Texas Education Agency
Charter School Program Grant Impact Report
School Years 2021–22 through 2023–24

Prepared by McREL International

Submitted to Texas Education Agency

July 2025

Contents

List of Tables

List of Figures

Glossary of Acronyms

CSP = Charter School Program

HLM = Hierarchical Linear Model

ICC = Intraclass Correlation Coefficient

NCES = National Center for Education Statistics

PEIMS = Public Education Information Management System

PSM = Propensity Score Matching

RLA = Reading Language Arts

RQ = Research Question

SD = Standard Deviation

STAAR® = State of Texas Assessments of Academic Readiness

SY = School Year

TEA = Texas Education Agency

ToT = Treatment on the Treated

WWC = What Works Clearinghouse

Executive Summary

Background

In 2020, the Texas Education Agency (TEA) was awarded a five-year, $90 million grant from the U.S. Department of Education under the Expanding Opportunity Through Quality Charter Schools Program (CSP). With this grant, TEA offers competitive and non-competitive subgrants to provide financial assistance for the planning, program design, and initial implementation of charter schools that support the growth of high-quality charter schools in Texas, especially those focused on improving academic outcomes for educationally disadvantaged students. These subgrants assist eligible applicants in opening and preparing for the operation of newly authorized charter schools and replicating/expanding high-quality charter school campuses. CSP funding allowed TEA to fund approximately 54 charter school campuses across four cohorts serving students from early childhood through Grade 12 across a variety of specialized foci.

This report provides results of the impact of the CSP grant on indicators hypothesized to predict academic success. Although not inclusive of all possible predictors of student success, the current study focused on the outcomes of attendance, discipline, mathematics achievement, and reading language arts achievement. The study was limited to the first cohort of CSP grantees (Cohort 1), to allow for sufficient time for outcomes to be observed. Outcomes were chosen based on TEA policy relevance and alignment with educational outcomes prioritized by The What Works Clearinghouse (WWC; 2024). The study addressed two research questions (see insert).

Research Question 1:

Did students in the CSP subgrantee schools show more positive final evaluation year (2024) outcomes of academic achievement, attendance, and discipline as compared to similar students in traditional (non-charter) campuses?

Research Question 2:

Did students in Cohort 1 CSP subgrantee campuses demonstrate a more positive growth trajectory in academic achievement, attendance, and discipline from school year 2021–22 to school year 2023–24 as compared to similar students in traditional (non-charter) campuses?

The Study

To evaluate impact, the McREL study team employed a research design that meets the highest level of standards feasible, **a matched-comparison, quasi-experimental design controlling for pre-existing biases that may be due to baseline differences between CSP subgrantees and non-grantees**. Pre-existing differences between these groups were controlled with a rigorous matching strategy called propensity score matching. Outcome and demographic data from CSP subgrantee campuses as well as comparison campuses were acquired from TEA to increase analytic precision. The study team collected student-level data from both CSP and traditional campuses annually from school year 2020–21 through school year 2023–24. **For each impact analysis, subgrantee campuses were matched with a comparison group of comparison campuses at both the student and school levels to allow researchers to provide an unbiased estimate of the impact of the CSP grant on multiple student outcomes across various Texas public school districts**. After establishing baseline equivalence, researchers examined outcomes.

Analytic Sample

The final analytic sample included six CSP schools for the State of Texas Assessments of Academic Readiness (STAAR) outcome analyses (456 CSP students STAAR-Mathematics and 548 CSP students for STAAR-Reading/Reading Language Arts) and 11 schools for attendance and discipline outcomes (1,357 CSP students for both outcomes). The number of matched comparison schools and students also varied, depending on the outcome (the number of comparison schools for STAAR-Mathematics was 161 and included 1,803 students; the number of comparison schools for STAAR-Reading/Reading Language Arts was 176 and included 2,343 students; the number of schools for the attendance outcome was 523 and included 7,538 students; and the number of schools for the discipline outcome was 530 and included 7,786 students).

Findings – Research Question 1

Analyses for the first research question revealed **no statistically significant differences** for any of the outcomes:

1. Attending a CSP grantee campus for three years did not impact attendance, discipline, STAAR-Reading Language Arts or STAAR-Mathematics, as indicated by *p*-values that did not approach statistical significance. Hedge's *g* effect sizes were small.

2. CSP students showed lower average attendance rate in the outcome year (2023–24) than their matched comparison peers; however, the Hedge's *g* effect size was small and the difference was not statistically significant.

3. CSP students showed, on average, more disciplinary instances in the outcome year (2023–24) than their matched comparison peers; however, the Hedge's *g* effect size was small and the difference was not statistically significant.

4. CSP students showed, on average, higher scores in the outcome year (2023–24) than their matched comparison peers on STAAR-Reading Language Arts but lower scores, on average, on STAAR-Mathematics outcomes. However, neither of these differences were statistically significant and Hedge's *g* effect sizes were small.

5. Attending a CSP grantee campus did not differentially impact students identified as economically disadvantaged versus students not identified as economically disadvantaged.

Findings – Research Question 2

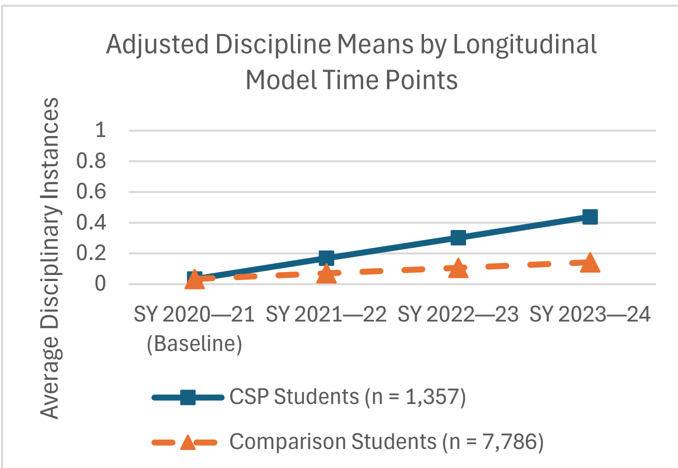
Analyses for the second research question revealed the following (graphs are provided only for statistically significant findings):[1]

1

Both CSP and comparison students showed slightly negative growth in attendance rates over time (school year 2020–21 through school year 2023–24) (no significant difference between the two groups).

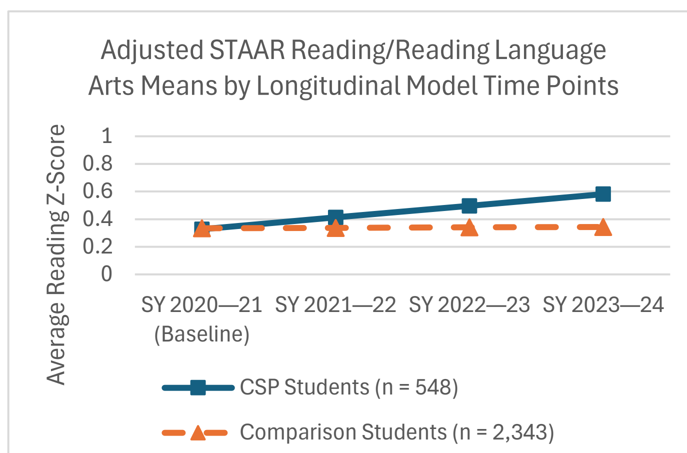
2

Both CSP and comparison students showed an increase in disciplinary instances over time; however, the increase was significantly greater for CSP students than their matched comparison peers as indicated by a statistically significant interaction (time X CSP).

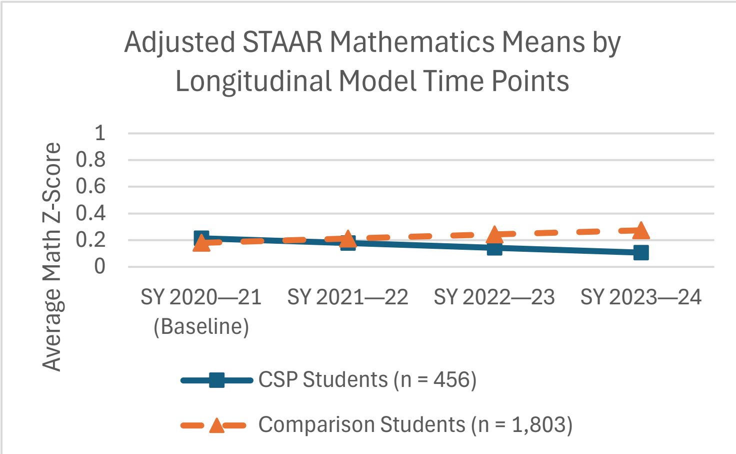


3

CSP students showed significantly more growth over time in STAAR-Reading/Reading Language Arts than their matched comparison peers (statistically significant).



[1] For more information on the source of the graphs, please see Figures 1-3 of the report.



4

CSP students showed significantly less growth over time in STAAR-Mathematics than their matched comparison peers (statistically significant).

5

Attending a CSP school did not differentially impact the growth of students identified as economically disadvantaged versus students not identified as economically disadvantaged.

Conclusions and Implications

The impact analyses show **mixed results** regarding the efficacy of the TEA CSP grant on student outcomes. The following provides conclusions and implications of this study.

Effects on Attendance. Neither of the analytic models showed a statistically significant effect of CSP grantee campus on the attendance outcome. Indeed, both groups showed a slight decrease in attendance rates over time.

Effects on Discipline. Both models showed a difference between CSP and comparison students on discipline, where comparison students had fewer disciplinary instances. This result was not statistically significant in the baseline controlled final year outcome model (Research Question 1; RQ1), however when examining individual growth over the course of the four-year period, the longitudinal growth model (Research Question 2; RQ2), revealed a statistically significant growth difference, with the average number of disciplinary instances for CSP students increasing over time. The reason this effect was observed is not clear. There is insufficient evidence to support an assertion that students in CSP campuses have more *actual* disciplinary instances—rather researchers can only say that CSP students have more *recorded* disciplinary instances. It is possible that CSP campuses have a greater focus on discipline and are more diligent in documenting discipline instances or it may be that CSP campuses operationally define disciplinary infractions differently.

Effects on Student Achievement Outcomes. For STAAR-Mathematics, there was also an observed difference between CSP and comparison students on their ultimate outcome, where CSP students performed lower than comparison students. This was only statistically significant in the longitudinal growth model.

On the other hand, CSP students showed higher STAAR-Reading/Reading Language Arts performance than their comparison counterparts. Again, however, this difference was only detected as statistically significant in the longitudinal growth model that modeled performance over four years of testing. The opposite effects on STAAR-Reading/Reading Language Arts and STAAR-Mathematics suggest that CSP campuses may be effective in supporting student success in reading, more so than traditional public school campuses but not as successful compared to matched traditional schools in mathematics. Researchers recommend follow-up studies to determine whether this effect is generalizable to other cohorts and perhaps leveraging qualitative data to explore the possibility that CSP campuses have a greater focus on reading/reading language arts and less of a focus on mathematics and whether teacher qualifications and experience at the CSP schools—especially in mathematics—may differ from those evident in their traditional school counterparts.

Finally, the CSP grant funded (1) opening and preparing for the operation of newly authorized charter schools and (2) replicating high-quality charters. Because replication campuses are able to rely on existing models and structures, these campuses may have advantages that contribute to their successes. Future studies should examine whether replication campuses outperform newly authorized campuses.

Limitations

Although the rigor of the study design and analyses provides confidence in the findings, the final sample of CSP schools in Cohort 1 was admittedly small. The study was limited to an examination of the first cohort of grantees and further limited to those schools that had the required student-level data for addressing the research questions. This resulted in a total of six schools that were eligible to be included in the examination of achievement outcomes and 11 schools that were eligible to be included in the examination of behavioral outcomes. Moreover, the study was limited in the total number of years that this first cohort was observed. As it is unclear how much time it takes for student outcomes to change at a charter school once they leave an underperforming school, it may be of interest to model a time by enrollment interaction. Furthermore, because the study was limited to a single cohort over just four years, differences among CSP cohort outcomes as well as longer term outcomes associated CSP attendance were unable to be examined. As such, researchers advise replication of these analyses with other cohorts of grantees as well as tracking the outcomes of each cohort and of students over time.

Key Analytic Terms:

Baseline equivalence-A measure of the similarity between the intervention and comparison groups at the baseline of a study. When two groups are similar at baseline it is reasonable to conclude that any differences in the outcomes that are measured at the end of a study (follow-up) are caused by the intervention.

Effect size- A measure of the strength or magnitude of the effect of a program on an outcome (or the strength or magnitude of the association between a program and an outcome) relative to a benchmark.

Hedges' g- An effect size metric as a standardized measure of intervention effects. This metric represents the mean difference between intervention and comparison groups in standard deviation units. That is, the raw mean difference is divided by the variability within the groups, placing effect sizes on a common scale.

Longitudinal data analysis- A research design that measures the same variables of interest repeatedly over a period of time for the same group of participants. This design allows researchers to examine change within individuals and contextual factors that account for interindividual differences.

Propensity score matching-A quasi-experimental design that allows units receiving a treatment (e.g., students; schools) to be matched with and compared to units not receiving a treatment based on the probability that a unit received a particular treatment, given a set of researcher-identified variables related to self-selected treatment participation. Propensity scores are used to adjust outcome analyses to account for self-selection bias, thereby mimicking a randomized controlled study.

Significance- Statistical test that indicates the probability that a relationship among variables as large or larger as that found in a sample could have been drawn randomly from a population in which there is no relationship.

Quasi-experimental design- A research design that attempts to test the causal impact of an independent variable without utilizing random selection and/or random assignment of participants. Families in Texas choose whether their child attends a CSP or a traditional public school; the assignment was not determined by researchers.

Definitions for key analytic terms based on The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation and the What Works Clearinghouse Procedures and Standards Handbook, Version 5.0.

Background

Charter schools, which are part of a broader landscape of education reform, have undergone considerable change over the past decades. While debates about charters still exist, the nature of those debates has largely shifted, with an acceptance that schools of choice are going to remain a feature of the education landscape, coupled with a recognition that charter schools are dominating the national reform agenda in the United States (Mullen & Bartlett, 2022). In theory, charter schools have advantages over public schools because of increased flexibility in pursuing innovation with respect to pedagogy. From the vantage point of pro-charter school advocates, charter schools afford families and caregivers more power in an education marketplace—and the nature of the marketplace incentivizes public schools to compete for students by improving their programs and services. There is also evidence to suggest that charter school attendance can have a longer-term impact on variables such as high school graduation, attendance at two-year post-secondary institutions, and average annual earnings (Sass, Zimmer, Gill, & Booker, 2016). Still open for debate and discussion, however, is whether charter schools accomplish goals of improving student outcomes, necessitating that efforts to support charter schools ought to be accompanied by thorough examinations of direct and associated outcomes.

In Texas, legislation to allow publicly funded charter schools was first passed in 1995 and the rate of growth has been remarkable. In the 2020–21 school year, the National Alliance for Public Charter Schools indicated that Texas was home to an estimated 983 charter schools, serving over 430,000 students.[2] In 2020, the Texas Education Agency (TEA) was awarded a five-year, $90 million grant from the U.S. Department of Education under the Expanding Opportunity Through Quality Charter Schools Program (CSP). This grant was specifically established to provide:

> *...funds to create promising new public charter schools, replicate high-quality public charter schools, and disseminate information about effective practices within charter schools. Federal funds are also available to help charter schools find suitable facilities, reward high-quality charter schools that form exemplary collaborations with traditional public schools, and invest in other national initiatives that support charter schools. Each year, the CSP publishes notices inviting applications in the Federal Register for CSP's federal discretionary grant programs.*[3]

With this grant, TEA offers competitive and non-competitive subgrants to provide financial assistance for the planning, program design, and initial implementation of charter schools that support the growth of high-quality charter schools in Texas, especially those focused on improving academic outcomes for educationally disadvantaged students. These subgrants assist eligible applicants in opening and preparing for the operation of newly authorized charter schools and replicating/ expanding high-quality charter schools.

The CSP funding has allowed TEA to fund approximately 54 charter school campuses across four cohorts.[4] These charter school campuses serve students in early childhood through Grade 12 across a variety of specialized foci (including college preparatory/readiness; career preparation; technology

---

[2] National Alliance for Public Charter Schools, Texas Charter Schools.

[3] U.S. Department of Education: Charter School Programs.

[4] An approximate number is provided for a variety of reasons, including that some charter school campuses were initially funded but returned their funds and other charter school campuses were awarded but not serving students at time of this report.

pathways; science technology, engineering and mathematics; science technology, engineering, arts, and mathematics; International Baccalaureate; community schools; dual language; visual/performing/fine arts; and neurodiversity [such as autism]). Across these campuses, budgets have largely been earmarked for supplies and materials and professional/contracted services (though most funded sites also used their funding for payroll, capital outlay, and other operational costs).

Study Purpose

This report provides results of the impact of the CSP grant on indicators hypothesized to predict academic success. Although not inclusive of all possible predictors of student success, the current study focused on the outcomes of attendance, discipline, mathematics achievement, and reading language arts achievement. These outcomes were chosen based on TEA policy relevance and alignment with educational outcomes prioritized by The What Works Clearinghouse (WWC, 2024).

To test for evidence of impact, the study team used a matched-comparison, quasi-experimental design that controls for pre-existing biases in impact estimates that may be due to baseline differences between CSP subgrantees and comparison grantees. Pre-existing differences between these groups were controlled with a rigorous matching strategy called propensity score matching (PSM), which is a computer-based algorithm that minimizes the overall distance between groups of cases by matching on similar characteristics (Rosenbaum & Rubin, 1985). To use this strategy, outcome and demographic data from CSP subgrantee campuses as well as comparison grantees from the same Texas public school regions were acquired from TEA to increase analytic precision. Specifically, the study team analyzed student-level data from both CSP and traditional campuses annually from school year 2020–21 through school year 2023–24. The sampled cohort received grant funding from SY 2021 to SY 2023; three of the CSP campuses in this evaluation were Charter School Program Grant Subchapter D and the remaining campuses were Charter school Program Grant Subchapter C and D.

For each impact analysis, Cohort 1 CSP subgrantee campuses were matched with a comparison group of comparison campuses at both the student and school levels. Details of the PSM matching process are provided in the Methodology section and in Appendix A. The ultimate goal of the study is to provide an unbiased estimate of the impact of the CSP grant on multiple student outcomes across various Texas public school districts to inform policy makers at both the state and federal level on whether funding for charter school programs provides greater opportunity for student success.

Key Analytic Terms:[5]

- **Baseline equivalence**-A measure of the similarity between the intervention and comparison groups at the baseline of a study. When two groups are similar at baseline it is reasonable to conclude that any differences in the outcomes that are measured at the end of a study (follow-up) are caused by the intervention

[5] Definitions for key analytic terms based on The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation and the What Works Clearinghouse Procedures and Standards Handbook, Version 5.0.

- **Cox's index**- The effect size measure for dichotomous variables preferred by the What Works Clearinghouse. This metric aims to yield effect sizes comparable with Hedges' g for continuous outcomes.
- **Effect size**- A measure of the strength or magnitude of the effect of a program on an outcome (or the strength or magnitude of the association between a program and an outcome) relative to a benchmark.
- **Hedges' g**- An effect size metric as a standardized measure of intervention effects. This metric represents the mean difference between intervention and comparison groups in standard deviation units. That is, the raw mean difference is divided by the variability within the groups, placing effect sizes on a common scale.
- **Hierarchical linear modeling**- A statistical method used to analyze data with a nested structure—such as students within schools—by accounting for the non-independence of observations within clusters. The approach models variation at each level separately (e.g., student and school), recognizing that individuals within the same cluster (e.g., school) are likely to be more similar to each other than to individuals in different clusters.
- Intraclass Correlation Coefficient (ICC)-A parameter ranging from 0-1, representing the proportion of the total variable attributable to the between-school variance.
- **Longitudinal data analysis**- A research design that measures the same variables of interest repeatedly over a period of time for the same group of participants. This design allows researchers to examine change within individuals and contextual factors that account for interindividual differences.
- **Propensity score matching**-A quasi-experimental design that allows units receiving a treatment (e.g., students; schools) to be matched with and compared to units not receiving a treatment based on the probability that a unit received a particular treatment, given a set of researcher-identified variables related to self-selected treatment participation. Propensity scores are used to adjust outcome analyses to account for self-selection bias, thereby mimicking a randomized controlled study.
- **Significance**- Statistical test that indicates the probability that a relationship among variables as large or larger as that found in a sample could have been drawn randomly from a population in which there is no relationship.
- Treatment on Treated (ToT) Analysis- A ToT analysis estimates the causal effect of the program on students who received the treatment over the course of the entire study period (SY 2021–22 through SY 2023–24). This means that students who start in a Cohort 1 CSP campus in SY 2021-22 but leave at some point prior to the ultimate outcome period (SY 2023–24) will be removed from the analysis (not included in either the treatment group or the control group). This contrasts with an Intent to Treat analysis which would consider a student in the treatment group and would analyze that student's outcomes even if they did not remain in the CSP campus over time.
- **Quasi-experimental design**- A research design that attempts to test the causal impact of an independent variable without utilizing random selection and/or random assignment of participants. Families in Texas choose whether their child attends a CSP campus or a traditional public school; the assignment was not determined by researchers.

Methodology

The following section provides the primary research questions for the study followed by an explanation of the sample and the analytic approach for answering each question.

Research Questions

Research Question 1 (RQ1): Did students in Cohort 1 CSP subgrantee campuses show more positive final evaluation year (2024) outcomes of academic achievement, attendance, and discipline as compared to similar students in traditional (non-charter) schools?

Research Question 2 (RQ2): Did students in Cohort 1 CSP subgrantee schools demonstrate a more positive growth trajectory in academic achievement, attendance, and discipline from school year 2021–22 to school year 2023–24 as compared to similar students in traditional (non-charter) schools?

Sample

CSP Group Campuses and Students

Of the original seventeen campuses in the CSP grant Cohort 1, eleven campuses were identified as eligible for the impact analyses presented in this report. Criteria for inclusion of campuses was that the campus began serving students in the 2021–22 school year and have student demographic and outcome data necessary to support the analytic design.[6] While all eleven of these eligible campuses had outcome data for analyses of impact on attendance and discipline, only six had impact data for analyses of impact on student achievement on State of Texas Assessments of Academic Readiness (STAAR®) assessments (STAAR-Mathematics and STAAR-Reading/Reading Language Arts). Specifically, because STAAR grade level assessments are only administered to students in Grades 3–8, six of the schools were dropped from the achievement analyses because they did not serve the necessary grade levels to have student achievement outcome data across all four years of the study.

Criteria for CSP student selection for inclusion in the analytic sample included the following:

1) the student must have baseline data (prior to CSP grant funding) from the 2020–21 school year on the outcome of interest,

2) the student must have remained in the same CSP campus from Year 1 of grant funding (school year 2021–22) through Year 3 of grant funding (school year 2023–24),

3) the student must have demographic data from the fall of Year 1 (school year 2021–22), and

4) the student must have outcome data for the respective analysis from Year 1 (school year 2021–22) through Year 3 (school year 2023–24).

[6] Five schools were excluded from the analysis. Three of the original 17 Cohort 1 campuses were excluded because they began serving students in Year 2 of the grant (school year 2022–23). One of the original 17 Cohort 1 campuses was excluded because they ended their charter contract in September 2023. Finally, one of the original 17 Cohort 1 schools was excluded because it is an alternative school serving a small number of students who are not tested with STAAR.

It is important to note that, for baseline year data (school year 2020–21), an included student could have been enrolled in any Texas public school campus because the CSP campuses were either not in existence during the baseline year or were not receiving any CSP funding.

Comparison (comparison) Group Campuses and Students

The starting population of potential comparison campuses included 8,954 campuses, which were all remaining Texas public school campuses after Cohort 1 CSP campuses were accounted for. From this population, campuses were excluded from the comparison group based on the following criteria:

1) the campuses had alternative instructional designation in the 2022 TEA accountability ratings source file (680 campuses removed),[7]

2) the campus had a charter school designation according to 2022 TEA accountability ratings source file (821 campuses removed),[8]

3) the campus was in a Texas public school region that did not include any CSP campuses (6,174 campuses removed),

4) the campus had a National Center for Education Statistics (NCES) rural locale classification that did not match any of the CSP campuses (286 campuses removed),[9, 10]

5) the campus had a grade type designation of 'B' in the 2022 TEA accountability ratings source file that was not represented in the CSP campus group (3 campuses removed),[11] or

6) the campus in the 2022 Accountability Ratings data source closed and therefore had no enrollment in 2024 (the ultimate outcome year) (24 campuses removed).

Criteria for comparison group student selection for inclusion in the analytic sample included the following:

1) the student must have baseline data (prior to CSP grant funding) from school year 2020–21 on the outcome of interest,

2) the student must have demographic data from the fall of Year 1 (school year 2021–22), and

3) the student must have outcome data for the respective analysis from Year 1 (school year 2021–22) through Year 3 (school year 2023–24).

As with the CSP campuses, an included student could have been enrolled in any Texas public school campuses in the baseline year (school year 2020–21). Additionally, due to the typical progression of

---

[7] Texas Accountability Rating System, Texas Education Agency, 2022.

[8] Because the goal of this study was to compare the outcomes of CSP campuses to traditional public school campuses, non-CSP campuses with charter school designation were removed from the comparison group.

[9] Campus and District Type Data Search.

[10] No Cohort 1 campuses were designated as rural in the NCES database.

[11] Grade type designations in the TEA accountability ratings file include E = Elementary, M = Middle, S = Secondary, and B = Both. No Cohort 1 campuses were designated as B.

students in traditional public schools to move campuses after elementary school to a middle school and then from a middle school to a high school, unlike CSP students, the comparison group students did not have to remain in the same campus from Year 1 of grant funding (school year 2021–22) through Year 3 of grant funding (school year 2023–24). However, to control for potential achievement impacts from mobility in a given school year, comparison students must have remained in the same campus from the fall enrollment to the final 6-week attendance period in a given school year to be included in the analysis. The comparison sample of campuses and students was further reduced using the PSM strategy.

Propensity Score Matching

PSM was used to control for baseline differences between CSP and comparison groups in two phases, first at the campus level and then at the student level. In both matching phases, matching was done using either an exact matching strategy or using logistic regression to obtain a propensity score, depending on the covariate. Exact matching required that comparison campuses were exactly matched to CSP campuses based on a categorical variables deemed important ensuring balance among the groups. PSM estimates the probability of treatment assignment (propensity score) for each campus based on observed covariates using logistic regression, followed by matching on their propensity scores to create groups that are similar in terms of their likelihood of being in the treatment group.

Campus-level matching

For campus level matching, the eleven CSP campuses identified for inclusion in the impact analyses were matched with traditional Texas public school campuses based on demographic characteristics from the fall of the first school year of CSP funding (school year 2021-22). The comparison population of campuses was from the entire population of Texas public schools after various exclusions previously discussed in the Sample section of this report. To ensure the best matches possible for each CSP campus, the study team used the replacement method of PSM, where any comparison school could be matched to more than one CSP campus. Using this method helps to ensure that every CSP campus obtains the best matches regardless of whether the comparison school matched a different CSP campus.

Three categorical variables were used to force campuses into exact matches as follows:

- TEA Region
- Grade level(s) served as categorized by TEA (Elementary, Middle, Secondary).
- Major Locale as categorized by NCES

After forced exact matches, the matching algorithm used propensity scores to match on the following continuous variables:

- Total Enrollment
- Percent Black students
- Percent Hispanic students
- Percent Other Race students[12]

[12] Other Race includes Asian, American Indian, Two or More Races, and Pacific Islander

- Percent students identified as economically disadvantaged
- Percent students identified as Emergent Bilingual/English Learner
- Percent students identified as at-risk
- Percent students identified as special education

Technical details on the campus-level matching are provided in Appendix A.

Student-level matching

After comparison campuses were identified in the campus-level matching phase, the study team acquired from TEA student level data for students in both CSP campuses and matched comparison campuses who were respectively enrolled during the 2021–22 school year. As discussed in the Sample section of the report, students were retained for the student-level matching phase if they met various criteria for inclusion, including having outcome data across all years of the study from 2020–21 (baseline year) through 2023–24 (ultimate outcome year). Student-level matching was conducted separately for each outcome (attendance, discipline, STAAR-Mathematics, and STAAR-Reading/Reading Language Arts).

Student grade level from the 2021–22 school year was the only variable specified for exact matching.

The following student-level variables from the 2021–22 school year were included in the PSM:

- Black status
- Hispanic status
- Other race status
- Economically disadvantaged status
- Emergent bilingual/English learner status
- At-risk status
- Special education status

Technical details on the student-level matching are provided in Appendix A.

The final analytic sample Ns (students and schools) for each outcome after campus- and student-level matching disaggregated by CSP and comparison groups are shown in Table 1.

Table 1. Final analytic samples for each outcome after campus- and student-level matching

Outcome	Total Ns	CSP Ns	Comparison Ns
Attendance	Students: 8,895 Schools: 534	Students: 1,357 Schools: 11	Students: 7,538 Schools: 523
Discipline	Students: 9,143 Schools: 541	Students: 1,357 Schools: 11	Students: 7,786 Schools: 530
STAAR-Mathematics	Students: 2,259 Schools: 167	Students: 456 Schools: 6	Students: 1,803 Schools: 161
STAAR-Reading/Reading Language Arts	Students: 2,891 Schools: 182	Students: 548 Schools: 6	Students: 2,343 Schools: 176

Source. Attendance and discipline data were provided by Texas Education Agency (TEA) at the student level from the end-of-year reporting period. Primary source: Public Education Information Management System data, 2020–21. STAAR-Mathematics and STAAR-Reading performance data were provided by TEA at the student-level. Source: Texas Education Agency, 2020–21. Data for the full population of potential schools and Ns were provided by TEA. The final Ns were calculated by the study team using exclusion criteria and subsequent propensity score matching.
Note. N = number of students or schools. CSP = Charter School Program. STAAR = State of Texas Assessments of Academic Readiness.

Appendix A, Figures 1–4, show how the final sample was reached from the initial population of students/schools for each outcome (attendance, discipline, STAAR-Mathematics and STAAR-Reading/Reading Language Arts), including reasons for removal of student/schools at each step. Appendix A, Table 5 shows the breakdown of student numbers by grade level for each impact analysis outcome.

Analytic Approach to address Research Question 1

The first impact analysis addressed RQ1 to determine the impact of the CSP grant on student achievement, attendance, and discipline in Cohort 1 CSP subgrantee campuses as compared to students in matched comparison traditional campuses. Once the matching was complete, hierarchical linear model (HLM) regression analyses were conducted, controlling for the nesting of students within schools, and included both student- and school-level baseline and demographic covariates from the year prior to grant implementation (2020–2021) and outcome variables from the most recent year of STAAR and PEIMS data (2023–2024). Relevant student- and school-level demographic variables were added to the models for statistical control. School- and student-level covariates included those used in the PSM as described previously. Technical details of the analytic approach are provided in Appendix B.

Analytic Approach to address Research Question 2

The impact analyses for RQ2 used all available student outcome data from school year 2020–21 (baseline) to school year 2023–2024 to model the relationship between receiving a CSP subgrant and changes (or growth) in student outcomes across the years. The rigorous methodology used for predicting change over time is a multilevel longitudinal growth model, which is essentially an HLM model that includes time at level 1, where time is represented by years of data collection for individual student growth (Singer & Willett, 2003). Thus, to estimate relationships with student-level outcomes, time was nested in students, and students were nested in schools (a three-level model). The value of a longitudinal growth model is that it allows for modeling of individual growth of the same students over time. Covariates and outcomes were the same as those assessed in the impact analyses for RQ1. Each of the RQ2 impact analyses compared student outcomes in the CSP subgrantee campuses to students in traditional public schools in the same regions. Technical details of the analytic approach are provided in Appendix B. Tables B-1 and B-2 show intraclass correlations for each of the analytic models. Figures B-1 through B-4 show results of examinations of linear fit for the longitudinal growth models.

Findings

The following sections present the findings from this study along with tables or figures associated with each finding. The first section provides baseline equivalence results. The second section provides the results of the baseline to outcome impact analysis (RQ1). The third section provides the results of the longitudinal growth model (RQ2).

Baseline Equivalence

The first step in the analysis was to establish baseline equivalence by computing descriptive, regression, and effect size statistics on each of the impact analyses. Results are presented in Table 2. As shown in Table 2, baseline equivalence was achieved for all covariates through propensity score matching and model specification.

Table 2. Results of Baseline Equivalence Analysis

Student Characteristic	Attendance Rates (CSP N = 1,357; Comp N = 7,538)	Disciplinary Instances (CSP N = 1,357; Comp N = 7, 786)	STAAR-Reading (CSP N = 548; Comp N = 2,343)	STAAR-Mathematics (CSP N = 456; Comp N = 1,803)
	Effect Size	**Effect Size**	**Effect Size**	**Effect Size**
At Risk	-0.11	-0.09	0.03	-0.25
Economically Disadvantaged	-0.12	-0.14	0.02	-0.04
Emergent Bilingual/English Learner	-0.03	0.01	0.03	0.08
Special Education	-0.14	-0.18	0.12	-0.01
Race				
Black	0.14	0.20	-0.04	0.00
Hispanic	-0.12	-0.13	-0.01	-0.07
Other	0.21	0.21	-0.04	-0.02
Baseline Attendance Rate	-0.12			
Baseline Disciplinary Instances		0.01		
Baseline STAAR-Reading			-0.05	
Baseline STAAR-Mathematics				0.07

Source. Attendance and discipline data were provided by Texas Education Agency (TEA) at the student level from the end-of-year reporting period. Primary source: Public Education Information Management System data, 2020–21. STAAR-Mathematics and STAAR-Reading performance data were provided by TEA at the student-level. Source: Texas Education Agency, 2020–21. Baseline demographic data were provided by TEA at the student level from fall (October) 2021. Primary source: Public Education Information Management System data, 2021–22.

Note. The table effect size statistics represent *Hedges' g* for continuous variables and *Cox's index* for binary variables, following What Works Clearinghouse (WWC) standards. An effect size of 0.05 or smaller indicates baseline equivalence, while effect sizes between 0.05 and 0.25 require statistical adjustment in impact analyses. Effect sizes larger than 0.25 indicate non-equivalence. Effect sizes shown are from post-propensity score matching (PSM) comparisons. Variables needing statistical adjustment were included in the hierarchical linear modeling (HLM) and longitudinal models, which satisfies the WWC adjustment requirement. N = number of students. CSP = Charter School Program students; Comp = comparison students. STAAR = State of Texas Assessments of Academic Readiness. At Risk, Economically Disadvantaged, Emergent Bilingual, Special Education, Black, Hispanic, and Other Race are coded as 1 for yes and 0 for no. Other Race includes Asian, American Indian, Two or More Races, and Pacific Islander. Attendance rates were determined by dividing the days present for each student by the days member for each student. Days member indicates the number of days a student was enrolled in their key campus; days present is the number of days the student attended their key campus. Disciplinary Instances are a count of the number of disciplinary actions documented by TEA for each student in a given school year. STAAR-Reading and STAAR-Mathematics are standardized scaled scores.

Findings: Research Question 1

> *Did students in Cohort 1 CSP subgrantee schools show more positive final evaluation year (2024) outcomes of academic achievement, attendance, and discipline as compared to similar students in traditional (non-charter) schools?*

- The 2023–24 outcome analyses, controlling for 2020–21 baseline and fall 2021–22 demographic characteristics revealed that attending a CSP campus for three years did not impact attendance, discipline, STAAR-Reading Language Arts or STAAR-Mathematics, as indicated by *p*-values that did not approach statistical significance. Hedge's *g* effect sizes were small (see Table 3).
- The observed difference for the attendance outcome was a slight negative effect where CSP students showed lower average attendance rate in the outcome year (2023–24) than their matched comparison peers (not statistically significant) (see Table 3).
- The observed difference for the disciplinary instances outcome was a slight negative effect where CSP students showed, on average, more disciplinary instances in the outcome year (2023–24) than their matched comparison peers (not statistically significant) (see Table 3).
- The observed difference for the STAAR-Reading Language Arts outcome was a slight positive effect where CSP students showed, on average, higher scores in the outcome year (2023–24) than their matched comparison peers (not statistically significant) (see Table 3).
- The observed difference for the STAAR-Mathematics outcome was a slight negative effect where CSP students showed, on average, lower scores in the outcome year (2023–24) than their matched comparison peers (not statistically significant) (see Table 3).
- Exploratory analyses were conducted to examine whether attending a CSP campus differentially impacted students identified as economically disadvantaged versus students not identified as economically disadvantaged by adding an interaction term to the analytic models (CSP attendance X Economic Status). These analyses did not reveal statistically significant moderation of economic status on any of the four outcomes.

Table 3. *Results of Research Question 1 (Baseline-Outcome) Impact Analyses*

Outcome	CSP			Comparison			Adj Mean Diff	Test Statistic (*t*-value)	Sig. Level (*p*-value)	Effect Size (Hedge's *g*)
	Adj Mean	SD	N	Adj Mean	SD	N				
Attendance Rates	0.95	0.07	1,357	0.95	0.06	7,538	-0.00[a]	-0.56	0.58	-0.05
Disciplinary Instances	0.79	1.50	1,357	0.79	1.33	7,786	0.01	0.11	0.91	0.01
STAAR-Reading Language Arts	0.25	0.97	548	0.16	1.00	2,343	0.09	0.83	0.41	0.09
STAAR-Mathematics	0.01	1.05	456	0.04	0.91	1,803	-0.03	-0.19	0.85	-0.03

Source. Attendance and discipline data were provided by Texas Education Agency (TEA) at the student level from the end-of-year reporting period. Primary source: Public Education Information Management System data, 2020–21 and 2023–24. STAAR-Mathematics and STAAR-Reading Language Arts performance data were provided by TEA at the student-level. Source: Texas Education Agency, 2020–21 and 2023–2024. Baseline demographic data were provided by TEA at the student level from fall (October) 2021. Primary source: Public Education Information Management System data, 2021–22.

Note. [a] Decimal places limit the ability to convey the slight negative effect; including an extra decimal place reveals an adjusted mean difference of 0.003. Adj stands for Adjusted. SD stands for Standard Deviation. N is the number of students. Sig stands for Significance. Hedge's g effect size is calculated as the difference between the adjusted CSP student mean and the adjusted comparison (comparison) student mean, divided by the unadjusted pooled standard deviation, which is conceptualized as an average standard deviation across the two groups. N = number of students. CSP = Charter School Program students. STAAR = State of Texas Assessments of Academic Readiness. STAAR-Reading (a previous version of STAAR) is the baseline measure (school year 2020–21) for the ultimate STAAR-Reading Language Arts outcome (school year 2023–24). Attendance rates were determined by dividing the days present for each student by the days member for each student. Days member indicates the number of days a student was enrolled in in their key campus; days present is the number of days the student attended their key campus. Disciplinary instances are a count of the number of disciplinary actions for each student in a given school year. STAAR-Reading/Reading Language Arts and STAAR-Mathematics are standardized scaled scores.

Findings: Research Question 2

Did students in Cohort 1 CSP subgrantee campuses demonstrate a more positive growth trajectory in academic achievement, attendance, and discipline from school year 2021–22 to school year 2023–24 as compared to similar students in traditional (non-charter) campuses?

- Students at both CSP and comparison campuses showed slightly negative growth in attendance rates over time (school year 2020–21 through school year 2023–24) with no significant difference between the two groups (see Table 4).
- Both CSP and comparison students showed an increase in their disciplinary instances over time; however, the increase was significantly greater for CSP students than their matched comparison peers as indicated by a statistically significant time X CSP interaction ($t = 5.41$, $p < .001$) (see Table 4 and Figure 1).
- CSP students showed significantly more growth over time in STAAR-Reading/Reading Language Arts than their matched comparison peers as indicated by a statistically significant

time X CSP interaction ($t = 3.26$, $p = .001$). The trend for CSP students was positive while the trend for comparison students was flat (see Table 4 and Figure 2).

- CSP students showed significantly less growth over time in STAAR-Mathematics than their matched comparison peers as indicated by a statistically significant time X CSP interaction ($t = -2.37$, $p = .02$). The trend for CSP students was negative while the trend for comparison students was positive (see Table 4 and Figure 3).
- The longitudinal growth models were more precise than the basic baseline-outcome analyses as indicated by the revelation of statistically significant growth trend interactions for discipline, STAAR-Reading/Reading Language Arts, and STAAR-Mathematics.
- Exploratory analyses were conducted to examine whether attending a CSP school differentially impacted the growth of students identified as economically disadvantaged versus students not identified as economically disadvantaged by adding an interaction term to each of the longitudinal growth models (Time X CSP Attendance X Economic Status). These analyses did not reveal statistically significant moderation of economic status on any of the four outcomes.

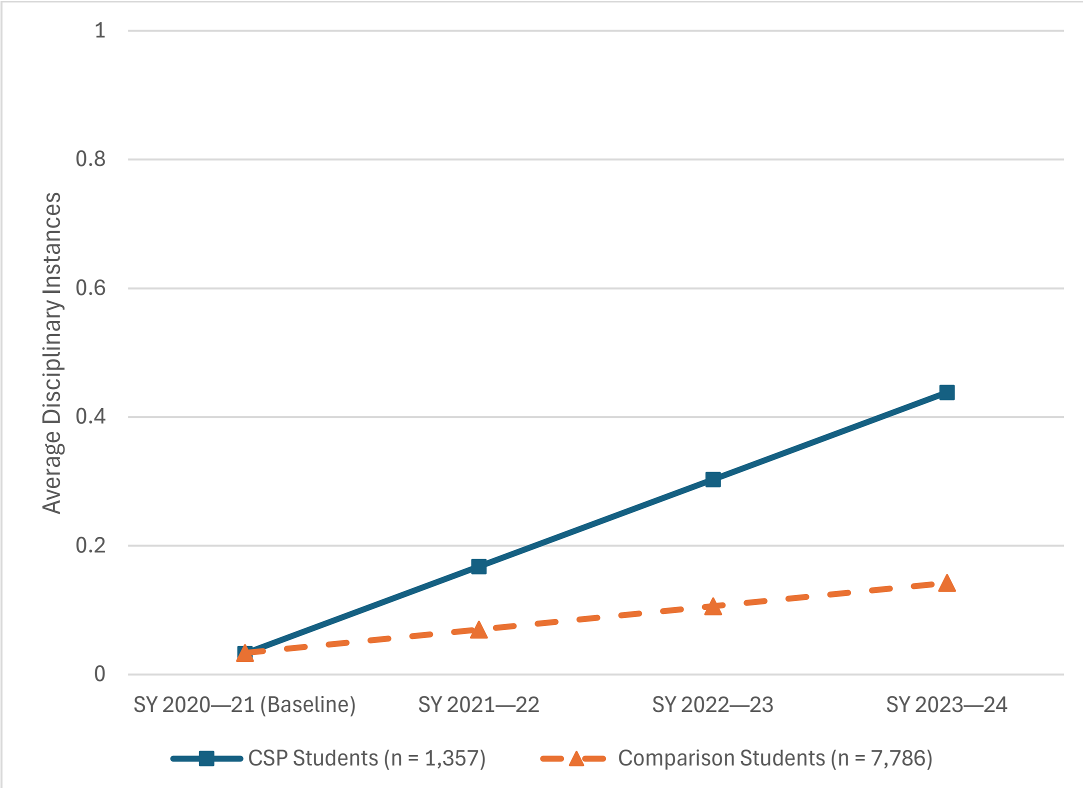
Table 4. Results of Research Question 2 (Longitudinal Growth Model) Impact Analyses

Outcome Area	N CSP	N Comp	Intercept	Time Coefficient	Time x CSP Coefficient	Test Statistic (*t*-value)	Significance Level (*p*-value)	Effect Size
Attendance Rate	1,357	7,538	0.96***	-0.003	-0.00	-0.80	0.42	-0.03
Discipline	1,357	7,786	-0.004	0.04	0.10	5.41	0.00***	0.22
STAAR-Reading/Reading Language Arts	548	2,343	1.03***	0.003	0.08	3.26	0.001***	0.32
STAAR-Mathematics	456	1,803	0.86***	0.03	-0.07	-2.37	0.02**	- 0.22

Source. Attendance and discipline data were provided by Texas Education Agency (TEA) at the student level from the end-of-year reporting period. Primary source: Public Education Information Management System data, 2020–21 through 2023–24. STAAR-Mathematics and STAAR-Reading/Reading Language Arts performance data were provided by TEA at the student-level. Source: Texas Education Agency, 2020–21 through 2023–2024. Baseline demographic data were provided by TEA at the student level from fall (October) 2021. Primary source: Public Education Information Management System data, 2021–22. *Note.* STAAR = State of Texas Assessments of Academic Readiness. STAAR-Reading (a previous version of STAAR) was used for school years 2020–21 and 2021–22. STAAR-RLA (an updated version of STAAR-Reading) was used for school years 2022–23 and 2023–24. N = Number of students; CSP = CSP group students; Comp = Comparison group students. Attendance rates were determined by dividing the days present for each student by the days member for each student. Days member indicates the number of days a student was enrolled in in their key campus; days present is the number of days the student attended their key campus. Disciplinary Instances are a count of the number of disciplinary actions documented by TEA for each student in a given school year. STAAR-Reading/Reading Language Arts and STAAR-Mathematics are standardized scaled scores. The effect size was calculated by computing the difference between adjusted mean pre- to post- change scores for CSP versus comparison schools, standardized by the pooled unadjusted standard deviation. Statistical significance is denoted by asterisks: ** $p < 0.05$, *** $p < 0.01$.

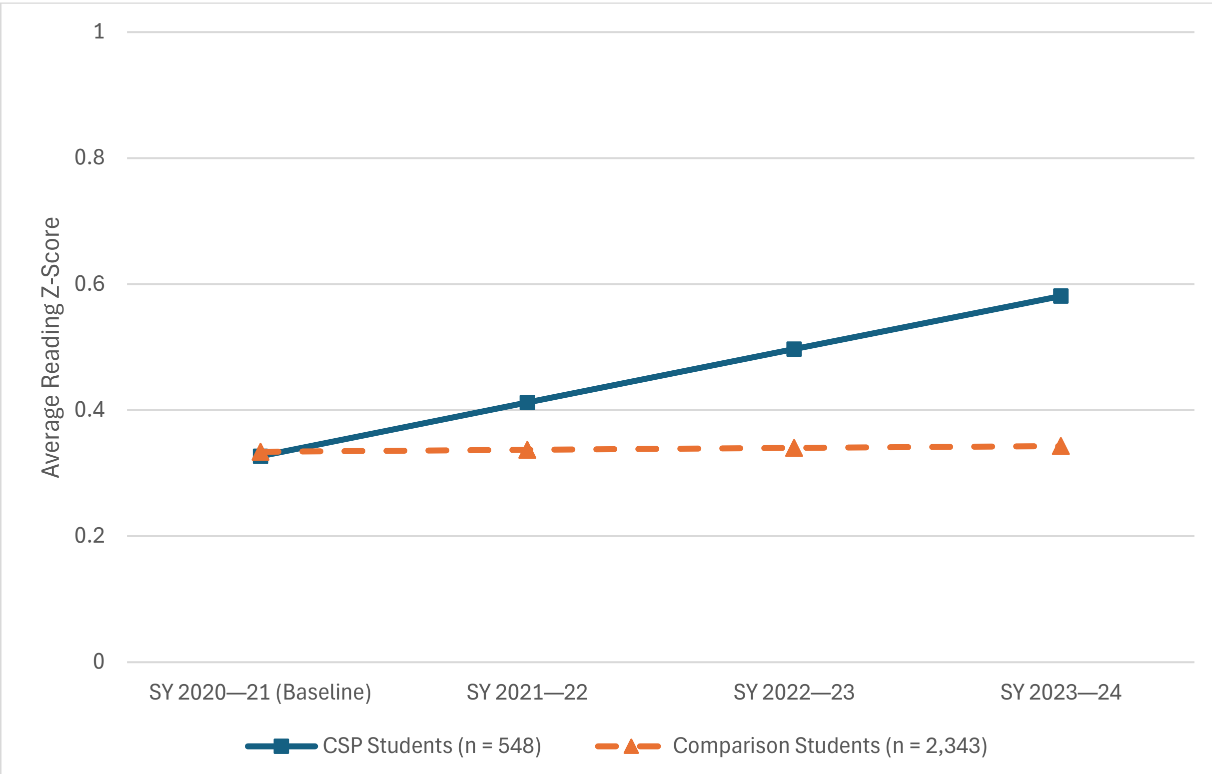
Figures 1–3 show the growth trends for CSP versus comparison students for outcomes with statistically significant outcomes: discipline, STAAR-Reading/Reading Language Arts, and STAAR-Mathematics, respectively.

Figure 1. Adjusted Discipline Means by Longitudinal Model Time Points



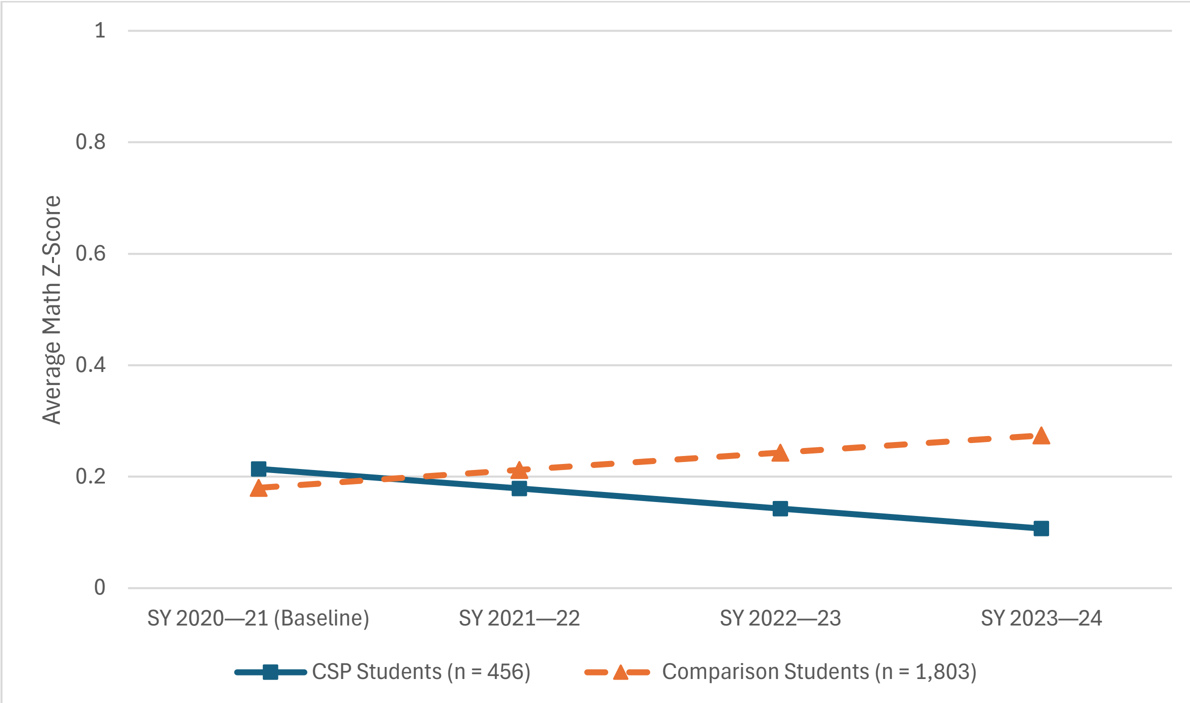
Source. Discipline data were provided by Texas Education Agency (TEA) at the student level from the end-of-year reporting period. Primary source: Public Education Information Management System data, 2020–21 through 2023–24. Baseline demographic data were provided by TEA at the student level from fall (October) 2021. Primary source: Public Education Information Management System data, 2021–22. *Note*. The y-axis represents the average number of disciplinary instances (counts of disciplinary actions for each student in a given school year) across students in each group. SY = School year. CSP = Charter School Program. n = number of students. Model results are statistically significant ($t = 5.41$, $p = 0.00$).

Figure 2. Adjusted STAAR-Reading/Reading Language Arts Means by Longitudinal Model Time Points



Source. STAAR-Reading/Reading Language Arts performance data were provided by Texas Education Agency (TEA) at the student-level. Source: Texas Education Agency, 2020–21 through 2023–24. Baseline demographic data were provided by TEA at the student level from fall (October) 2021. Primary source: Public Education Information Management System data, 2021–22.
Note. SY = School Year. CSP = Charter School Program. STAAR = State of Texas Assessments of Academic Readiness. n = number of students. In 2023, the STAAR test was redesigned to better align with classroom instruction, which necessitated resetting of standards and scales from 2022 to 2023. Therefore, STAAR-Reading was used for SYs 2020–21 and 2021–22 and STAAR-Reading Language Arts was used for SYs 2022–23 and 2023–24. The change in passing standards between the 2022 and 2023 tests should be kept in mind when interpreting results in this report. The Y-axis represents z-scores for the reading outcomes calculated using $z = (x-\mu)/\sigma$, where x is the student scaled score, μ is the grade level mean, and σ is the grade level standard deviation. This means that each student's score was transformed relative to the mean and standard deviation of their grade cohort, allowing for comparisons across grades. Model results are statistically significant ($t = 3.26$, $p = 0.001$).

Figure 3. Adjusted STAAR-Mathematics Means by Longitudinal Model Time Points



Source. STAAR-Mathematics performance data were provided by TEA at the student-level. Source: Texas Education Agency, 2020–21 through 2023–24. Baseline demographic data were provided by TEA at the student level from fall (October) 2021. Primary source: Public Education Information Management System data, 2021–22.
Note. SY = School Year. CSP = Charter School Program. STAAR = State of Texas Assessments of Academic Readiness. n = number of students. The Y-axis represents z-scores for the mathematics outcomes calculated using $z = (x-\mu)/\sigma$, where x is the student scaled score, μ is the grade level mean, and σ is the grade level standard deviation. This means that each student's score was transformed relative to the mean and standard deviation of their grade cohort, allowing for comparisons across grades. Model results are statistically significant (t = -2.37, p = 0.02).

Conclusions

Summary of Findings

Results of the impact analyses show mixed results regarding the efficacy of the TEA CSP grant program on student outcomes. For all outcomes, direction of effects are consistent for both the baseline controlled final year outcome models (RQ1) and the longitudinal growth models (RQ2).

Specifically, neither model showed a statistically significant effect of CSP on the attendance outcome. For the discipline outcome, both models showed a difference between CSP and comparison students where comparison students had fewer disciplinary instances in the ultimate outcome year (school year 2023–24) when controlling for baseline (school year 2020–21); the result was not statistically significant in the baseline controlled final year outcome model (RQ1). However, when examining individual growth over the course of the four-year period, the longitudinal growth model (RQ2), revealed a statistically significant growth difference, where the average number of disciplinary instances for CSP students increased over time, while the average number of disciplinary instances for comparison students remained relatively stable. This is supported by the statistically significant time by CSP interaction on the discipline outcome.

For STAAR-Mathematics, there was also an observed difference between CSP and comparison students on their ultimate outcome, where CSP students performed lower than comparison students; however, only the longitudinal model assessing growth over time was able to detect statistical significance.

Conversely, CSP students showed higher STAAR-Reading/Reading Language Arts performance than comparison students. Again, this difference was only detected as statistically significant using the longitudinal growth model that examined individual-level change in performance across four years of data.

Implications and Recommendations

For the statistically significant negative effect on discipline growth over time, where CSP students showed a larger increase in disciplinary instances over time than comparison students, it is unclear why this effect occurred. That is, researchers cannot say with any degree of certainty that students in CSP campuses have more *actual* disciplinary instances—rather we can only say that CSP students have more *recorded* disciplinary instances. It is possible that CSP campuses have a greater focus on discipline and are more diligent in documenting discipline instances. It is also possible that CSP campuses operationally define disciplinary infractions differently, suggesting the need for a more nuanced approach to cross-site comparisons.

The opposite effects on STAAR-Reading/Reading Language Arts and STAAR-Mathematics are noteworthy. Specifically, the results suggest that CSP campuses were effective for supporting student success in reading, more so than traditional public school campuses; however, they fell behind traditional schools in mathematics. These effects are worth conducting follow-up studies to determine whether this effect is generalizable to other cohorts. It may also be worth collecting qualitative data like educator interviews, surveys, and/or classroom or school observations to explore potential reasons for these effects, for example, whether CSP campuses have a greater focus on reading/reading language arts and less of a focus on mathematics. Moreover, CSP data collection could include an examination of teacher qualifications and experience—including participation in targeted professional development—to determine the extent to which the CSP staff have the requisite backgrounds to drive student learning outcomes in mathematics.

Finally, the CSP grant funded (1) opening and preparing for the operation of newly authorized charter schools and (2) replicating high-quality charters. Because replication campuses are able to rely on existing models and structures, these campuses may have advantages that contribute to their successes. Future studies should examine whether replication campuses outperform newly authorized charter school campuses.

Limitations

Although the rigorous design of this impact study allows for confidence in the findings, the final sample of CSP campuses was small. The study was limited to an examination of the first cohort of grantees and further limited to those schools that had the required data for addressing the research questions. This resulted in a total of six schools that were eligible to be included in the examination of achievement outcomes and 11 schools that were eligible to be included in the examination of behavioral outcomes. Moreover, the study was limited in the total number of years that this first cohort was observed. Additionally, it is unclear how much time it takes for student outcomes to change, once they leave an underperforming school. The current study was unable to examine this time by enrollment interaction. To understand whether findings are consistent over cohorts, whether there are longer term outcomes associated with attendance at a grantee campus (as suggested by prior research on charter schools; see for example, Sass, Zimmer, Gill, & Booker, 2016), and the anticipated length of time for students to experience desired outcomes, researchers advise replication of these analyses with other cohorts of grantees as well as tracking the outcomes of each cohort and of students over time.

References

Mullen, C. A., & Bartlett, T. C. (2022). Charter movement controversy: an American public charter school case study. *Education Inquiry*, *15*(3), 367–384. https://doi.org/10.1080/20004508.2022.2112420

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). Routledge.

Rosenbaum, P. R. & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38. doi: 10.2307/2683903.

Sass, R. T., Zimmer, R. W., Gill, B. P., & Booker, T. K. (2016). Charter high schools' effects on long-term attainment and earnings. *Journal of Policy Analysis and Management*, *35*(3), 683–706.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press. doi: 10.1093/acprof:oso/9780195152968.001.0001

What Works Clearinghouse. (2024, December). *Study Review Protocol, Version 5.1*. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Document/1297

Appendix A: Analytic Sample

Campus-level matching process specifications

- To ensure the best matches possible for each Charter School Program (CSP) campus, the study team used the replacement method of propensity score matching (PSM), where any comparison school could be matched to more than one CSP campus. Using this method helps to ensure that every CSP campus obtains the best matches regardless of whether the comparison school matched a different CSP campus.
- Three categorical variables were used to force campuses into exact matches (variable names in matched dataset in parentheses):
    1) Region (REGION; Note: Only Regions 1, 18, and 20 included)
    2) Grade level(s) served as categorized by TEA (GRDTYPE – Elementary, Middle, Secondary).
    3) Major Locale (LOCALE_AGG: values of 1, 2, 3, – corresponding to City, Town, and Suburban).
- After the above forced exact matches, the matching algorithm used propensity scores to match on the following continuous variables (variable names in matched dataset in parentheses):
    - Total Enrollment (ENROLL)
    - Percent Black (PER_BLACK)
    - Percent Hispanic (PER_HISPANIC)
    - Percent Other Race (PER_OR)
    - Percent Economically Disadvantaged (PER_ECDIS)
    - Percent Emergent Bilingual/English Learner (PER_EBEL)
    - Percent At-risk (PER_RISK)
    - Percent Special Education (PER_SPED)

The final matching ratio was determined by trying several different ratios and assessing the quality of the matches. Specifically, based on PSM literature, standardized mean differences should be within or close to the range of -0.25 to 0.25 to be considered a good match on a particular variable. A 1:10 matching ratio was determined to be the highest ratio that would keep the standardized mean differences within or close to this recommended range for all the variables. Note that some CSP campuses have less than the specified K=10 matched comparisons because there were not enough adequate comparisons for those campuses. Based on inspection of the output from the PSM model using the following criteria suggested by Rubin (2001), the study team concluded that the campus-level PSM process resulted in an adequate matched dataset.[13]

- The ratio of the variances of the propensity scores in the two groups must be close to 1.0. Rubin (2001) suggests that the variance ratios should be between 0.5 and 2.0.

[13] Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2*, 169-188.

- The difference in the means of the propensity scores in the two groups being compared must be small. Rubin (2001) suggests that the standardized differences of means should be less than 0.25.
- For the percent of balance improvement, the larger the percent, the better the PSM results. The percent improvement in balance is defined as follows:

  $100*((|a| - |b|) / |a|)$

  where a is the balance before and b is the balance after matching

The matched campus-level dataset includes:

- A variable labeled CSP that specifies whether the school is a Cohort 1 CSP grantee campus or a comparison campus as follows:
  - Cohort 1 CSP (CSP = "Yes")
  - Comparison (CSP = "No")
- Variables used in the matching (see list above)
- Variables used to determine inclusion in the matching as follows:
  - REGION
  - LOCAL_AGG
  - GRDTYPE
  - CFLATED
  - CFLDAEP
  - CFLJJ
  - CFLRTF
  - CFLCHART
  - SUBCHAPTER_C
- Campus identifying information as follows:
  - Campus Number (CAMPUS)
  - Campus Name (CAMPNAME
  - District Number (DISTRICT)
  - District Name (DISTNAME)
- Matching specifications as follows:
  - Propensity Score for each record (PROP_SCORE)
  - Match weight (MATCH_WGT)
  - Matched ID number that specifies the matched groups (MATCH_ID)
- The final matched dataset includes **12** CSP campuses and **80** Comparison campuses for a Total of **92** campuses.

Student-level matching process specifications

The study team performed student-level PSM on four pre-matching analytic files to create four final matched analytic files as follows:

1) Final matched attendance outcome file
2) Final matched discipline outcome file
3) Final matched State of Texas Assessments of Academic Readiness (STAAR): STAAR-Reading/Reading Language Arts (RLA) outcome file
4) Final matched STAAR-Mathematics outcome file

For each PSM, the research team used an initial 1:10 matching ratio (with replacement) of Cohort 1 CSP campus students to Comparison campus students and evaluated the matches. For attendance and discipline, the ratio was able to be increased to 1:15 with replacement. For STAAR outcomes, it remained at the 1:10 ratio to preserve quality matching. Student-level data were matched exactly within campus groups that were constructed in the school-level matching process.

Each of the PSMs will include the student-level characteristics listed in Table A-1.

Table A-1. Student-level characteristics for student level matching (all outcomes)

Student Variable	Description	Code or Statistic	Match Type[a]
BLACK	Black status (SY 2021–22)	Yes/No	PS
HISPANIC	Hispanic status (SY 2021–22)	Yes/No	PS
OTHER_RACE	Other race status (SY 2021–22)	Yes/No	PS
ECONOMIC	Economically disadvantaged status (SY 2021–22)	Yes/No	PS
EBEL	Emergent bilingual/English learner status (SY 2021–22)	Yes/No	PS
AT_RISK	At-risk status (SY 2021–22)	Yes/No	PS
SPED	Special education status (SY 2021–22)	Yes/No	PS
GRADE	EE–12 for attendance and discipline outcomes (SY 2021–22) 3–8 for STAAR-Reading and Mathematics outcomes (SY 2021–22)	Yes/No	Exact

Source. Baseline demographic data were provided by TEA at the student level from fall (October) 2021. Primary source: Public Education Information Management System data, 2021–22.
Note. [a]Match Type is either Exact or PS; PS = propensity score. SY = School year. STAAR = State of Texas Assessments of Academic Readiness.

In addition to the student level characteristics in Table A-1, the Attendance outcome PSM included baseline (School Year [SY] 2020–21) attendance rate (days present/days member); the Discipline outcome PSM included baseline (SY 2020–21) number of disciplinary actions; and the STAAR-

Reading/Reading Language Arts and STAAR-Mathematics outcomes (standardized scaled scores) included baseline (SY 2020–21) STAAR scores from both Reading and Mathematics assessments.

Tables A-2 through A-6 provide numbers of students and schools by outcome (total, CSP, and comparison). Table A-7 provides the number of CSP and comparison students for the matched dataset for each outcome area by grade level.

Table A-2. Attendance Ns

Step	Total Ns	CSP Ns	Comparison Ns
Begin with Attendance Outcome File	Students: 27,907 Schools: 974	Students: 1,394 Schools: 11	Students: 26,513 Schools: 972
Remove control students (n = 164) that attended a CSP any time after baseline	Students: 27,743 Schools: 971	Students: 1,394 Schools: 11	Students: 26,349 Schools: 960
1:15 PSM **with** replacement	Students: 8,895 Schools: 534	Students: 1,357 Schools: 11	Students: 7,538 Schools: 523

Source. Attendance data were provided by TEA at the student level from the end-of-year reporting period. Primary source: Public Education Information Management System data, 2020–21.

Note. N = Number of students. CSP = Charter School Program. PSM = Propensity Score Match.

Table A-3. Discipline Ns

Step	Total Ns	CSP Ns	Comparison Ns
Begin with Discipline Outcome File	Students: 27,907 Schools: 974	Students: 1,394 Schools: 11	Students: 26,513 Schools: 972
Remove control students (n = 164) that attended a CSP any time after baseline	Students: 27,743 Schools: 971	Students: 1,394 Schools: 11	Students: 26,349 Schools: 960
1:15 PSM with replacement	Students: 9,143 Schools: 541	Students: 1,357 Schools: 11	Students: 7,786 Schools: 530

Source. Discipline data were provided by TEA at the student level from the end-of-year reporting period. Primary source: Public Education Information Management System data, 2020–21.

Note. N = Number of students. CSP = Charter School Program. PSM = Propensity Score Match.

Table A-4. STAAR-Reading Ns

Step	Total Ns	CSP Ns	Comparison Ns
Begin with STAAR-Reading/Reading Language Arts Outcome File	Students: 6,930 Schools: 304	Students: 567 Schools: 6	Students: 6,363 Schools: 302
Remove control students (n = 125) that attended a CSP any time after baseline	Students: 6,805 Schools: 303	Students: 567 Schools: 6	Students: 6,238 Schools: 297
1:10 PSM **with** replacement	Students: 2,891 Schools: 182	Students: 548 Schools: 6	Students: 2,343 Schools: 176

Source. STAAR-Reading performance data were provided by TEA at the student-level. Source: Texas Education Agency, 2020–21.

Note. STAAR = State of Texas Assessments of Academic Readiness. N = Number of students. CSP = Charter School Program. PSM = Propensity Score Match.

Table A-5. STAAR-Mathematics Ns

Step	Total Ns	CSP Ns	Comparison Ns
Begin with STAAR-Mathematics Outcome File	Students: 5,757 Schools: 285	Students: 474 Schools: 6	Students: 5,283 Schools: 283
Remove control students (n = 121) that attended a CSP any time after baseline	Students: 5,636 Schools: 284	Students: 474 Schools: 6	Students: 5,162 Schools: 278
1:10 PSM **with** replacement	Students: 2,276 Schools: 168	Students: 456 Schools: 6	Students: 1,803 Schools: 162

Source. STAAR-Mathematics performance data were provided by TEA at the student-level. Source: Texas Education Agency, 2020–21.

Note. STAAR = State of Texas Assessments of Academic Readiness. N = Number of students. CSP = Charter School Program. PSM = Propensity score match.

Table A-6. Numbers of Students for Matched Dataset by Outcome Area and Grade Level

Grade Level	Attendance Rates		Disciplinary Instances		STAAR-Reading/Reading Language Arts		STAAR-Mathematics	
	CSP N	Comp N	CSP N	Comp N	CSP N	Comp N	CSP N	Comp N
1	32	181	32	183	0	0	0	0
2	164	804	164	725	0	0	0	0
3	202	1,147	202	1,169	0	0	0	0
4	180	1,214	180	1,299	0	0	0	0
5	179	1,110	179	1,259	0	0	0	0
6	94	634	94	654	83	378	85	404
7	75	577	75	549	62	391	61	332
8	431	1,871	431	1,948	403	1,574	310	1,067
Total	**1,357**	**7,538**	**1,357**	**7,786**	**548**	**2,343**	**456**	**1,803**

Source. Data for the full population of potential schools and Ns were provided by Texas Education Agency (TEA) from Public Education Information Management System (PEIMS). The final Ns were calculated by the study team using exclusion criteria and subsequent propensity score matching.

Note. STAAR = State of Texas Assessments of Academic Readiness. N = Number of students. CSP = Charter School Program; Comp = Comparison.

Appendix B: Detailed Analytic Methods

The study team used the following major types of impact analysis to estimate the causal effect of Charter School Program (CSP) campus membership on various student outcomes:

1) **To address Research Question 1, we conducted a Treatment on the Treated (ToT)[14] impact analyses on school year (SY) 2023–24 student outcomes controlling for (1) baseline SY 2020–21 covariates for the outcome of interest and (2) student demographic characteristics from the fall of school year 2021–22.**
   - These analyses were conducted separately with the four different analytic files from each propensity score match (PSM) file: attendance, discipline, State of Texas Assessments of Academic Readiness (STAAR)- Reading/Reading Language Arts,[15] and STAAR-Mathematics.
     i. The attendance outcome was expressed as days present/days member (attendance rate percentage) for each student
     ii. The discipline outcome was expressed as number of disciplinary instances (including students with values of 0)
     iii. Both the STAAR-Reading/Reading Language Arts and STAAR-Mathematics outcomes were expressed as scaled scores—standardized (transformed to z-scores) in order to include multiple grade levels in the same analysis.
   - For each analysis, the study team used a two-level Hierarchical Linear Model (HLM), with students nested in schools.
     i. The first (student) level included student-level covariates used in the student-level matching and the baseline covariate of the outcome of interest (e.g., for the attendance outcome, baseline SY 2020–21 attendance rate was included as a covariate along with the student-level demographics).
     ii. The second (school) level included a TREATMENT indicator (1/0, where 1 indicates that the student remained in the Cohort 1 CSP campus across all 3 three years (SY 2021–22 through SY 2023–24, and 0 indicates that the student did not attend any CSP campus across all three years). It also included school-level covariates used in the school level matching described in Appendix B.
   - An additional exploratory analysis included a cross-level TREATMENT X ECONOMIC interaction term to determine if the causal effect of CSP membership was

[14] A ToT analysis estimates the causal effect of the program on students who received the treatment over the course of the entire study period (SY 2021–22 through SY 2023–24). This means that students who start in a Cohort 1 CSP campus in SY 2021-22 but leave at some point prior to the ultimate outcome period (SY 2023–24) will be removed from the analysis (not included in either the treatment group or the control group). This contrasts with an Intent to Treat analysis which would consider a student in the treatment group and would analyze that students' outcomes even if they did not remain in the CSP campus over time.

[15] In 2023, the STAAR test was redesigned to better align with classroom instruction, which necessitated resetting of standards and scales from 2022 to 2023. The change in passing standards between the 2022 and 2023 tests should be kept in mind when interpreting results in this report

moderated by whether students were identified as economically disadvantaged in the baseline year (SY 2021–22).

The general equation for the impact analyses on SY 2023–24 outcomes is as follows:

Level-1:

$$Y_{ij} = \beta_{0j} + \sum_1^P \beta_{pj} COV_{ij} + \varepsilon_{ij},$$

where:

Y_{ij} is the outcome of student *i* (i=1, … I) in campus *j* (j=1, …, J);

β_{0j} is the regression-adjusted mean value of the student outcome for campus *j*;

β_{pj} is the regression coefficient of the student-level covariate *P* for campus *j*;

COV_{ij} are a list of student associated covariates for student *i* in campus *j*;

ε_{ij} is the residual error term for student *i* in campus *j*, which is assumed to be independently and identically distributed.

Level-2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} CSP_{0j} + \sum_2^x \gamma_{0x} COV_{0j} + u_{0j};$$

$$\beta_{pj} = \gamma_{p0},$$

where:

γ_{00} is the grand mean of the regression-adjusted student outcome score for the comparison campuses;

γ_{01} is the grand mean difference of the regression-adjusted student outcome between the treatment (CSP) and the comparison campuses;

CSP_{0j} is an indicator variable equal to 1 if the student is in a CSP campus *j* and 0 otherwise;

γ_{0x} is the regression coefficient of the x^{th} covariate at the campus level;

COV_{0j} are a list of campus-level covariates including a baseline measure for campus *j*;

u_{0j} is the residual error term for campus *j*, which is assumed to be independently and identically distributed;

γ_{p0} is the grand mean regression coefficient for p^{th} student-level covariate across *J* campuses.

2) **To address Research Question 2, we conducted a ToT longitudinal growth analyses to determine the causal effect of CSP campus membership on individual student growth on student outcomes across four years (from SY 2020–21 to SY 2023–24).**
   - These analyses were conducted separately on all four outcomes: attendance, discipline, STAAR-Reading/Reading Language Arts, and STAAR-Mathematics.
   - For each analysis, the study team used a three-level longitudinal growth model with time nested in students and students nested in schools.
      i. The first level included a TIME indicator where 0 represents SY 2020–21, 1 represents SY 2021–22, 2 represents SY 2022–23, and 3 represents SY 2023–24.
      ii. The second (student) level included student-level covariates for fall of SY 2021–22 demographic characteristics.
      iii. The third (school) level included indicators for TREATMENT and school level covariates used in the school level matching described in Appendix B. A cross-level TIME X TREATMENT interaction was the indicator of interest to determine if membership in a CSP school predicts a positive growth trajectory in the outcome over time.
   - This analysis required adding to the impact dataset outcomes from all school years from SY 2020–21 (baseline – Time 0) through SY 2023–24 (final outcome – Time 3) as opposed to the previous analysis that included only baseline (SY 2020–21) covariates and ultimate outcome (SY 2023–24). It also required transforming the dataset to long (person period) format where students had multiple records representing each time period (Time 0–3).
   - An additional exploratory analysis included a three-way TIME X TREATMENT X ECONOMIC interaction term to determine if the causal effect of CSP membership on individual student growth over time was moderated by whether students were economically disadvantaged.

The equations for the longitudinal model where time is nested within students and students are nested within schools with treatment applied at the school level, covariates at both student and school levels, and a time by treatment interaction are as follows:

Level 1:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(TIME_{1ij}) + \varepsilon_{ijt},$$

where:

Y_{ijt} is the outcome for timepoint *t*, student *i* in school *j*

π_{1ij} is the intercept for student *i* in school *j*

$TIME_{1ij}$ is the measurement occasion

ε_{ijt} is the residual error term at timepoint t which is assumed to be independently and identically distributed

Level 2:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(StudentCov_{ij}) + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}(StudentCov_{ij}) + r_{1ij}$$

where:

$StudentCov_{ij}$ is the student level covariate

β_{00j} is the intercept for school *j*

β_{01j} is the effect of the student covariate on the intercept

β_{10j} is the average slope for school *j*

β_{11j} is the effect of the student covariate on the slope

r_{0ij} and r_{1ij} are student level random effects assumed to be independently and identically distributed

Level 3:

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(SchoolCov_j) + \gamma_{002}(CSP_j) + \mu_{00j}$$

$$\beta_{00j} = \gamma_{100} + \gamma_{101}(SchoolCov_j) + \gamma_{102}(CSP_j) + \mu_{10j}$$

where:

$SchoolCov_j$ is the school-level covariate

CSP_j is the CSP condition (1 = CSP, 0 = comparison), assigned at school level

γ_{000} is the grand mean intercept across schools

γ_{001} is the effect of the school level covariate on the intercept

γ_{002} is the treatment (CSP) effect on the intercept

γ_{100} is the grand mean slope across schools

γ_{101} is the effect of school-level covariate on slope

γ_{102} is the CSP X Time interaction (effect on slope)

μ_{00j} and μ_{10j} are school random effects assumed to be independently and identically distributed

Important considerations for impact analyses

1) Students in comparison campuses at baseline were removed from the analysis if they ever attended any CSP campus across the three-year period from SY 2021–22 to SY 2023–24. This ensured that there was no contamination of the comparison group resulting from exposure to a campus that received a CSP grant.
2) Comparison group students were removed from the analysis if they were mobile during any given school year included in the analysis. Specifically, to eliminate bias to the analysis due to within-school-year student mobility in the comparison group but not in the treatment group, comparison students must have been represented in the same comparison campus in the fall (October) *and* the last 6-week attendance period in a given school year. The school in which the student is represented at both time periods in a given school year was considered the student's "key" school for nesting of students in schools in the analytic models. Comparison group students could be represented in different campuses across school years due to the natural movement of students across traditional schools from elementary to middle school and middle school to high school. Comparison group students who moved schools from one school year to the next were not removed from the analysis.
3) Attendance rates (days present/days member) for the attendance outcome were calculated using data only from campuses for which a student was represented in both the fall (October) and the last 6-week attendance period ("key" school in a given school year).

Intraclass Correlation Coefficients (ICCs)

The intraclass correlations (ICCs) for the baseline-outcome models addressing Research Question (RQ) 1 are shown in Table B-2.

Table B-1. Intraclass Correlation Coefficients (ICC) from Baseline-Outcome HLM

	Attendance	Discipline	STAAR-Reading Language Arts	STAAR-Mathematics
Adjusted ICC	0.08	0.05	0.11	0.16

Note. HLM = Hierarchical Linear Model. ICC = Intraclass correlation coefficient. STAAR = State of Texas Assessments of Academic Readiness.

- For discipline and attendance, the ICCs are considered low (Hox, 2010),[16] which means that schools explain only a small part of the variability in the outcome. Most of the variation is at the individual (student) level.

- For STAAR-Reading Language Arts and STAAR-Mathematics, the ICCs increase into the moderate range (Hox, 2010), meaning that students within the same school tend to be more similar to each other than to students in different schools — more so than for attendance and discipline; however, the majority of the variance remains at the student level.

[16] Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). Routledge.

- Note that, especially for STAAR-Reading Language Arts and STAAR-Mathematics, there is noticeable clustering effect whereby school context plays a nontrivial role in shaping student outcomes (and therefore warrants the hierarchical modelling applied in this study).

Table B- 2. Intraclass Correlation Coefficients (ICC) from Longitudinal HLM

	Attendance	Discipline	STAAR-Reading/Reading Language Arts	STAAR-Mathematics
Adjusted ICC	0.18	0.53	0.12	0.26

Note. HLM = Hierarchical Linear Model. ICC = Intraclass correlation coefficient. STAAR = State of Texas Assessments of Academic Readiness.

- In the longitudinal growth models, ICCs for attendance and reading are approaching moderate levels, suggesting that students within the same school tend to be more similar to each other in terms of growth over time than to students in different schools; however, the majority of the variance in growth remains at the student level.
- For longitudinal growth in STAAR-Mathematics, the ICC of 0.26 reveals that there is substantial clustering — schools account for a significant proportion of the variability.
- For change in discipline over time, the ICC of 0.53 indicates that approximately half of the variance in the outcome is due to differences between schools, while the other half is due to a combination of differences between students within schools and within-student changes over time. This is considered a large clustering effect (Hox, 2010).