Y.M. & Kuhfeld, M. (2020). NWEA 2020 MAP Growth achievement status and growth norms for students and schools. NWEA.

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1990-2022 Reading Assessment.

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), various years, 1990-2022 Mathematics Assessment.

Welsh, R. (2022). School discipline in the age of COVID-19: Exploring patterns, policy, and practice considerations. *Peabody Journal of Education*, 97(3), 291-308.

Xu, Z., Goldhaber, D., & Dao, R. (2025). Student engagement in charter and traditional public schools before and after the COVID-19 pandemic. CALDER Working Paper No. 331-1225. https://doi.org/10.63302/HATI5550.

Tables and Figures

Table 1. Descriptive Statistics for Full Student Population and Analytic Sample

	Guilford Students Enrolled in 2022-23 (Grade 3-9)					Analytic Sample (Grades 3-9)†				
	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N
Participation in Interventions										
Any Tutoring (SY 21-22)	0.08		0.00	1.00	36,864	0.09		0.00	1.00	29,817
Any Summer School (2022)	0.09		0.00	1.00	36,864	0.10		0.00	1.00	29,817
Any Tutoring (SY 22-23)	0.18		0.00	1.00	36,864	0.17		0.00	1.00	29,817
Any Summer School (2023)	0.05		0.00	1.00	36,864	0.05		0.00	1.00	29,817
Student Demographics										
Male	0.52		0.00	1.00	36,864	0.52		0.00	1.00	29,817
English Language Learners	0.11		0.00	1.00	36,864	0.12		0.00	1.00	29,817
Gifted	0.18		0.00	1.00	36,864	0.20		0.00	1.00	29,817
Students with Disabilities	0.14		0.00	1.00	36,864	0.14		0.00	1.00	29,817
Race: Asian	0.07		0.00	1.00	36,864	0.07		0.00	1.00	29,817
Race: Black	0.43		0.00	1.00	36,864	0.43		0.00	1.00	29,817
Race: Hispanic	0.19		0.00	1.00	36,864	0.19		0.00	1.00	29,817
Race: White	0.26		0.00	1.00	36,864	0.26		0.00	1.00	29,817
Race: Other	0.06		0.00	1.00	36,864	0.05		0.00	1.00	29,817
Prior Test Scores										
MAP Growth RIT Score (Norm-Standardized)										
Math - Fall 2021	-0.23	1.10	-5.32	4.36	29,281	-0.21	1.09	-5.32	4.36	26,716
Reading - Fall 2021	-0.08	1.06	-3.94	3.33	23,789	-0.07	1.05	-3.94	3.33	21,688
EOG Percentile Score										
Math - Spring 2021	0.53	0.28	0.00	1.00	20,860	0.53	0.28	0.00	1.00	20,428
Reading - Spring 2021	0.45	0.3	0.00	1.00	20,890	0.45	0.28	0.00	1.00	20,462
DIBELS Percentile Score										

Fall 2021	0.40	0.29	0.000	1.00	9,125	0.40	0.29	0.00	1.00	8,421
Outcomes (SY 2022-23)										
Number of Days Absent	15.54	17.55	0.00	172.00	36,864	15.35	17.36	0.00	158.00	29,817
Chronically Absent	0.28		0.00	1.00	36,864	0.27		0.00	1.00	29,817
Number of Discipline Referrals	0.68	2.31	0.00	69.00	36,864	0.66	2.30	0.00	69.00	29,817
Received at Least One Discipline Referral	0.21		0.00	1.00	36,864	0.21		0.00	1.00	29,817
Fighting	0.06		0.00	1.00	36,864	0.06		0.00	1.00	29,817
Disruptive Behavior	0.11		0.00	1.00	36,864	0.10		0.00	1.00	29,817
Cutting School/Class	0.06		0.00	1.00	36,864	0.06		0.00	1.00	29,817
Non-Violent Aggressive Behavior	0.06		0.00	1.00	36,864	0.06		0.00	1.00	29,817
Number of Days Suspended	0.60	2.83	0.00	66.00	36,864	0.58	2.79	0.00	66.00	29,817
Received at Least One Suspension	0.09		0.00	1.00	36,864	0.09		0.00	1.00	29,817
Unweighted GPA	2.51	1.13	0.00	4.00	36,577	2.52	1.12	0.00	4.00	29,817
Retained Next Year	0.01		0.00	1.00	36,864	0.01		0.00	1.00	29,817
Non-Test Factor	0.00	1.00	-9.41	1.68	36,577	0.02	0.99	9.41	1.68	29,817
Baseline Outcome Measures (SY 2020-21)										
Number of Days Absent	14.41	21.52	0.00	174.00	30,220	14.16	21.07	0.00	169.00	29,817
Number of Discipline Referrals	0.03	0.36	0.00	30.00	30,220	0.03	0.35	0.00	30.00	29,817
Number of Days Suspended	0.01	0.16	0.00	10.00	30,220	0.01	0.15	0.00	10.00	29,817
Unweighted GPA	2.30	1.18	0.00	4.00	21,756	2.30	1.18	0.00	4.00	21,595
Retained Next Year	0.00		0.00	0.00	30,220	0.00	0.00		0.00	29,817
Non-Test Factor	0.00	1.01	-25.99	1.65	30,011	0.00	1.01	-25.99	1.65	29,817

† Analytic sample shown is for models estimating impacts of SY 2021-22 Tutoring, using 2022-23 as the outcome period. See Appendix Table A1 for summary statistics of other analytic samples used.

Notes: Table reports descriptive statistics for students in grades 3-9 enrolled in Guilford County Schools in the 2022-23 school year. Students included in the analytical sample have non-missing outcome variables (e.g., number of days absent, number of disciplinary referrals, number of days suspended, unweighted GPA, and grade retention) from baseline (2020-21) and outcome (2022-23) school years; students in grades 3-4 are allowed to have missing GPA at baseline, as they are too young to have been assigned course grades in that year. Intervention categories are not mutually exclusive; students may participate in multiple interventions.

Table 2. Characteristics of Academic Recovery Intervention Participants

	Intervention Participants									
	SY 21-22 Tutoring		Summer School 22		SY 22-23 Tutoring		Summer School 23		Full Analytic Sample †	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Grade Level										
Elementary (grades 3-5 in outcome year)	0.63		0.57		0.64		0.53		0.42	
Middle (grades 6-8 in outcome year)	0.31		0.33		0.31		0.47		0.42	
High (grade 9 in outcome year)	0.07		0.09		0.05		0.00		0.16	
Student Demographics										
Male	0.49		0.53		0.49		0.53		0.52	
English Language Learners (ELL)	0.14		0.20		0.16		0.20		0.12	
Gifted	0.06		0.01		0.06		0.01		0.20	
Special Education	0.12		0.26		0.11		0.24		0.14	
Race: Asian	0.05		0.05		0.06		0.05		0.07	
Race: Black	0.49		0.55		0.48		0.54		0.43	
Race: Hispanic	0.19		0.23		0.21		0.24		0.19	
Race: White	0.22		0.13		0.20		0.11		0.26	
Race: Other	0.05		0.05		0.05		0.06		0.05	
Prior Test Scores										
Average RIT Score (Norm-Standardized)										
Math - Fall 21	-0.62	0.92	-1.08	0.90	-0.58	0.99	-1.08	0.92	-0.21	1.09
Reading - Fall 21	-0.39	0.91	-0.94	0.89	-0.43	0.96	-0.94	0.89	-0.07	1.05
Average EOG Percentile Score										
Math - Spring 21	0.45	0.25	0.32	0.20	0.45	0.27	0.31	0.22	0.53	0.28
Reading - Spring 21	0.31	0.25	0.27	0.20	0.25	0.23	0.21	0.17	0.45	0.28
Average DIBELS Percentile Score										

Fall 2021	25.90	23.10	17.88	18.54	28.28	25.73	14.76	17.17	39.81	29.11
Outcomes (in year following intervention)										
Number of Days Absent	14.56	14.24	15.33	15.63	14.50	14.11	15.08	15.22	15.35	17.36
Chronically Absent	0.27		0.29		0.27		0.29		0.27	
Number of Discipline Referrals	0.53	1.76	0.88	2.58	0.69	2.49	1.18	3.40	0.66	2.30
Received at Least One Discipline Referral	0.20		0.27		0.21		0.30		0.21	
Number of Days Suspended	0.31	1.75	0.76	3.17	0.46	2.46	0.77	3.33	0.58	2.79
Received at Least One Suspension	0.07		0.12		0.08		0.13		0.09	
Unweighted GPA	2.29	0.98	1.87	0.94	2.28	1.01	1.86	0.94	2.52	1.12
Retained Next Year	0.00		0.01		0.00		0.00		0.01	
Non-Test Factor	-0.14	0.96	-0.35	1.05	-0.16	1.01	-0.40	1.09	0.02	0.99
Baseline Outcome Measures (in year prior)										
Number of Days Absent	16.08	22.41	16.58	15.55	16.12	14.61	14.73	13.88	14.16	21.07
Number of Discipline Referrals	0.02	0.21	0.62	2.05	0.23	1.11	0.90	3.06	0.03	0.35
Number of Days Suspended	0.00	0.12	0.48	2.24	0.14	1.08	0.52	2.34	0.01	0.15
Unweighted GPA	2.22	1.04	1.78	0.91	2.33	0.99	1.72	0.96	2.30	1.18
Retained Next Year	0.00		0.02		0.00		0.01		0.00	
Non-Test Factor	-0.07	0.94	-0.35	1.09	-0.07	0.93	-0.36	1.10	0.00	1.01
Observations	2,832		3,261		5,977		2,194		29,817	

† Full analytic sample shown is for models estimating effects of SY 2021-22 tutoring, using 2022-23 as outcome period and 2020-21 as baseline period.

Notes: Descriptive statistics are shown for intervention participants with non-missing baseline and outcome non-test factor measures. Intervention categories are not mutually exclusive; students may participate in multiple interventions.

Table 3. Association Between Non-Test Outcomes and Test Scores

	Math Test Score Percentile			Reading Test Score Percentile		
	(1)	(2)	(3)	(4)	(5)	(6)
	Chronic absence	Any suspensions	GPA (math and reading)	Chronic absence	Any suspensions	GPA (math and reading)
End of Grade (EOG) percentile	-0.234***	-0.146***	2.353***	-0.119***	-0.0817***	1.754***
	(0.00642)	(0.00475)	(0.0139)	(0.00653)	(0.00444)	(0.0156)
Year: 2018	0.0152*	0.00148		0.0109	-0.00515	
	(0.00619)	(0.00470)		(0.00575)	(0.00409)	
Year: 2019	0.0447***	0.0109*		0.0341***	-0.00459	
	(0.00631)	(0.00477)		(0.00586)	(0.00410)	
Year: 2021	0.223***	-0.121***	-0.424***	0.153***	-0.0911***	-0.0215
	(0.00694)	(0.00343)	(0.0148)	(0.00649)	(0.00305)	(0.0162)
Year: 2022	0.307***	0.0510***	-0.176***	0.267***	0.0227***	-0.131***
	(0.00703)	(0.00514)	(0.0139)	(0.00676)	(0.00434)	(0.0141)
Year: 2023	0.240***	0.0577***	-0.277***	0.189***	0.0253***	-0.171***
	(0.00700)	(0.00522)	(0.0139)	(0.00663)	(0.00441)	(0.0142)
2018 x EOG percentile	-0.0204*	-0.000920		-0.0184*	0.00904	
	(0.00922)	(0.00672)		(0.00922)	(0.00619)	
2019 x EOG percentile	-0.0555***	-0.0158*		-0.0404***	0.0109	
	(0.00937)	(0.00679)		(0.00938)	(0.00623)	
2021 x EOG percentile	-0.244***	0.138***	0.0602**	-0.115***	0.0759***	-0.790***
	(0.0102)	(0.00495)	(0.0226)	(0.0103)	(0.00458)	(0.0278)
2022 x EOG percentile	-0.252***	-0.0572***	0.117***	-0.185***	-0.00436	0.0101
	(0.0107)	(0.00729)	(0.0204)	(0.0110)	(0.00648)	(0.0234)
2023 x EOG percentile	-0.200***	-0.0586***	0.249***	-0.108***	0.00138	0.0326
	(0.0107)	(0.00745)	(0.0204)	(0.0108)	(0.00669)	(0.0235)
Constant	0.244***	0.129***	1.522***	0.189***	0.0982***	1.848***
	(0.00429)	(0.00330)	(0.00946)	(0.00405)	(0.00294)	(0.00949)
Observations	175989	175989	113378	185988	185988	121422
R-squared	0.1048	0.0504	0.4519	0.0522	0.0197	0.2015

*p<.05 **p<.01 ***p<.001

Notes: Results shown are from regressions of non-test outcomes in a school-year on math or reading test scores in spring of the same school year, year fixed effects, and test scores interacted with year. 2017 is the reference year for chronic absence and suspensions. 2019 is the reference year for GPA, as course grade data are not available for 2016-17 or 2017-18.

Table 4. Estimated Impacts of School Year Tutoring on Non-Test Factor (NTF) Index Across Multiple Model Specifications

	Tutoring 2021-22						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	NTF	NTF	NTF	NTF	NTF	NTF	NTF
Any Tutoring Sessions	-0.0649*	-0.0995***	-0.0593*	0.0211	0.0311	0.0210	0.0352
	(0.0292)	(0.0271)	(0.0266)	(0.0248)	(0.0190)	(0.0148)	(0.0191)
Asian (ref = White)	0.241***		0.240***	0.231***	0.224***	0.231***	0.229***
	(0.0288)		(0.0386)	(0.0395)	(0.0529)	(0.00199)	(0.0533)
Black (ref = White)	-0.382***		-0.231***	-0.136***	-0.146***	-0.136***	-0.147***
	(0.0468)		(0.0417)	(0.0352)	(0.0400)	(0.00162)	(0.0348)
Hispanic (ref = White)	-0.0961*		0.0365	0.0475	0.0582	0.0475***	0.0575
	(0.0420)		(0.0370)	(0.0336)	(0.0378)	(0.00192)	(0.0350)
Other race (ref = White)	-0.276***		-0.161*	-0.0980	-0.104	-0.0980***	-0.107*
	(0.0652)		(0.0637)	(0.0579)	(0.0554)	(0.00214)	(0.0539)
Male	-0.225***		-0.205***	-0.209***	-0.197***	-0.209***	-0.202***
	(0.0349)		(0.0311)	(0.0343)	(0.0341)	(0.000761)	(0.0338)
Gifted	0.513***		0.288***	0.0542	0.0738*	0.0542***	0.0677*
	(0.0553)		(0.0418)	(0.0325)	(0.0324)	(0.000279)	(0.0299)
ELL	-0.0532		0.0262	0.219***	0.222***	0.219***	0.236***
	(0.0551)		(0.0603)	(0.0637)	(0.0596)	(0.000991)	(0.0545)
Control for baseline NTF (cubic polynomial)		X	X	X	X	X	X
Control for prior test scores (MAP, EOG, Dibels)				X	X	X	X
School FE					X		
Grade FE						X	
School-by-grade FE							X
Observations	29817	29817	29817	29817	29817	29817	29817
R-squared	0.1601	0.2060	0.2644	0.3213	0.3685	0.3213	0.3816

*p<.05 **p<.01 ***p<.001

Notes: Estimated coefficients shown are only for participating in tutoring and demographic variables; the value-added models they come from additionally control for participation in other available recovery interventions and the combination of other controls reflected in the table. NTF index impacts estimates are in standard deviations of the NTF index measure. All tutoring effect estimates are for the impact of participating in at least one session of tutoring. Standard errors are clustered at the school-by-grade level with the exception of models in columns 5 and 6, which are clustered at the school level and grade level, respectively.

Table 5. Estimated Impacts of School Year Tutoring on Individual Non-Test Outcomes

		Outcomes								
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		NTF index	Chronically absent	Log absences	Any disciplinary referrals	Log disciplinary referrals	Any suspensions	Log days suspended	GPA	Observations
Intervention:										
Tutoring 2021-22 *(outcome: 2022-23)*	All (3-9)	0.0352	-0.00761	-0.0202	-0.00720	-0.0194	-0.0105	-0.0286**	-0.0106	29817
		(0.0191)	(0.00908)	(0.0177)	(0.00766)	(0.01000)	(0.00584)	(0.00940)	(0.0233)	
	Elem (3-5)	0.00978	-0.00627	-0.0322	-0.00509	-0.00339	-0.00236	-0.00713	-0.0257	12413
		(0.0255)	(0.0121)	(0.0204)	(0.00944)	(0.0105)	(0.00559)	(0.00672)	(0.0309)	
	MS (6-8)	0.0726*	-0.00696	0.0259	-0.0148	-0.0513*	-0.0226	-0.0562*	0.0119	12569
		(0.0329)	(0.0164)	(0.0345)	(0.0141)	(0.0196)	(0.0134)	(0.0221)	(0.0373)	
	HS (9)	0.0490	-0.0157	-0.117	0.0125	0.00154	-0.0171	-0.0645	-0.0683	4835
		(0.0445)	(0.0229)	(0.0678)	(0.0293)	(0.0396)	(0.0190)	(0.0426)	(0.0443)	
Tutoring 2022-23 *(outcome: 2023-24)*	All (3-9)	-0.0147	-0.00839	-0.0310*	-0.00207	-0.00248	0.00237	0.000317	-0.0407*	28942
		(0.0155)	(0.00751)	(0.0147)	(0.00648)	(0.00881)	(0.00457)	(0.00754)	(0.0165)	
	Elem (3-5)	-0.0361	0.00620	-0.00858	-0.000180	0.00553	0.00264	0.00164	-0.0472	12034
		(0.0208)	(0.00953)	(0.0201)	(0.00789)	(0.00881)	(0.00412)	(0.00508)	(0.0267)	
	MS (6-8)	0.00885	-0.0339**	-0.0466*	0.00179	-0.00789	0.00254	-0.00296	-0.0348	12253
		(0.0266)	(0.0126)	(0.0208)	(0.0121)	(0.0186)	(0.0104)	(0.0176)	(0.0245)	
	HS (9)	0.0254	0.00533	-0.111	-0.0328	-0.0320	-0.000773	0.00625	-0.0499	4655
		(0.0415)	(0.0260)	(0.0558)	(0.0212)	(0.0315)	(0.0137)	(0.0335)	(0.0294)	

*p<.05 **p<.01 ***p<.001

Notes: Estimates shown are from value-added models controlling for participation in available recovery interventions, student characteristics, prior student test scores, prior non-test factor (NTF) index, and school-by-grade fixed effects. NTF index impacts estimates are in standard deviations of the NTF index measure. Chronically absent, has any disciplinary referrals, and has any suspensions are binary indicators; reported estimates for these models should be interpreted as percentage point changes in the outcome. Log absences, log disciplinary referrals, and log days suspended are the natural logs (plus 1) of the given outcome; reported estimates for these models should be interpreted as percent changes in the outcome. All estimates are for the impact of participating in at least one session of the given intervention. The samples used when modeling effects to GPA exclude grades 3 and 4: for tutoring 2021-22, the number of observations is 21,588 for all grades and 4,184 for elementary; for tutoring 2022-23, the number of observations is 21,051 for all grades and 4,143 for elementary.

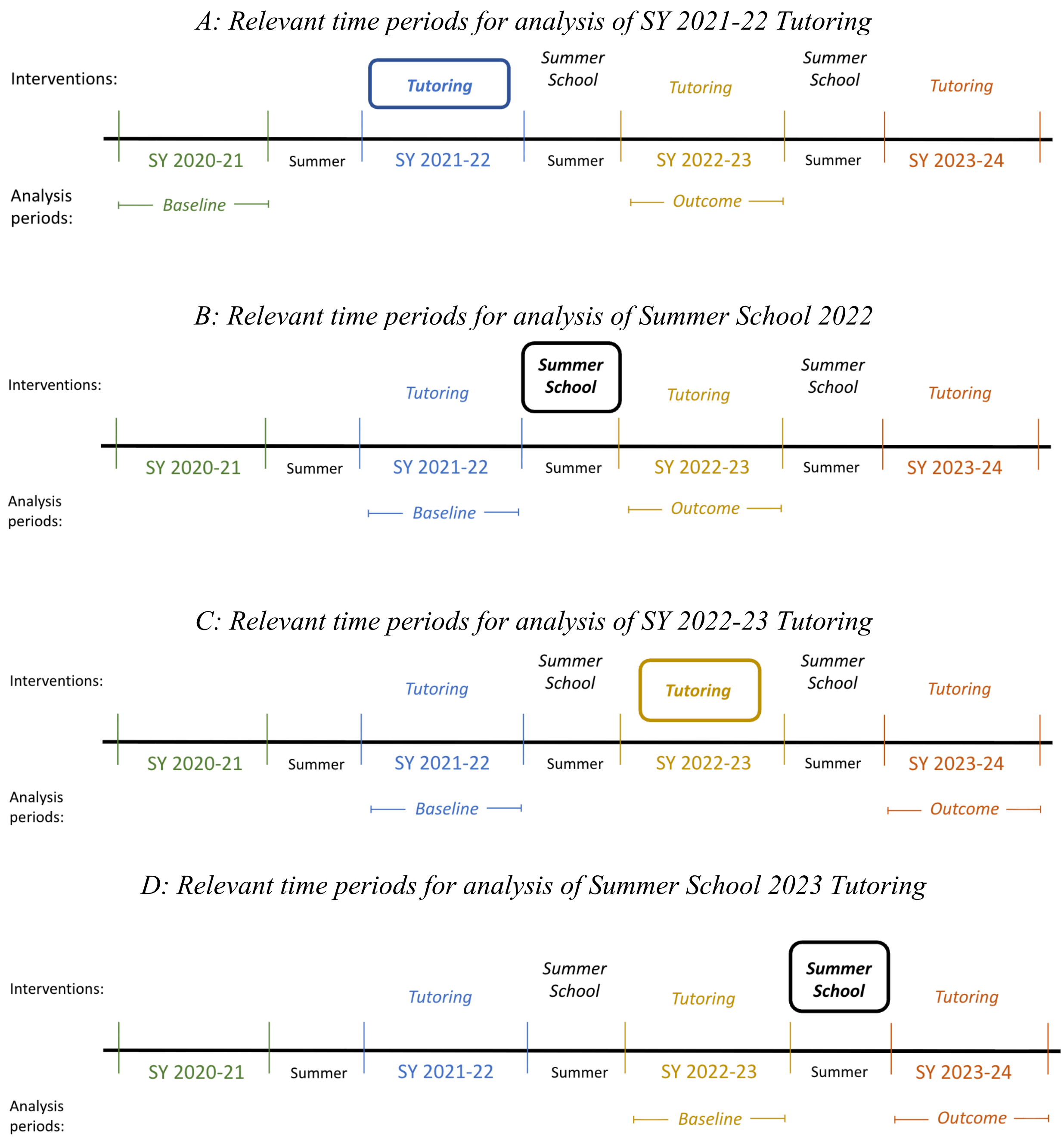
Table 6. Estimated Impacts of Summer Programming on Individual Non-Test Outcomes

		Outcomes								
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		NTF index	Chronically absent	Log absences	Any disciplinary referrals	Log disciplinary referrals	Any suspensions	Log days suspended	GPA	Observations
Intervention:										
Summer School 2022 (outcome: 2022-23)	All (3-9)	0.0457*	-0.0604***	-0.134***	-0.00148	-0.00866	-0.00603	-0.00605	-0.0572*	32469
		(0.0193)	(0.00918)	(0.0169)	(0.00898)	(0.0117)	(0.00693)	(0.0127)	(0.0227)	
	Elem (3-5)	0.0601*	-0.0686***	-0.135***	0.00481	0.00252	-0.00151	-0.00224	-0.0200	13442
		(0.0266)	(0.0149)	(0.0264)	(0.0109)	(0.0118)	(0.00694)	(0.00706)	(0.0324)	
	MS (6-8)	0.0591	-0.0613***	-0.146***	-0.0102	-0.0290	-0.0178	-0.0309	-0.0424	13802
		(0.0319)	(0.0136)	(0.0256)	(0.0156)	(0.0217)	(0.0137)	(0.0252)	(0.0338)	
	HS (9)	-0.0482	-0.0300	-0.0961*	0.00410	0.0128	0.0120	0.0538	-0.145***	5225
		(0.0563)	(0.0181)	(0.0406)	(0.0308)	(0.0464)	(0.0226)	(0.0554)	(0.0386)	
Summer School 2023 (outcome: 2023-24)	All (3-9)	0.0417	-0.0386*	-0.119***	0.0108	-0.00284	-0.00920	-0.0246	-0.0740	31265
		(0.0298)	(0.0162)	(0.0317)	(0.0121)	(0.0178)	(0.0113)	(0.0172)	(0.0393)	
	Elem (3-5)	0.0530	-0.0485*	-0.167***	0.0127	0.00899	-0.00130	-0.00690	-	13074
		(0.0468)	(0.0221)	(0.0393)	(0.0169)	(0.0196)	(0.00829)	(0.00960)	-	
	MS (6-8)	0.0345	-0.0320	-0.0849	0.0108	-0.0110	-0.0156	-0.0382	-0.0720	13218
		(0.0393)	(0.0232)	(0.0470)	(0.0172)	(0.0278)	(0.0191)	(0.0297)	(0.0394)	

*p<.05 **p<.01 ***p<.001

Notes: Estimates shown are from value-added models controlling for participation in available recovery interventions, student characteristics, prior student test scores, prior non-test factor (NTF) index, and school-by-grade fixed effects. Each impact estimate is from a separate regression. NTF index impacts estimates are in standard deviations of the NTF index measure. Chronically absent, has any disciplinary referrals, and has any suspensions are binary indicators; reported estimates for these models should be interpreted as percentage point changes in the outcome. Log absences, log disciplinary referrals, and log days suspended are the natural logs (plus 1) of the given outcome; reported estimates for these models should be interpreted as percent changes in the outcome. All estimates are for the impact of participating in at least one session of the given intervention. The samples used when modeling effects to GPA exclude grades 3 and 4: for summer school 2022, the number of observations is 23,563 for all grades and 4,536 for elementary; for summer school 2023, the number of observations is 22,659 for all grades and 4,468 for elementary.

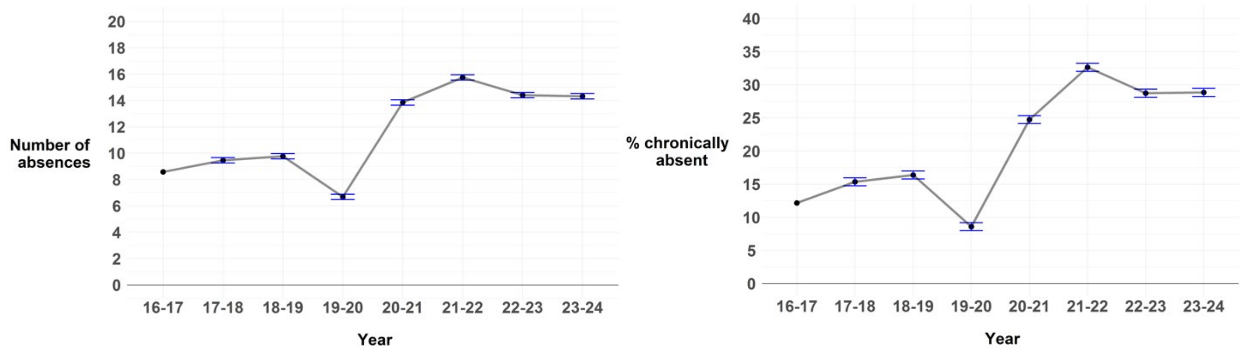
Figure 1. Time Periods for Impact Estimate Models



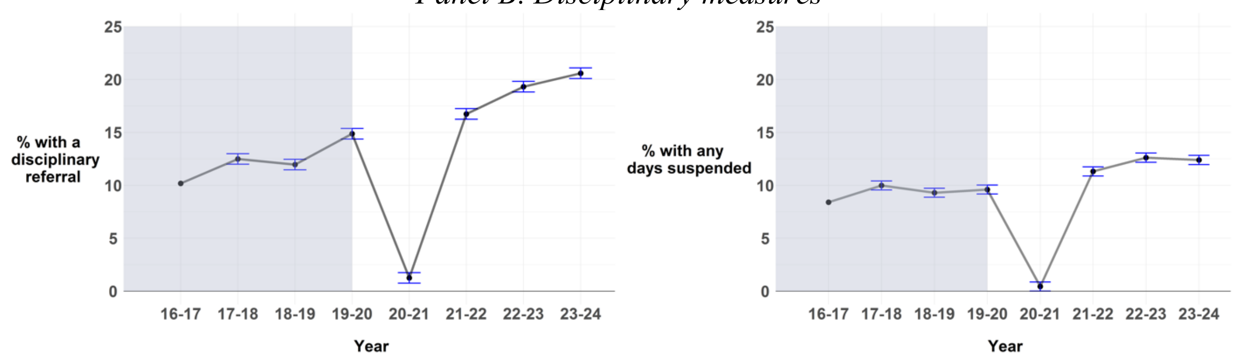
Notes: Intervention(s) being evaluated are shown in bold.

Figure 2. Overall Trends in Absenteeism and Disciplinary Measures, Grades 3-8

Panel A. Absenteeism



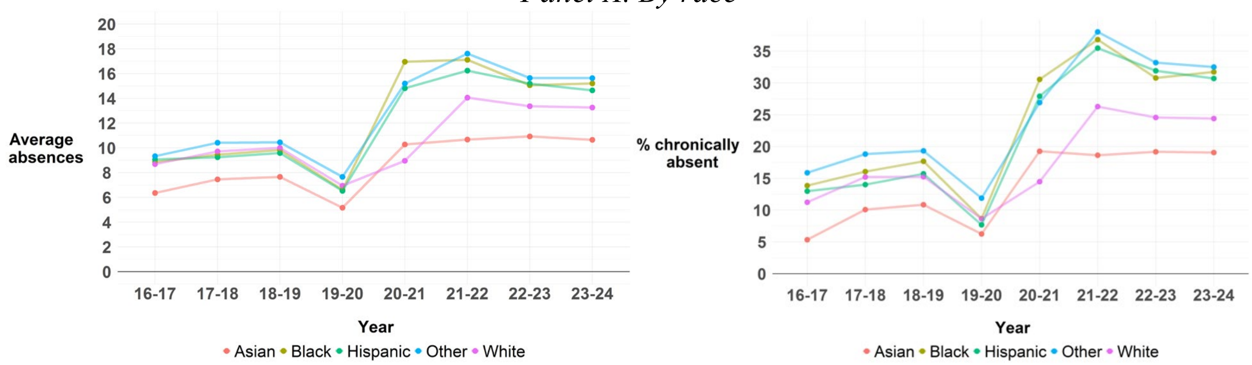
Panel B. Disciplinary measures



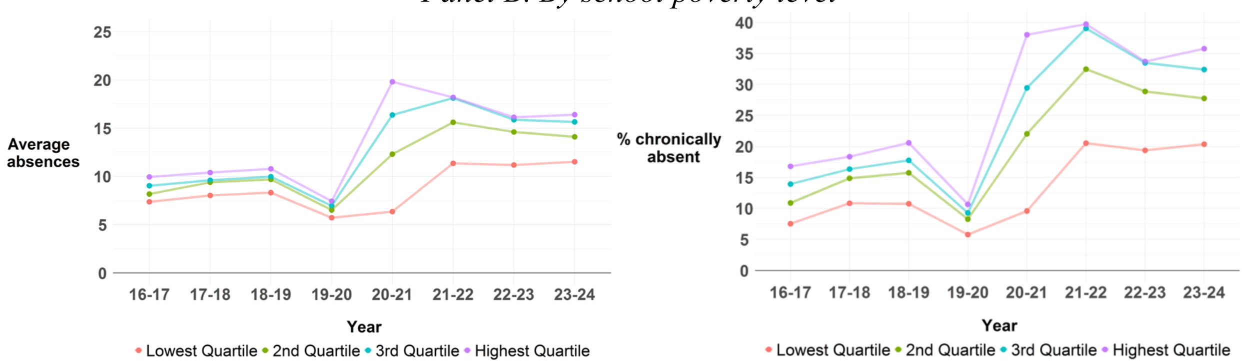
Notes: Data points shown represent predicted values for a student in the average grade level and of average student characteristics (ELL, SWD, gifted status) and race. Estimates are from regressions of outcome measures on mean-centered grade level, year, mean-centered student characteristic dummies, and mean-centered race dummies. 95% confidence intervals on estimated year effects are shown in blue. Chronic absence is defined as missing 10% or more of instructional days enrolled. The color shift in 2019-20 represents a change that occurred that year in GCS' reporting system for school discipline, which was likely to affect reporting of disciplinary referrals in particular.

Figure 3. Trends in Absenteeism by Race and School Poverty Level, Grades 3-8

Panel A. By race



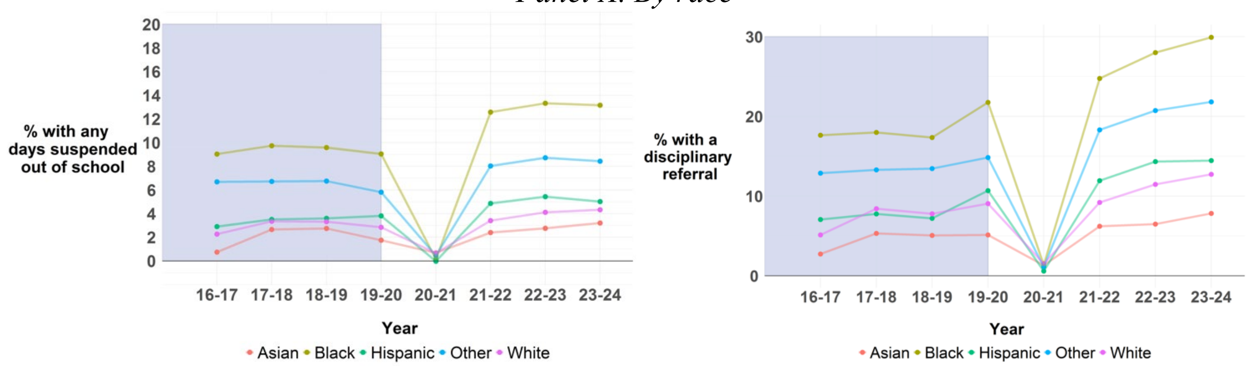
Panel B. By school poverty level



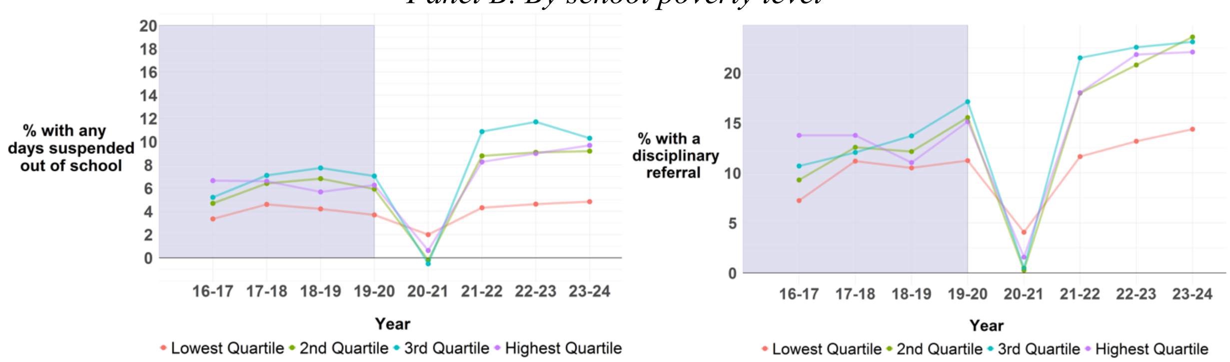
Notes: Data points shown represent predicted values for a student in the average grade level and of average student characteristics (ELL, SWD, gifted status). Estimates for Panel A are from a regression of attendance measures on grade level, student characteristics, race, year, and race-year interactions. Estimates for Panel B are from a regression of attendance measures on grade level, student characteristics, school poverty level, year, and school poverty level-year interactions. Chronic absence is defined as missing 10% or more of instructional days enrolled.

Figure 4. Trends in School Discipline by Race and School Poverty Level, Grades 3-8

Panel A. By race

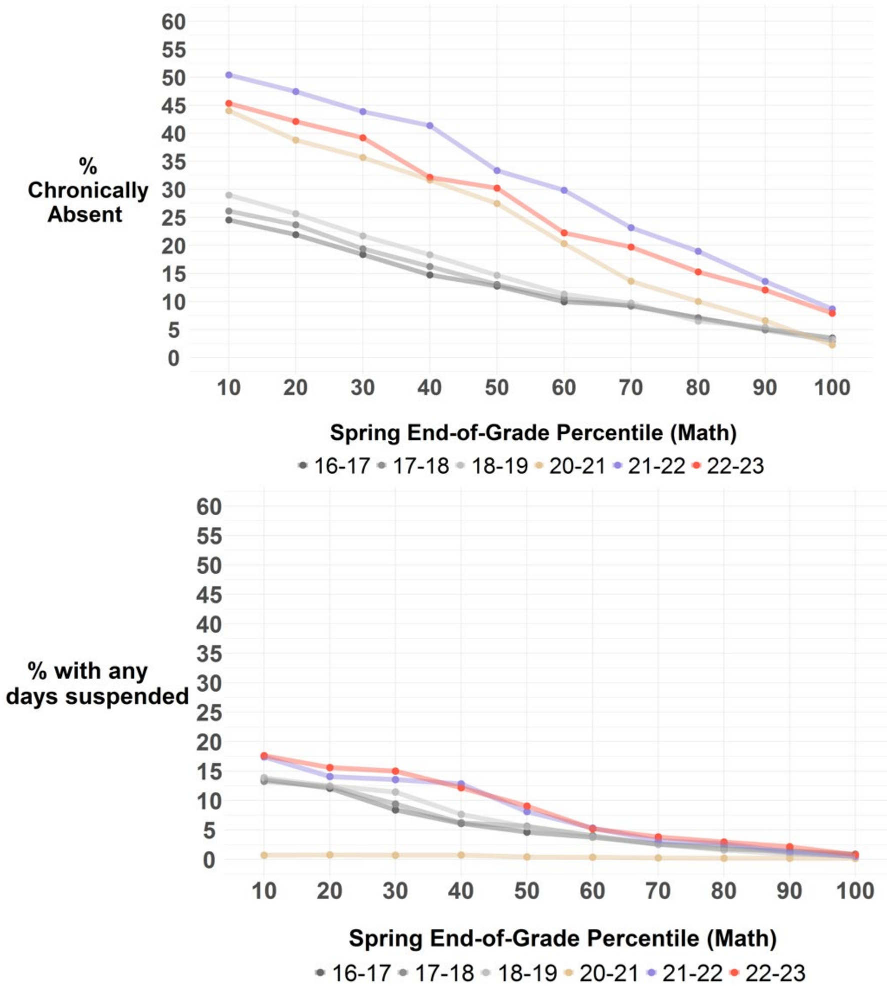


Panel B. By school poverty level



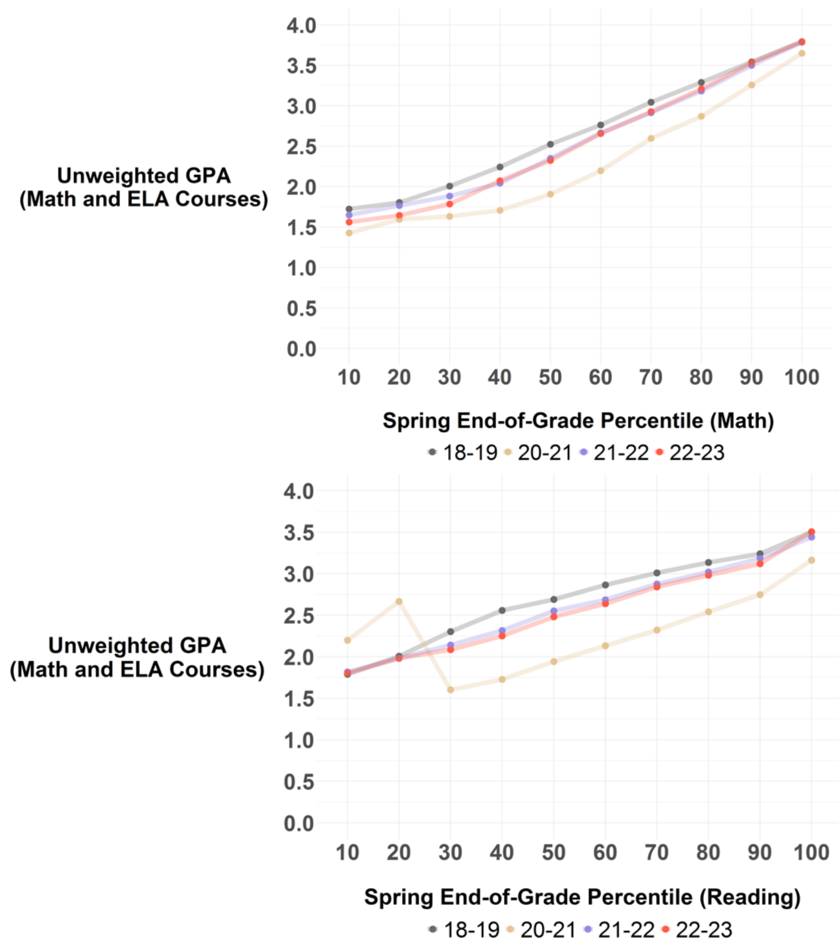
Notes: Data points shown represent predicted values for a student in the average grade level and of average student characteristics (ELL, SWD, gifted status). Estimates for Panel A are from a regression of discipline measures on grade level, student characteristics, race, year, and race-year interactions. Estimates for Panel B are from a regression of discipline measures on grade level, student characteristics, school poverty level as measured by the percentage of free and reduced-price lunch eligible students, year, and school poverty level-year interactions. The color shift in 2019-20 represents a change that occurred that year in GCS' reporting system for school discipline, which was likely to affect reporting of disciplinary referrals in particular.

Figure 5. Association Between Test Score Performance and Measures of Attendance and Discipline by Year, Grades 3-8



Notes: Figures show the average percent of students who are either chronically absent or have at least one suspension at each decile of the spring End-of-Grade math scores in the same school year. Comparable figures for End-of-Grade ELA scores look similar and are not shown to conserve space.

Figure 6. Association Between Course Grades and Test Score Performance by Year, Grades 3-8



Notes: Figures show the average grade point average for math and ELA courses at each decile of the spring End-of-Grade math and reading scores in the same school year.

Appendix of Supplemental Tables

Appendix Table A1. Descriptive Statistics for Full Student Population and Additional Analytic Samples

	Guilford Students Enrolled in 2022-23 (Grade 3-9)					Analytic Sample for Models of Summer 2022 Program Impacts (Grades 3-9)					Analytic Sample for Models of SY 2022-23 Tutoring Impacts (Grades 3-9)					Analytic Sample for Models of Summer 2023 Program Impacts (Grades 3-8)				
	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N
Participation in Interventions																				
Any Tutoring (SY 21-22)	0.09		0.00	1.00	36,864	0.10		0.00	1.00	32,469	0.11		0.00	1.00	28,944	0.10		0.00	1.00	31,266
Any Summer School (2022)	0.09		0.00	1.00	36,864	0.10		0.00	1.00	32,469	0.11		0.00	1.00	28,944	0.11		0.00	1.00	31,266
Any Tutoring (SY 22-23)	0.18		0.00	1.00	36,864	0.17		0.00	1.00	32,469	0.21		0.00	1.00	28,944	0.21		0.00	1.00	31,266
Any Summer School (2023)	0.05		0.00	1.00	36,864	0.05		0.00	1.00	32,469	0.07		0.00	1.00	28,944	0.07		0.00	1.00	31,266
Student Demographics																				
Male	0.52		0.00	1.00	36,864	0.52		0.00	1.00	32,469	0.51		0.00	1.00	28,944	0.51		0.00	1.00	31,266
English Language Learners	0.11		0.00	1.00	36,864	0.11		0.00	1.00	32,469	0.12		0.00	1.00	28,944	0.12		0.00	1.00	31,266
Gifted	0.18		0.00	1.00	36,864	0.20		0.00	1.00	32,469	0.20		0.00	1.00	28,944	0.19		0.00	1.00	31,266
Students with Disabilities	0.14		0.00	1.00	36,864	0.14		0.00	1.00	32,469	0.13		0.00	1.00	28,944	0.14		0.00	1.00	31,266
Race: Asian	0.07		0.00	1.00	36,864	0.07		0.00	1.00	32,469	0.07		0.00	1.00	28,944	0.07		0.00	1.00	31,266
Race: Black	0.43		0.00	1.00	36,864	0.43		0.00	1.00	32,469	0.42		0.00	1.00	28,944	0.42		0.00	1.00	31,266
Race: Hispanic	0.19		0.00	1.00	36,864	0.19		0.00	1.00	32,469	0.19		0.00	1.00	28,944	0.19		0.00	1.00	31,266
Race: White	0.26		0.00	1.00	36,864	0.26		0.00	1.00	32,469	0.26		0.00	1.00	28,944	0.26		0.00	1.00	31,266
Race: Other	0.06		0.00	1.00	36,864	0.05		0.00	1.00	32,469	0.05		0.00	1.00	28,944	0.05		0.00	1.00	31,266
Prior Test Scores																				
MAP Growth RIT Score (Norm-Standardized)																				
Math - Fall 2021	-0.23	1.10	-5.32	4.36	29,281	-0.22	1.10	-5.32	4.358	29,087	-0.21	1.11	-5.32	4.421	26,223	-0.21	1.11	-5.32	4.421	25,932
Reading - Fall 2021	-0.08	1.06	-3.94	3.33	23,789	-0.08	1.05	-3.94	3.334	23,653	-0.05	1.07	-3.94	3.639	18,847	-0.05	1.07	-3.94	3.875	18,622
EOG Percentile Score																				
Math - Spring 2021	0.53	0.28	0.00	1.00	20,860	0.53	0.28	0.00	1.00	20,488	0.53	0.29	0.01	1.00	14,976	0.53	0.29	0.00	1.00	14,949
Reading - Spring 2021	0.45	0.28	0.00	1.00	20,890	0.45	0.28	0.00	1.00	20,515	0.41	0.28	0.00	1.00	14,989	0.41	0.28	0.00	1.00	14,959

DIBELS Percentile Score																				
Fall 2021	39.67	29.16	0.00	100.00	9,125	39.75	29.17	0.00	100.00	9,034	39.25	30.20	0.00	100.00	11,894	39.36	30.19	0.00	100.00	11,794
Outcomes (in year following intervention)																				
Number of Days Absent	15.54	17.55	0.00	172.00	36,864	15.48	17.42	0.00	158.00	32,469	14.66	15.51	0.00	139.00	28,944	14.73	15.54	0.00	139.00	31,266
Chronically Absent	0.28		0.00	1.00	36,864	0.28		0.00	1.00	32,469	0.26		0.00	1.00	28,944	0.27		0.00	1.00	31,266
Number of Discipline Referrals	0.68	2.31	0.00	69.00	36,864	0.67	2.30	0.00	69.00	32,469	0.76	2.59	0.00	66.00	28,944	0.76	2.59	0.00	66.00	31,266
Received at Least One Discipline Referral	0.21		0.00	1.00	36,864	0.21		0.00	1.00	32,469	0.22		0.00	1.00	28,944	0.22		0.00	1.00	31,266
Number of Days Suspended	0.60	2.83	0.00	66.00	36,864	0.59	2.80	0.00	66.00	32,469	0.57	2.87	0.00	75.00	28,944	0.57	2.89	0.00	75.00	31,266
Received at Least One Suspension	0.09		0.00	1.00	36,864	0.09		0.00	1.00	32,469	0.09		0.00	1.00	28,944	0.09		0.00	1.00	31,266
Unweighted GPA	2.51	1.13	0.00	4.00	36,577	2.52	1.13	0.00	4.00	32,469	2.56	1.10	0.00	4.00	28,944	2.55	1.10	0.00	4.00	31,266
Retained Next Year	0.01		0.00	1.00	36,864	0.01		0.00	1.00	32,469	0.00		0.00	0.00	28,944	0.00		0.00	0.00	31,266
Non-Test Factor	-0.00	1.00	-9.41	1.68	36,577	0.01	1.00	-9.41	1.68	32,469	0.01	1.00	-9.05	1.63	28,944	0.01	1.00	-9.05	1.63	31,266
Baseline Outcome Measures (in year prior)																				
Number of Days Absent	14.41	21.52	0.00	174.0	30,220	16.00	16.46	0.00	156.00	32,469	15.14	14.73	0.00	139.00	28,944	14.25	14.53	0.00	157.00	31,266
Number of Discipline Referrals	0.03	0.36	0.00	30.00	30,220	0.45	1.72	0.00	44.00	32,469	0.35	1.47	0.00	44.00	28,944	0.55	2.06	0.00	46.00	31,266
Number of Days Suspended	0.01	0.16	0.00	10.00	30,220	0.39	2.14	0.00	69.00	32,469	0.25	1.58	0.00	40.10	28,944	0.40	2.16	0.00	61.00	31,266
Unweighted GPA	2.30	1.18	0.00	4.00	21,756	2.56	1.09	0.00	4.00	28,035	2.63	1.06	0.00	4.00	21,062	2.56	1.09	0.00	4.00	27,110
Retained Next Year	0.00		0.00	0.00	30,220	0.01		0.00	1.00	32,469	0.00		0.00	1.00	28,944	0.01		0.00	1.00	31,266
Non-Test Factor	-0.00	1.01	-25.99	1.65	30,011	0.00	1.01	-16.01	1.72	32,469	0.02	0.98	-16.52	1.72	28,944	0.02	0.99	-14.03	1.68	31,266

Appendix Table A2. Estimated Correlation Between Intervention Participation and Outcome Period Intervention Participation

	All students (3-8)		Elementary students		Middle school students		High school students	
	(1) SY 22-23 Tutoring	(2) SY 23-24 Tutoring	(3) SY 22-23 Tutoring	(4) SY 23-24 Tutoring	(5) SY 22-23 Tutoring	(6) SY 23-24 Tutoring	(7) SY 22-23 Tutoring	(8) SY 23-24 Tutoring
SY 21-22 Tutoring	0.0692***		0.107***		0.0233		-0.0120	
	(0.0109)		(0.0152)		(0.0152)		(0.0201)	
Summer School 22	0.0229**		0.0410**		0.0112		-0.00777	
	(0.00861)		(0.0150)		(0.0103)		(0.0169)	
SY 22-23 Tutoring		n/a		n/a		n/a		n/a
Summer School 23		n/a		n/a		n/a		n/a

* p<.05 ** p<.01 *** p<.001

Notes: Coefficients shown are from models predicting participation in at least one session of tutoring during the analysis outcome period, controlling for participation in interventions in the prior period as well as student characteristics, prior student test scores, prior non-test factor (NTF) index, and school-by-grade fixed effects. Each estimate is from a separate regression. Data on participation in SY 2023-24 tutoring are unavailable at this time.

Appendix Table A3. Estimated Impacts of School Year Tutoring on Individual Non-Test Outcomes, Using Individual Non-Test Outcome Measures as Baseline Controls

		Outcomes							
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Chronically absent	Log absences	Any disciplinary referrals	Log disciplinary referrals	Any suspensions	Log days suspended	GPA	Observations
Intervention:									
Tutoring 2021-22 *(outcome: 2022-23)*	All (3-9)	-0.0139	-0.0189	-0.0108	-0.0248*	-0.0127*	-0.0329**	-0.00297	29817
		(0.00982)	(0.0175)	(0.00775)	(0.0106)	(0.00608)	(0.00997)	(0.0246)	
	Elem (3-5)	-0.00399	-0.0270	-0.00549	-0.00349	-0.00181	-0.00603	-0.0175	12413
		(0.0127)	(0.0192)	(0.00945)	(0.0106)	(0.00569)	(0.00681)	(0.0313)	
	MS (6-8)	-0.0234	0.0284	-0.0219	-0.0623**	-0.0282*	-0.0665**	0.0226	12569
		(0.0177)	(0.0346)	(0.0145)	(0.0211)	(0.0140)	(0.0236)	(0.0397)	
	HS (9)	-0.0409	-0.149*	0.000801	-0.0173	-0.025	-0.0819	-0.0724	4835
		(0.0274)	(0.0666)	(0.0289)	(0.0413)	(0.0194)	(0.0419)	(0.0430)	
Tutoring 2022-23 *(outcome: 2023-24)*	All (3-9)	-0.00610	-0.0275*	-0.000671	0.000285	0.00183	-0.000850	-0.0196	28942
		(0.00752)	(0.0127)	(0.00641)	(0.00860)	(0.00460)	(0.00767)	(0.0176)	
	Elem (3-5)	0.00871	-0.0247	-0.00000182	0.00601	0.00303	0.00290	-0.0383	12034
		(0.00960)	(0.0162)	(0.00784)	(0.00854)	(0.00427)	(0.00515)	(0.0270)	
	MS (6-8)	-0.0328**	-0.0124	0.00369	-0.00403	0.00047	-0.00803	-0.00771	12253
		(0.0122)	(0.0200)	(0.0120)	(0.0182)	(0.0103)	(0.0178)	(0.0260)	
	HS (9)	0.0102	-0.111*	-0.0255	-0.0188	-0.000944	0.00522	-0.0263	4655
		(0.0266)	(0.0486)	(0.0212)	(0.0321)	(0.0142)	(0.0340)	(0.0352)	

*p<.05 **p<.01 ***p<.001

Notes: Estimates shown are from value-added models controlling for participation in available recovery interventions, student characteristics, prior student test scores, prior continuous measure of the non-test outcome, and school-by-grade fixed effects. For example, models with chronic absence and log absences as the dependent variable include baseline attendance rate as an independent variable. Each impact estimate is from a separate regression.

Appendix Table A4. Estimated Impacts of Summer Programming on Individual Non-Test Outcomes, Using Individual Non-Test Outcome Measures as Baseline Controls

		Outcomes							
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Chronically absent	Log absences	Any disciplinary referrals	Log disciplinary referrals	Any suspensions	Log days suspended	GPA	Observations
Intervention:									
Summer School 2022 *(outcome: 2022-23)*	All (3-9)	-0.0142	-0.0325*	-0.000751	-0.00805	-0.00618	-0.00461	0.00578	32469
		(0.00797)	(0.0145)	(0.00915)	(0.0123)	(0.00708)	(0.0130)	(0.0209)	
	Elem (3-5)	-0.0128	-0.0185	-0.00174	-0.00816	-0.00551	-0.00729	0.0200	13442
		(0.0128)	(0.0215)	(0.0112)	(0.0124)	(0.00720)	(0.00763)	(0.0315)	
	MS (6-8)	-0.0226	-0.0605**	-0.00638	-0.0256	-0.0177	-0.0308	0.0314	13802
		(0.0115)	(0.0226)	(0.0157)	(0.0223)	(0.0137)	(0.0254)	(0.0309)	
	HS (9)	0.00471	0.000656	0.0235	0.0488	0.0258	0.0844	-0.0858*	5225
		(0.0196)	(0.0385)	(0.0319)	(0.0494)	(0.0231)	(0.0566)	(0.0341)	
Summer School 2023 *(outcome: 2023-24)*	All (3-9)	-0.0386*	-0.0487	0.0134	-0.0000734	-0.00311	-0.0123	0.0385	31265
		(0.0181)	(0.0247)	(0.0107)	(0.0158)	(0.0114)	(0.0174)	(0.0371)	
	Elem (3-5)	-0.0643**	-0.0647	0.00800	0.00171	-0.00300	-0.00910	--	13074
		(0.0204)	(0.0356)	(0.0190)	(0.0224)	(0.00876)	(0.0108)	--	
	MS (6-8)	-0.0202	-0.0368	0.0190	-0.000251	-0.00390	-0.0154	0.0403	13218
		(0.0279)	(0.0338)	(0.0119)	(0.0220)	(0.0192)	(0.0297)	(0.0375)	

*p<.05 **p<.01 ***p<.001

Notes: Estimates shown are from value-added models controlling for participation in available recovery interventions, student characteristics, prior student test scores, prior continuous measure of the non-test outcome, and school-by-grade fixed effects. For example, models with chronic absence and log absences as the dependent variable include baseline attendance rate as an independent variable. Each impact estimate is from a separate regression.

Appendix Table A5. Estimated Impacts of SY 2021-22 Tutoring on Non-Test Outcomes in 2022-23, by Student Subgroup

	Non-Test Factor Index				Log of Days Absent				Log of Disciplinary Referrals				Log of Days Suspended			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)
Overall Effect:	0.0352	0.00978	0.0726*	0.0490	-0.0202	-0.0322	0.0259	-0.117	-0.0194	-0.00339	-0.0513*	0.00154	-0.0286**	-0.00713	-0.0562*	-0.0645
	(0.0191)	(0.0255)	(0.0329)	(0.0445)	(0.0177)	(0.0204)	(0.0345)	(0.0678)	(0.01000)	(0.0105)	(0.0196)	(0.0396)	(0.00940)	(0.00672)	(0.0221)	(0.0426)
Subgroup:																
Male	0.0252	-0.0462	0.131*	0.0845	-0.0213	-0.00435	0.00203	-0.210*	-0.0197	0.0223	-0.0932**	-0.0137	-0.0317*	0.0028	-0.0848*	-0.0585
	-0.0319	-0.0432	-0.054	-0.0537	-0.0256	-0.0317	-0.0442	-0.102	-0.0163	-0.0172	-0.035	-0.0399	-0.0142	-0.0111	-0.0339	-0.0579
Race: Asian	0.0339	-0.0174	0.0933	0.680***	0.0627	0.0658	0.125	-0.786	-0.0335	-0.00541	-0.0642	-0.556***	-0.021	0.00775	-0.0583	-0.324
	-0.0479	-0.0584	-0.0911	-0.166	-0.0957	-0.119	-0.176	-0.701	-0.0225	-0.0164	-0.059	-0.133	-0.0155	-0.0063	-0.0421	-0.175
Race: Black	0.046	0.0197	0.098	-0.00357	-0.0425	-0.0651	0.00536	-0.09	-0.0233	-0.005	-0.0717	0.0475	-0.0285	-0.00568	-0.0663	-0.0242
	-0.0345	-0.0479	-0.0568	-0.09	-0.0282	-0.0332	-0.0558	-0.0963	-0.0201	-0.0202	-0.0408	-0.075	-0.0171	-0.0131	-0.0366	-0.0835
Race: Hispanic	-0.0115	-0.0276	0.0246	-0.0626	0.0189	-0.00169	0.0429	0.0379	-0.00412	0.00762	-0.0297	0.0222	-0.000462	-0.00111	-0.00525	0.031
	-0.0319	-0.037	-0.059	-0.145	-0.0416	-0.05	-0.0746	-0.202	-0.0186	-0.0148	-0.0362	-0.126	-0.0194	-0.00879	-0.0494	-0.0826
Race: White	0.00943	-0.0193	0.0316	0.0951*	-0.00959	-0.027	0.0699	-0.14	-0.0151	0.00306	-0.0292	-0.0462	-0.0231*	-0.00569	-0.0425	-0.0686*
	-0.0313	-0.0432	-0.0545	-0.0454	-0.0345	-0.0455	-0.0623	-0.112	-0.016	-0.0154	-0.0403	-0.035	-0.00976	-0.00929	-0.0233	-0.0334
Students w/Disabilities	-0.0103	0.0273	-0.0178	-0.375	0.0078	0.00685	0.0156	-0.0114	0.00289	-0.0231	-0.00793	0.337	-0.00468	-0.0225	-0.000131	0.156
	-0.0597	-0.0739	-0.106	-0.262	-0.0572	-0.0751	-0.0961	-0.268	-0.0319	-0.028	-0.0728	-0.192	-0.0302	-0.0185	-0.0756	-0.227
Gifted	-0.0258	-0.0565	-0.031	0.0155	0.0885	0.118	0.134	-0.0496	-0.00958	-0.00353	-0.0142	0.00162	-0.0339**	-0.011	-0.0117	-0.115**
	-0.0379	-0.081	-0.0512	-0.0798	(0.0584)	(0.114)	(0.0845)	(0.109)	-0.026	-0.0334	-0.0398	-0.0565	-0.0125	-0.00912	-0.0153	-0.0413
ELL Student	-0.0126	0.00734	-0.0191	-0.202	0.0629	0.0277	0.105	0.217	-0.00367	-0.00982	-0.0178	0.155	0.0202	0.00919	0.0226	0.142
	-0.0344	-0.0444	-0.0555	-0.237	-0.045	-0.0539	-0.0893	-0.207	(0.0189)	(0.0133)	(0.0442)	(0.175)	-0.0172	-0.0081	-0.0396	-0.202

*p<.05 **p<.01 ***p<.001

Notes: Estimates shown are from value-added models controlling for participation in SY 2021-22 tutoring (and other available recovery interventions during that time), student characteristics, prior student test scores, prior non-test factor (NTF) index, and school-by-grade fixed effects. The baseline and outcome terms are SY 2020-21 and SY 2022-23, respectively. Each impact estimate is from a separate regression. Subgroup estimates are from regression models restricted to students in that subgroup; overall effects are from models including all students. All estimates are for the impact of participating in at least one session of tutoring in math or ELA during the 2021-22 school year.

Appendix Table A6. Estimated Impacts of Summer School 2022 on Non-Test Outcomes in 2022-23, by Student Subgroup

	Non-Test Factor Index				Log of Days Absent				Log of Disciplinary Referrals				Log of Days Suspended			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)
Overall Effect:	0.0457*	0.0601*	0.0591	-0.0482	-0.134***	-0.135***	-0.146***	-0.0961*	-0.00866	0.00252	-0.0290	0.0128	-0.00605	-0.00224	-0.0309	0.0538
	(0.0193)	(0.0266)	(0.0319)	(0.0563)	(0.0169)	(0.0264)	(0.0256)	(0.0406)	(0.0117)	(0.0118)	(0.0217)	(0.0464)	(0.0127)	(0.00706)	(0.0252)	(0.0554)
Subgroup:																
Male	0.0315	0.0657	0.0443	-0.114	-0.135***	-0.134***	-0.149***	-0.106	-0.00211	-0.000255	-0.0282	0.0686	0.00758	-0.00516	-0.0104	0.0979
	-0.0327	-0.046	-0.0501	-0.105	-0.0243	-0.039	-0.0344	-0.0673	-0.0194	-0.0198	-0.0315	-0.0823	-0.0213	-0.0129	-0.0419	-0.081
Race: Asian	-0.0599	0.027	-0.148	-0.184	0.127	0.00195	0.176	0.563	-0.0224	-0.0139	0.032	-0.22	0.0645	-0.00998	0.0995	0.313
	-0.0789	-0.0666	-0.119	-0.438	-0.125	-0.12	-0.198	-0.59	-0.0312	-0.0187	-0.0542	-0.216	-0.0568	-0.0126	-0.114	-0.276
Race: Black	0.0681*	0.0945*	0.0699	-0.0249	-0.142***	-0.164***	-0.137***	-0.0896	-0.0172	-0.00226	-0.0254	-0.0414	-0.0133	-0.00654	-0.0504	0.077
	-0.0304	-0.044	-0.0482	-0.0842	-0.0234	-0.0354	-0.0362	-0.0646	-0.0187	-0.0182	-0.0353	-0.0602	-0.02	-0.0123	-0.0355	-0.0885
Race: Hispanic	0.0561	0.0821*	0.0599	-0.0628	-0.156***	-0.0981	-0.236***	-0.094	-0.0182	-0.0124	-0.0442	0.0546	0.00139	-0.0112	0.0178	-0.00482
	-0.0355	-0.0382	-0.0625	-0.135	-0.0367	-0.0538	-0.0551	-0.0969	-0.0222	-0.0206	-0.0407	-0.0915	-0.0272	-0.00859	-0.0542	-0.116
Race: White	0.000148	-0.027	0.0314	-0.00244	-0.0912	-0.0626	-0.0918	-0.182	0.015	0.00189	0.0005	0.116	-0.0199	0.0185	-0.0462	-0.0952
	-0.046	-0.0672	-0.0699	-0.148	-0.0496	-0.0715	-0.0844	-0.121	-0.0326	-0.0275	-0.0507	-0.177	(0.0242)	(0.0212)	(0.0491)	(0.103)
Students w/Disabilities	-0.0185	-0.029	0.0151	-0.0768	-0.102**	-0.0923	-0.0823	-0.193	0.0149	0.0381	0.00235	-0.0326	0.0272	0.0158	0.0159	0.1
	-0.0482	-0.0748	-0.0728	-0.125	-0.0385	-0.0607	-0.0535	-0.114	-0.0259	-0.0329	-0.046	-0.075	-0.0308	-0.0233	-0.0599	-0.122
Gifted	0.154	0.336	0.0402	0.178	-0.0696	-0.33	-0.191	0.245	-0.0698	-0.125*	-0.0391	-0.0613	-0.129	-0.0626*	0.0122	-0.330*
	-0.101	-0.173	-0.145	-0.191	-0.151	-0.304	-0.175	-0.268	-0.0646	-0.0511	-0.063	-0.163	-0.0695	-0.0297	-0.1	-0.149
ELL Student	0.019	0.0289	0.0153	-0.0172	-0.146**	-0.0906	-0.218**	-0.165	-0.00357	0.0107	-0.0135	-0.0314	0.0248	0.0111	0.05	-0.00128
	-0.0376	-0.0392	-0.0758	-0.119	(0.0442)	(0.0584)	(0.0689)	(0.166)	-0.0218	-0.0204	-0.0436	-0.0846	-0.0323	-0.00995	-0.0736	-0.125

*p<.05 **p<.01 ***p<.001

Notes: Estimates shown are from value-added models controlling for participation in Summer School in 2022 (and other available recovery interventions during that time), student characteristics, prior student test scores, prior non-test factor (NTF) index, and school-by-grade fixed effects. The baseline and outcome terms are SY 2021-22 and SY 2022-23, respectively. Each impact estimate is from a separate regression. Subgroup estimates are from regression models restricted to students in that subgroup; overall effects are from models including all students. All estimates are for the impact of participating in at least one session of Summer School in 2022.

Appendix Table A7. Estimated Impacts of SY 2022-23 Tutoring on Non-Test Outcomes in 2023-24, by Student Subgroup

	Non-Test Factor Index				Log of Days Absent				Log of Disciplinary Referrals				Log of Days Suspended			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)	All (3-9)	Elem (3-5)	MS (6-8)	HS (9)
Overall Effect:	-0.0147	-0.0361	0.00885	0.0254	-0.0310*	-0.00858	-0.0466*	-0.111	-0.00248	0.00553	-0.00789	-0.0320	0.000317	0.00164	-0.00296	0.00625
	(0.0155)	(0.0208)	(0.0266)	(0.0415)	(0.0147)	(0.0201)	(0.0208)	(0.0558)	(0.00881)	(0.00881)	(0.0186)	(0.0315)	(0.00754)	(0.00508)	(0.0176)	(0.0335)
Subgroup:																
Male	-0.0358	-0.0578*	-0.00268	-0.0449	-0.0399**	-0.0000159	-0.0694***	-0.151*	0.00904	0.0108	0.00361	0.0238	0.0129	0.0082	0.00579	0.0717
	-0.0227	-0.0326	-0.0363	-0.0584	-0.0191	-0.0255	-0.0248	-0.0872	-0.0134	-0.0138	-0.0282	-0.045	-0.0118	-0.00884	-0.0243	-0.0604
Race: Asian	-0.0169	-0.0154	-0.00316	-0.0293	-0.0367	0.0108	-0.0114	-0.750*	0.0125	-0.00454	0.0146	0.173*	-0.00202	-0.00827	-0.00769	0.0817
	-0.0334	-0.05	-0.0441	-0.253	-0.0626	-0.0834	-0.0856	-0.415	-0.0174	-0.0195	-0.0307	-0.085	-0.0138	-0.00873	-0.0248	-0.188
Race: Black	-0.0238	-0.0471	-0.00781	0.0365	-0.0174	0.0118	-0.0395	-0.0869	0.00271	0.0174	-0.00827	-0.0316	0.00541	-0.00218	0.0227	-0.0181
	-0.0253	-0.0342	-0.0419	-0.078	-0.02	-0.0293	-0.0299	-0.053	-0.0144	-0.015	-0.0291	-0.0486	-0.0135	-0.00966	-0.0293	-0.0632
Race: Hispanic	0.0274	0.00261	0.0721	-0.000651	-0.0691*	-0.00352	-0.157***	-0.0777	-0.0199	-0.0230*	-0.0202	-0.014	-0.0238	-0.00142	-0.0591*	-0.0135
	-0.0268	-0.031	-0.0475	-0.129	-0.0362	-0.0409	-0.0527	-0.204	-0.0174	-0.0121	-0.0399	-0.063	-0.0148	-0.00644	-0.0314	-0.0873
Race: White	-0.0289	-0.0353	-0.0255	0.0176	-0.0332	-0.0655	0.0533	-0.19	0.000183	0.0123	-0.0177	-0.0233	0.0165*	0.00804	0.0248	0.032
	-0.0248	-0.0392	-0.0331	-0.0437	-0.0347	-0.0494	-0.0418	-0.117	-0.0138	-0.0161	-0.0275	-0.0308	-0.00845	-0.00712	-0.0214	-0.0322
Students w/Disabilities	-0.024	-0.115**	0.108	0.195	-0.00173	0.0345	-0.0959	0.0574	-0.021	0.0364	-0.0992	-0.176	-0.0141	0.00644	-0.0289	-0.138
	-0.047	-0.0568	-0.0843	-0.183	-0.038	-0.0482	-0.0614	-0.196	-0.0279	-0.0259	-0.0635	-0.115	-0.0223	-0.0147	-0.0556	-0.12
Gifted	-0.0172	-0.0389	-0.00715	0.0368	-0.0416	-0.0116	-0.0392	-0.159	0.0181	0.0179	0.0262	-0.00328	0.0055	0.00622	0.00296	0.00254
	-0.0253	-0.0425	-0.0329	-0.0732	-0.043	-0.0692	-0.0488	-0.112	-0.0166	-0.0191	-0.0337	-0.0448	-0.0113	-0.00986	-0.021	-0.0555
ELL Student	0.0105	-0.0426	0.0749	0.0558	-0.0773*	-0.0169	-0.122*	-0.211	-0.0211	-0.00824	-0.0344	-0.0411	-0.00757	0.0128	-0.0416	0.00455
	-0.04	-0.0396	-0.0797	-0.153	-0.046	-0.056	-0.0699	-0.182	-0.024	-0.0147	-0.0583	-0.0783	-0.0236	-0.00894	-0.0536	-0.118

*p<.05 **p<.01 ***p<.001

Notes: Estimates shown are from value-added models controlling for participation in SY 2022-23 Tutoring (and other available recovery interventions during that time), student characteristics, prior student test scores, prior non-test factor (NTF) index, and school-by-grade fixed effects. The baseline and outcome terms are SY 2021-22 and SY 2023-24, respectively. Each impact estimate is from a separate regression. Subgroup estimates are from regression models restricted to students in that subgroup; overall effects are from models including all students. All estimates are for the impact of participating in at least one session of tutoring in math or ELA in 2022-23. Because GPA measures in 2023-24 are not currently available, we are unable to calculate NTF index measures in 2023-24 at this time.

Appendix Table A8. Estimated Impacts of Summer School 2023 on Non-Test Outcomes in 2023-24, by Student Subgroup

	Non-Test Factor Index			Log of Days Absent			Log of Disciplinary Referrals			Log of Days Suspended		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8	(9)	(10)	(11)	(12)
	All (3-9)	Elem (3-5)	MS (6-8)	All (3-9)	Elem (3-5)	MS (6-8)	All (3-9)	Elem (3-5)	MS (6-8)	All (3-9)	Elem (3-5)	MS (6-8)
Overall Effect:	0.0417	0.0530	0.0345	-0.119***	-0.167***	-0.0849	-0.00284	0.00899	-0.0110	-0.0246	-0.00690	-0.0382
	(0.0298)	(0.0468)	(0.0393)	(0.0317)	(0.0393)	(0.0470)	(0.0178)	(0.0196)	(0.0278)	(0.0172)	(0.00960)	(0.0297)
Subgroup:												
Male	0.0588	0.0786	0.0462	-0.103**	-0.226***	-0.0285	-0.0171	0.00845	-0.0313	-0.0294	-0.00526	-0.0432
	-0.0437	-0.0813	-0.0497	-0.0453	-0.0628	-0.0581	-0.0251	-0.0356	-0.0343	-0.0269	-0.0179	-0.0421
Race: Asian	-0.0352	0.115	-0.203	-0.0562	-0.126	0.0639	0.0389	-0.00935	0.0912	0.103	0.00196	0.227
	-0.12	-0.0833	-0.255	-0.124	-0.149	-0.201	-0.0596	-0.0192	-0.139	-0.0955	-0.0121	-0.233
Race: Black	0.0387	0.0127	0.0579	-0.0812*	-0.0784	-0.0842	-0.00348	0.0137	-0.0152	-0.0537**	-0.0169	-0.0786*
	-0.0438	-0.0759	-0.0538	-0.0454	-0.0648	-0.0623	-0.0247	-0.034	-0.0352	-0.0268	-0.0171	-0.0418
Race: Hispanic	0.0293	0.0482	0.0146	-0.153**	-0.216**	-0.0976	0.00359	0.0119	-0.00224	0.00477	0.0134	-0.00186
	-0.0465	-0.0717	-0.0614	-0.0624	-0.084	-0.0871	-0.0239	-0.0227	-0.0384	-0.0273	-0.0223	-0.0469
Race: White	0.0874	0.0808	0.0987	-0.131	-0.215**	-0.0783	-0.0361	0.00799	-0.0731	-0.0433	-0.000512	-0.0812
	-0.0652	-0.0837	-0.102	-0.0995	-0.1	-0.175	-0.0316	-0.0377	-0.0505	-0.042	-0.0159	-0.0805
Students w/Disabilities	-0.0125	0.0266	-0.0404	-0.151***	-0.186*	-0.127*	0.0504	0.0147	0.0755	0.000464	0.00227	0.000106
	-0.0646	-0.0883	-0.0937	-0.0571	-0.0964	-0.0711	-0.0444	-0.0435	-0.0718	-0.0409	-0.0221	-0.0688
Gifted	-0.133	0.0665	-0.236	0.0244	-0.00273	0.0686	0.00514	-0.054	0.0476	0.0681	-0.0353	0.127
	-0.115	-0.127	-0.141	-0.24	-0.248	-0.33	-0.128	-0.0643	-0.166	-0.109	-0.0295	-0.155
ELL Student	0.081	0.104*	0.0713	-0.141**	-0.179*	-0.11	-0.0261	-0.014	-0.041	-0.0126	-0.00941	-0.0221
	-0.0519	-0.0616	-0.0838	-0.0684	-0.101	-0.0905	-0.0322	-0.015	-0.059	-0.0374	-0.00628	-0.0715

*p<.05 **p<.01 ***p<.001

Notes: Estimates shown are from value-added models controlling for participation in SY 2022-23 Tutoring (and other available recovery interventions during that time), student characteristics, prior student test scores, prior non-test factor (NTF) index, and school-by-grade fixed effects. The baseline and outcome terms are SY 2021-22 and SY 2023-24, respectively. Each impact estimate is from a separate regression. Subgroup estimates are from regression models restricted to students in that subgroup; overall effects are from models including all students. All estimates are for the impact of participating in at least one session of tutoring in math or ELA in 2022-23. Because GPA measures in 2023-24 are not currently available, we are unable to calculate NTF index measures in 2023-24 at this time.