Clark County School District Magnet Schools Assistance Program (MSAP)

Spring 2022 Interim Evaluation Summary Report

Jonathan Nakamoto, Alexis Grant, Jason Snipes, Sara Allender, Tanya Figueroa

April 2022

Grant Award #: U165A170032

© 2022 WestEd. All rights reserved.

Suggested citation: Nakamoto, J., Grant, A., Snipes, J., Allender, S., & Figueroa, T. (2022). *Clark County School District Magnet Schools Assistance Program (MSAP): Spring 2022 interim evaluation summary report*. WestEd.

WestEd is a nonpartisan, nonprofit research, development, and service agency that works with education and other communities throughout the United States and abroad to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has more than a dozen offices nationwide, from Massachusetts, Vermont, Georgia, and Washington, DC, to Arizona and California, with headquarters in San Francisco.

WestEd.org

Table of Contents

Executive Summary

WestEd has made substantial progress on the evidence of promise study for Clark County School District's (CCSD) 2017 Magnet Schools Assistance Program (MSAP) grant. WestEd collected the necessary student-level test score, course credit, and demographic data from CCSD for the study. For WestEd's first interim evaluation summary report to the U.S. Department of Education (Nakamoto et al., 2020), we used quasi-experimental designs (QEDs) for each of the three MSAP-funded schools to examine the impact of the schools in 2018–19 on students' performance on state mathematics, reading/English language arts (ELA), and science assessments as well as enrollment in and completion of STEM courses. As part of the QEDs, WestEd utilized propensity score matching and identified comparison groups of students who were equivalent at baseline to the students enrolled in the MSAP-funded schools. The identification of equivalent comparison groups is an important criterion for meeting What Works Clearinghouse (WWC) Design Standards (U.S. Department of Education, 2020). These analyses showed multiple positive and statistically significant impacts on achievement across the three schools during their first year of magnet implementation.

For WestEd's second interim evaluation summary report (Nakamoto et al., 2021), we again used QEDs and identified comparison groups of students who were equivalent at baseline to the students enrolled in the MSAP-funded schools. For this report, we identified different assessments (i.e., Measures of Academic Progress [MAP] assessments) to measure achievement in 2019–20 (because the state tests were not administered due to the COVID-19 pandemic). The second evaluation summary report found that the schools had several positive and statistically significant impacts on student performance during the schools' second year of magnet implementation.

For the current report (i.e., the third interim evaluation summary report), we added additional findings using the same MAP assessments from 2019–20 to the second interim report based on a randomized controlled trial (RCT) that relied on the lottery of one school in the study that was substantially oversubscribed. This school had a large group of 6th graders who applied for admission to the school prior to 2019–20 but did not gain admission to the school. The use of the RCT allows the study to receive the highest rating from the WWC Design Standards (i.e., "meets without reservations"). The findings from the RCT, which is a more rigorous design than the QEDs, replicated the positive and statistically significant impact on the 6th graders' mathematics achievement found by the QED.

WestEd's final summary report will be submitted at the end of CCSD's grant in fall 2022. The final summary report may include findings based on 2020–21 assessments if WestEd's review of the assessment data show the data are valid and reliable enough to use for the study. Additionally, the final summary report will include findings based on 2021–22 assessments. Before completing the final summary report, WestEd will explore whether the impact of the schools on the reduction of minority group isolation can be examined using a design that meets the WWC Design Standards.

Introduction

Magnet Program Overview

Clark County School District (CCSD) in Nevada received a 5-year U.S. Department of Education Magnet Schools Assistance Program (MSAP) grant in 2017 to develop and implement three magnet schools:

- Roger D. Gehring Academy of Science and Technology
- Lied STEM Academy
- Mike O'Callaghan i³ Learn Academy

The aim of the three magnet schools is to reduce, eliminate, or prevent minority group and socioeconomic isolation while improving student achievement and offering students the opportunity to attend a school with a Science, Technology, Engineering, and Math (STEM) program that aligns with their interests. An overview of the three schools and their themes is included in Exhibit 1. During the planning year of the grant (2017–18), each school conducted outreach and recruitment activities; developed and refined the school themes, initial curricula, and instructional approaches; identified and secured initial community partners; and created professional development plans. The first year of MSAP implementation for the three schools occurred in 2018–19.

Exhibit 1. Overview of Clark County School District's Magnet Schools Assistance Program Grant–Funded Schools

School name	Grades served	Magnet theme	Curricula and instructional details
Roger D. Gehring Academy of Science and Technology	1–5	STEM – Project Lead the Way (PLTW) Launch	• Magnet coursework via STEM labs in computer science, engineering, and biomedical science • Blended learning with 1:1 Chromebooks • Project-Based Learning (PBL)
Lied STEM Academy	6–8	STEM – PLTW	• Magnet "pathways" in architecture and construction technology; robotics, automation, and manufacturing; and computer science with IT and cybersecurity • Inquiry-based learning • Blended learning with 1:1 Chromebooks • PBL (e.g., STEM tank)
Mike O'Callaghan i³ Learn Academy	6–8	STEM – Blended Learning and Digital Media	• Magnet "tracks" in video game design, video production, web design, and robotics/engineering • Blended learning with 1:1 Chromebooks • PBL

Study Overview

The goal of WestEd's evidence of promise study for CCSD's MSAP grant is to rigorously evaluate the impact of the three MSAP-funded magnet schools on students' performance on mathematics, reading/English language arts (ELA), and science assessments and students' enrollment in and completion of STEM courses. Specifically, WestEd aims to employ quasi-experimental designs (QEDs) and randomized controlled trials (RCTs) that will meet WWC Design Standards (U.S. Department of Education, 2020). The QEDs and RCTs will produce evidence of promise if they identify statistically significant or substantively important (i.e., a difference of 0.25 standard deviations or larger) effects of the MSAP-funded schools on student outcomes. WestEd's goals for the evidence of promise study are aligned with CCSD's goal for the grant of improving student achievement.

Research Questions (Study Goals)

The three research questions guiding this interim report for our evidence of promise study are outlined below. Research questions 1 and 2 are addressed by the QEDs for each participating school. For Lied STEM Academy, research questions 1 and 2 are also addressed by the RCT. Research question 3 is addressed by school-level descriptive data.

1. What impact does attendance at an MSAP-funded magnet school have on students' performance on mathematics, reading/ELA, and science assessments?
2. What impact does attendance at an MSAP-funded magnet school have on students' enrollment in and completion of STEM courses?
3. Do MSAP-funded magnet schools reduce, eliminate, or prevent minority group isolation?

Additionally, prior to developing the final summary report for the evidence of promise study, WestEd will explore ways to rigorously evaluate the impact of the MSAP-funded magnet schools on reducing, eliminating, or preventing minority group isolation that could meet WWC Design Standards.

Methodology

Study and Sampling Design (Study Contrast)

Quasi-Experimental Design

For this interim report, WestEd used a QED for each school that included a matched comparison group of students and pretest measures of academic achievement (Shadish et al., 2002). WestEd identified the matched comparison students using propensity score matching (Stuart, 2010). The outcome measures, which are listed below, were collected in the winter of the 2019–20 school year (i.e., the second year of MSAP implementation). Unlike the analysis conducted for the interim evaluation summary report, scores for 2019–20 were not available for the Smarter Balanced Assessment Consortium (SBAC) tests and the Nevada Criterion Referenced Tests (CRT) because the tests were not administered in the spring of 2020 due to the COVID-19 pandemic. However, mathematics, reading, and science scores from the Measures of Academic Progress (MAP) assessments based on the winter 2020 test administrations, which were administered before the schools closed for in-person instruction, were available for the evaluation. The spring 2020 MAP assessments were not administered to all students and the schools expressed concerns about the reliability of the data that were collected.

The outcome measures by grade level for the QEDs were as follows:

- Grades 1–5: Achievement on the mathematics and reading MAP assessments
- Grades 3–5: Achievement on the science MAP assessment
- Grades 6–8: Achievement on the mathematics, reading, and science MAP assessments and number of STEM credits attempted and completed

The demographic characteristics used in the identification of the comparison groups and impact analyses were based on 2019–20 data. The pretest achievement measures were collected at the end of 2017–18 or 2018–19 and varied depending on when the students enrolled in the MSAP schools (see Exhibit 2). For students in Grades 2–5 and Grades 7–8 in 2019–20, pretest or baseline measures from 2017–18 were used because these students were enrolled in the MSAP schools in 2018–19. For students in Grades 1 and 6 in 2019–20, pretest measures from 2018–19 were used because these students were enrolled in the MSAP schools for the first time in 2019–20. The QED analyses with the achievement measures, therefore, allowed for the examination of the impact of attendance at the MSAP schools for 1.5 years for students in Grades 2–5 and Grades 7–8 in 2019–20 and for 0.5 years for students in Grades 1 and 6 in 2019–20.

Exhibit 2. Cohorts Included in the Quasi-Experimental Design

Grade level in 2019–20	Status in 2017–18	Status in 2018–19	Status in 2019–20
Grade 1	-	Baseline	Year 1 of Enrollment
Grade 2	Baseline	Year 1 of Enrollment	Year 2 of Enrollment
Grade 3	Baseline	Year 1 of Enrollment	Year 2 of Enrollment
Grade 4	Baseline	Year 1 of Enrollment	Year 2 of Enrollment
Grade 5	Baseline	Year 1 of Enrollment	Year 2 of Enrollment
Grade 6	-	Baseline	Year 1 of Enrollment
Grade 7	Baseline	Year 1 of Enrollment	Year 2 of Enrollment
Grade 8	Baseline	Year 1 of Enrollment	Year 2 of Enrollment

***Note*. Students in Grade 6 in 2019–20 at Lied were also included in the RCT.**

The sample of treatment students included in the analyses for each QED included all students enrolled in the MSAP schools during 2018–19 (for students in Grades 2–5 and Grades 7–8 in 2019–20) and 2019–20 (all grade levels) who had complete data on the demographic characteristics, pretest measures of achievement, and outcome measures. At Gehring, 81.62 percent of the 604 students enrolled in the school during 2019–20 had complete data. In addition, 85.69 percent of the 1,356 students enrolled at Lied during 2019–20 had complete data, and 86.93 percent of the 1,331 students enrolled at O'Callaghan during 2019–20 had complete data.

The pool of potential comparison students included all students enrolled in traditional elementary and middle schools in CCSD in 2018–19 and 2019–20, which excluded other magnet schools, charter schools, and other alternative schools. Additionally, students needed complete data on the demographic characteristics, pretest achievement measures, and outcome measures to be included in the pool of potential comparison students.

To increase the statistical power of the analyses, WestEd identified three comparison students for each MSAP student (Shadish et al., 2002). The individual grade-level analyses were all adequately powered (with power of .80) to detect effect sizes at or below 0.22. The sample sizes for the individual grade-level analyses ranged from a low of 91 treatment and 273 comparison students to a high of 532 treatment and 1,596 comparison students. WestEd used *PowerUp!* (Dong & Maynard, 2013) to calculate the minimum detectable effect sizes for the QEDs. For the smallest sample size (i.e., the Gehring Grade 3 analysis), the QED was powered (with power of .80) to detect effect sizes in the range of 0.17 to 0.22. For the largest sample size (i.e., the Lied Grade 6 analysis), the QED was powered (with power of .80) to detect effect sizes in the range of 0.06 to 0.07. The different proportions of variance in the outcomes explained by pretest measures and demographic characteristics resulted in the ranges for the different sample sizes.

Each of the QEDs includes only one treatment school, which could be considered a confounding factor because the treatment group contains a single study unit. However, according to the WWC, when the intervention under study is "a school with unique organization and governance," such as a magnet school, "the WWC does not consider this to be a confounding factor because the school and the intervention are the same" (U.S. Department of Education, 2020, p. 86). Consistent with this argument, other past studies of charter (Tuttle et al., 2015) and magnet schools (Bifulco et al., 2009) that met WWC Design Standards have used similar designs with single treatment schools.

Randomized Controlled Trial

In the grant application, WestEd also proposed to use RCTs based on CCSD's magnet school lottery to examine the impact of the magnet schools whenever possible. CCSD's use of an oversubscription lottery to determine admittance into its magnet schools when the number of interested students exceeds the capacity of the schools would allow WestEd to conduct an "opportunistic experiment" (Resch et al., 2014) and exploit the random assignment of students to a treatment group that has access to the magnet school or a control group that does not. However, there were not enough students who participated in the general lotteries[1] prior to 2018–19 to allow WestEd to implement an RCT. For example, prior to 2018–19 at Lied, 552 incoming 6th graders applied and put the school as their first choice. Students who applied and did not have preferences based on siblings or feeder schools were randomly assigned to be admitted to the school or to go on a waitlist to be used if students who were admitted chose not to attend Lied. After going through the waitlist to fill the available seats, only 17 students "lost" the lottery and were not offered admission. Since Lied filled all of the available seats with the first choice students, the students who put Lied as their second or third choice did not participate in the lottery. Additionally, there were not enough students who participated in Gehring and O'Callaghan's general lotteries prior to 2019–20 to allow WestEd to implement RCTs with these two schools for the current report. However, there were enough students who participated in Lied's general lottery prior to 2019–20 to allow WestEd to implement an RCT.

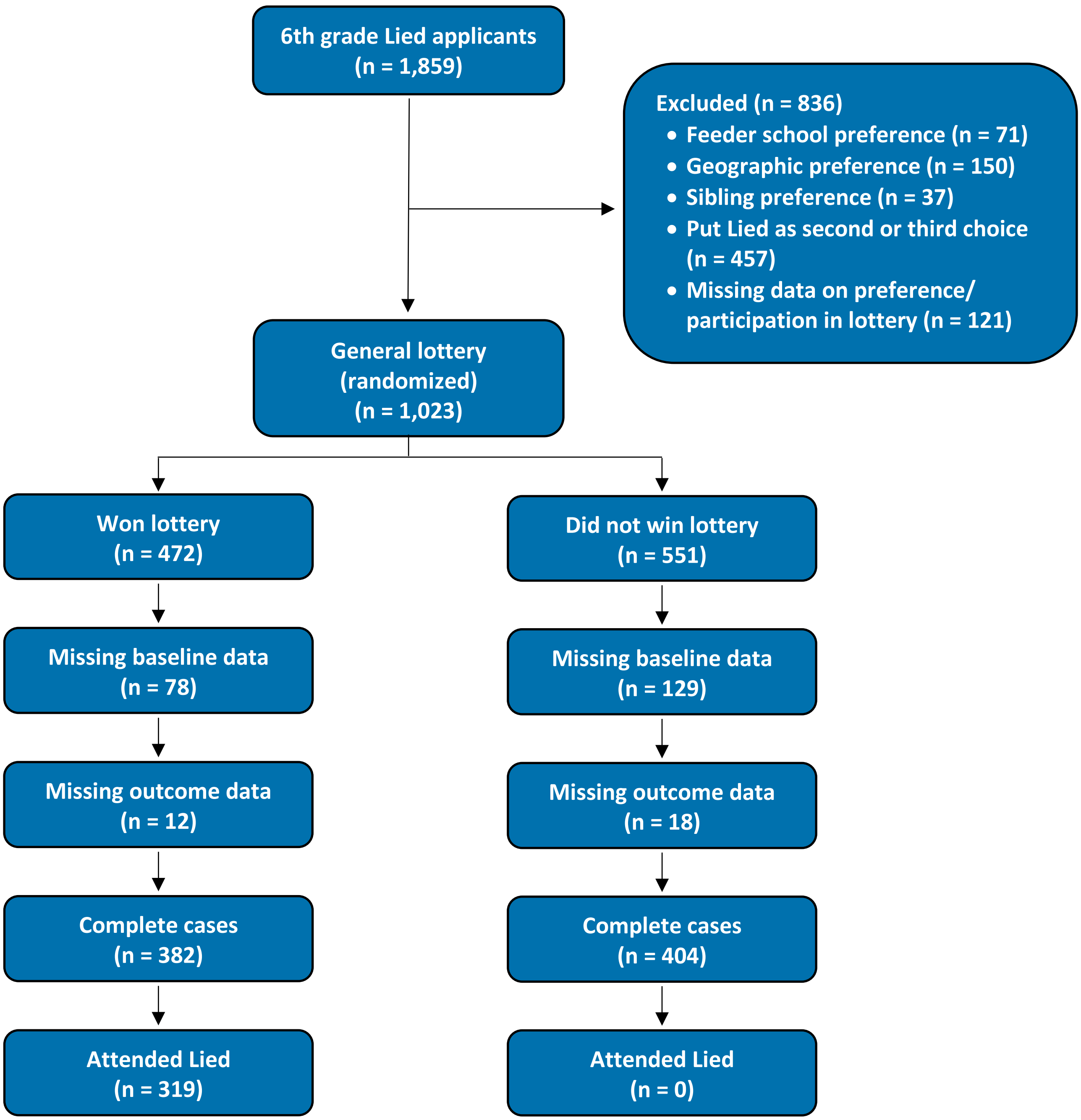
The flow diagram with text boxes and arrows shown in Exhibit 3 tracks the total number of Lied applicants (shown at the top of the diagram) through the randomization and analysis stages (shown at the bottom of the diagram) of the study. As shown in Exhibit 3, Lied had a total of 1,859 students apply as incoming 6th graders at the school in 2019–20. Eight hundred and thirty-six students were excluded from the general lottery for a number of reasons. Students who participated in the feeder school (n = 71), geographic (n = 150), or sibling preference (n = 37) lotteries, which allowed them to gain admission because the number of students selecting each of these preferences was under the number that CCSD reserved for each of these categories, were excluded from the study. Additionally, 457 students put Lied as their second or third choice. Since Lied filled the number of available seats with just the first choice applicants, the students who put Lied as their second or third choice were not randomized. Another 121 students who were missing data on preference/participation in lottery were excluded from the study.

The middle of the flow diagram shows that of 1,023 students that participated in the general lottery, nearly half were offered admission (i.e., they won the lottery; n = 472) and the remaining students were not offered admission (i.e., they did not win the lottery; n = 551). Of the students who won the lottery, 78 were missing baseline data and 12 were missing outcome data, resulting in 382 complete cases. Of the students who did not win the lottery, 129 were missing baseline data and 18 were missing outcome

[1] Students who have preferences based on siblings, attendance at feeder schools, or their geographic location participate in separate lotteries, which have high rates of admission.

data, resulting in 404 complete cases. At the bottom of the flow diagram, it shows that 319 of the students with complete cases who won the lottery attended Lied and zero students who did not win the lottery attended Lied.

Exhibit 3. Flow Diagram Outlining the Sample Selection for the Lied RCT in 2019–20



The overall attrition rate based on the general lottery was 23.17 percent (i.e., 237 students with incomplete data out of the 1,023 general lottery participants). The attrition rate for the lottery winners was 19.07 percent and the rate for students who did not win the lottery was 26.67 percent, making the differential attrition rate 7.60 percent. This combination of overall and differential attrition can be classified as low attrition if there are optimistic assumptions about the relationship between attrition and the study's outcomes (U.S. Department of Education, 2020). Students who did not win the lottery were more likely to have no 2019–20 data in CCSD's data files, indicating they did not enroll in one of the district's schools. Although some students likely moved out of the area, CCSD staff reported that the most likely reason a student would not enroll in a district school was that they opted to attend a charter school. We can assume, therefore, that students who did not win Lied's lottery were more likely to attend a charter school than students who won the lottery.

The outcome measures for the RCT were 2019–20 achievement on the mathematics, reading, and science MAP assessments and the number of STEM credits attempted and completed for students in Grade 6. Pretest measures from 2018–19 were used because these students were enrolled in Lied for the first time in 2019–20. The RCT analyses using the achievement measures, therefore, allowed for the examination of the impact of attendance at Lied for 0.5 years.

WestEd again used *PowerUp!* (Dong & Maynard, 2013) to calculate the minimum detectable effect sizes for the RCT. The RCT was powered (with power of .80) to detect effect sizes between 0.10 and 0.12. We used the sample size for first RCT analysis (n = 786) for the power analysis. The different proportions of variance in the outcomes explained by pretest measures and demographic characteristics resulted in the range for the minimum detectable effect sizes.

Treatment and Treatment Fidelity (Study Intervention)

At Gehring, the grant funds allowed for the instructional day to be extended by an additional 19 minutes for all students. These extra minutes were used for instructional enrichment activities to support the STEM curriculum, and teachers integrated STEM and project-based learning throughout the curriculum. The extent to which teachers implemented project-based learning, blended learning, and personalized learning strategies, which are goals of the grant, was assessed by survey items administered to teachers in the spring of each year (see the Appendix for the items). Additionally, all students participated daily in PLTW lessons, with frequent access to STEM labs. Finally, students could participate in a range of STEM-related extracurricular activities (i.e., robotics club, math club, tortoise habitat, gardening). A student survey administered in the spring of each year asked Gehring students how frequently they participated in each of these activities (see the Appendix for the items).

At Lied and O'Callaghan, each student enrolled in a STEM elective (e.g., a robotics, automation, and manufacturing course) that was taught by licensed teachers hired with MSAP grant funds. Consistent with Gehring, teachers at Lied and O'Callaghan integrated STEM and project-based learning throughout the curriculum. Each spring, the teachers at Lied and O'Callaghan completed the same survey items as those completed by teachers at Gehring, assessing project-based learning, blended learning, and personalized learning strategies. A student survey administered in the spring of each year asked students how frequently they participated in STEM-related extracurricular activities (see the Appendix for the items). During 2018–19 and 2019–20, the students at Lied could participate in a range of STEM-related extracurricular activities (e.g., robotics competition and computer programming competition).

Although no specific STEM-related extracurricular activities were identified at O'Callaghan during 2018–19, the school implemented several STEM-related extracurricular activities in 2019–20.

Data Collection (Measurement)

The data collection plan for WestEd's evidence of promise study changed for the 2019–20 school year because SBAC and science CRT assessments were not administered due to the COVID-19 pandemic. WestEd collected the assessment data and other data elements listed below that were needed to conduct the preliminary evidence of promise analyses that examined the impact on students in winter of 2019–20. The SBAC and science CRT assessments were not administered to all students in 2020–21 due to the COVID-19 pandemic. This will affect our ability to conduct the evidence of promise analyses that examine the impact on students after Year 3 of implementation. Although the MAP assessments were administered throughout 2020–21, students generally completed them at home, and school staff expressed concerns about the reliability of their scores due to "parent participation."

Measures of Academic Progress (MAP) Assessments

Starting in 2019–20, students in kindergarten to Grade 8 in CCSD completed the mathematics and reading MAP assessments in the fall, winter, and spring of each year. Additionally, students in Grades 3–8 completed the science MAP assessment starting in 2019–20. The winter 2020 scores were used as the outcome in the analysis because the spring 2020 assessments were not administered due to COVID-19 closures. Prior to 2019–20, MAP assessments were administered districtwide only to students in kindergarten to Grade 3. Spring data from 2017–18 and 2018–19 were used as some of the baseline measures of student achievement for this report. The goal for the MAP assessments is to monitor students' growth so that instruction can be informed and personalized. The assessments are completed online and have a computer-adapted format. Researchers from the Northwest Evaluation Association (NWEA), the organization that developed the MAP assessment, have found strong evidence for the validity and reliability of the assessments (Bjorklund-Young & Borkoski, 2016).

Smarter Balanced Assessment Consortium (SBAC) Tests

Students in Grades 3–8 in Nevada complete the ELA and mathematics SBAC tests in the spring of each school year. WestEd's initial plan called for these tests to be used as an outcome in the QED. However, these tests were not administered in 2019–20 due to COVID-19 closures. Spring data from 2017–18 and 2018–19 were used as the baseline measures of student achievement for students in Grades 5–8 for this interim report because MAP assessments were not administered in these grade levels prior to 2019–20. The tests, which are aligned with the Nevada Academic Content Standards, are completed online and have a computer-adapted format. There is strong published evidence regarding the validity and reliability of the tests (SBAC, n.d.).

Nevada State Science Assessments

Students in Grades 5 and 8 in Nevada complete the state science accountability assessments (i.e., the Nevada CRTs). WestEd's initial plan called for these assessments to be used as an outcome in the QED. Like the SBAC assessments, the science CRTs were not administered in 2019–20 due to COVID-19 closures. The assessments, which are based on the Nevada Academic Content Standards for Science, are completed online in the spring of each year. The science assessments are valid and reliable (Nevada Department of Education, 2019). Baseline Grade 5 CRT data from 2017–18 (i.e., prior to when the students in Grade 7 in 2019–20 enrolled in the magnet schools) and baseline Grade 5 CRT data from 2018–19 (i.e., prior to when the students in Grade 6 in 2019–20 enrolled in the magnet schools) were used as covariates in the analysis of the MAP assessments.

Science, Technology, Engineering, and Math (STEM) Course Data

CCSD provided WestEd with data files that included all STEM courses, including mathematics, science, and electives. The files included data on the number of credits attempted and earned for all students in the district. The outcomes for the study were the number of STEM credits attempted and earned in 2019–20 for students in Grades 6–8. WestEd summed the number of credits for each student. It should be noted that all schools in CCSD engaged in distance education due to the COVID-19 pandemic starting in March 2020, and this continued through the end of 2019–20. The full impact of the move to distance education on middle schools' grading practices and the course data is unknown. However, there is no reason to believe the practices during distance education differed systematically across the MSAP and comparison schools.

Student Demographic Data

CCSD provided WestEd with data files that included the following student demographic characteristics at the student level for 2019–20: eligibility for free or reduced-price lunch (FRL), race/ethnicity, gender, limited English proficient (LEP) status, and Individualized Education Program (IEP) status. Regardless of their FRL status in the student-level file, WestEd recoded the FRL status to "yes" for all students in schools in CCSD that participated in two programs that served breakfast and lunch at no charge to all students without requiring meal applications.

Lottery Data

CCSD provided student-level magnet school lottery data files for the three schools based on the applications for the 2018–19 and 2019–20 school years. The files indicated whether the students participated in the general lottery or the specific lotteries for siblings, feeder schools, and geographic locations. In addition, the files indicated whether the students were selected (i.e., won the lottery) or were not selected (i.e., did not win the lottery) and whether they put the three MSAP-funded schools as their first, second, or third choice.

Analysis Approach

Propensity Score Matching for the Quasi-Experimental Design

For the QED, WestEd removed all students from the dataset that had missing data on any of the variables included in matching and impact analyses. The use of complete case analysis is in accordance with the WWC standards (U.S. Department of Education, 2020). WestEd also removed any student from the dataset who attended a non-MSAP magnet school or charter school during 2018–19 or 2019–20 to allow for the comparison of students attending the MSAP-funded schools with students attending traditional public schools in CCSD.

WestEd utilized Stata's psmatch2 command (Leuven & Sianesi, 2003) to conduct the propensity score matching and identify the comparison groups of students (Stuart, 2010) for the QED. Propensity score matching is a multivariate matching algorithm that identifies one or more comparison students for each treatment student with similar pretest achievement measures and demographic characteristics. To conduct the propensity score matching, we utilized a logistic regression model with the pretest achievement measures (i.e., mathematics and reading/ELA scores for students in Grades 1–8 and science CRT scores for students in Grades 6 and 7) and demographic characteristics (i.e., FRL status, race/ethnicity, gender, LEP status, and IEP status) to calculate each student's propensity (on a scale from zero to one) to enroll in one of the magnet schools. The demographic characteristics were dummy-coded. Race/ethnicity had four dummy-coded variables that contrasted Hispanics/Latinos with African American/Black, Asian, White, and Other, which included Multiracial, Native American, and Native Hawaiian or Other Pacific Islander. Each treatment student was then matched with a comparison student with the closest propensity score (i.e., nearest-neighbor matching). The propensity score matching technique identifies groups of students who would likely have similar outcomes to the treatment students (i.e., the students enrolled in the MSAP-funded schools) if the treatment students had not enrolled in the MSAP-funded schools.

WestEd conducted the matching separately by school and grade level. The identification of same-grade matches for each treatment student is consistent with Stuart's (2010) recommendation for combining exact matching with propensity score matching. For the MAP assessment analyses, WestEd identified three matched comparison students for each treatment student. For the STEM credit analyses, which were conducted separately for the grant's performance objectives, WestEd identified one matched comparison student for each treatment student. To ensure that each comparison student was selected as a match only once, the matching was done without replacement (Dehejia & Wahba, 2002). That is, each comparison student was removed from the pool of potential comparison students after they were selected as a match.

It would have been optimal to use past performance on a standardized science assessment to match the treatment and comparison students in grades with a science assessment outcome (i.e., Grades 3–8), but this was not feasible because a pretest science assessment was only available when the students were in Grade 5. However, the prior reading/ELA test scores were excellent predictors of the science MAP assessment scores. For example, the correlation between the prior ELA SBAC from 2017–18 and the 5th grade science MAP score in 2019–20 was $r = .76$ with all of CCSD's 5th graders.

Randomized Controlled Trial

To examine impacts using the RCT based on Lied's lottery, we conducted an analysis using students who applied to Lied as incoming 6th graders and were included in the general lottery. As noted above, eligible students for this analysis were general lottery applicants to the 6th grade in Lied in 2019–20, they ranked Lied first, and they did not have any sibling, geographic, or feeder school preference. In total, 1,023 eligible students participated in the general lottery. Prior to analyses for the RCT, WestEd removed all students from the dataset that had missing data on any of the variables included in impact analyses. The use of complete case analysis is in accordance with the WWC standards (U.S. Department of Education, 2020).

WestEd conducted two sets of analyses as part of the RCT. The first analysis, which is termed an intent-to-treat (ITT) analysis (Shadish et al., 2002), compared all the outcomes of students who won the lottery (n = 382), regardless of whether they attended Lied (n = 319) or did not attend Lied (n = 63) in 2019–20, with those students who did not win the lottery (n = 404). The ITT analysis provides an unbiased estimate of the impact of winning the Lied lottery on student outcomes. The second analysis, termed a treatment-on-treated (TOT) analysis (Shadish et al., 2002), compared only those students who won the lottery and attended Lied (n = 319) with those students who did not win the lottery (n = 404). The TOT analysis, however, is quasi-experimental because students who won the lottery but opted to not attend Lied may be different from those who won and opted to attend Lied.

Baseline Equivalence Testing

The next step in the QEDs and RCT was to assess the baseline equivalence of the pretest measures of the outcomes for each analytic sample (i.e., for each grade level within schools) to ensure that the treatment and comparison/control groups were equivalent before the treatment students enrolled in the MSAP-funded magnet schools. For the baseline balance testing, we compared the mean scores of the treatment and comparison/control groups and calculated the effect size (i.e., Hedges' *g*) indexing these differences. When the treatment and comparison/control groups have differences that are below effect sizes of +/- 0.25, they are considered to be equivalent (U.S. Department of Education, 2020). Additionally, we assessed whether the differences between the groups were statistically significant on the pretest measures of the outcomes using *t* tests and on the demographic characteristics using logistic regression.

The propensity score matching approach resulted in comparison groups that were equivalent at baseline (i.e., less than +/- 0.25 standard deviations; U.S. Department of Education, 2020) on all of the pretest achievement measures. The baseline analyses with the samples included in the analysis of the MAP assessments for the QEDs are shown in Exhibits 4, 5, and 6. The baseline analyses for the samples for the STEM credit comparison groups were consistent with the results in Exhibits 4, 5, and 6, and all effect sizes were less than +/- 0.25 standard deviations. The pretest achievement measures varied across grades and depended on the administration schedule of the assessments, the need to have pretest achievement from before the students enrolled in the MSAP schools, and which data were available to WestEd at the time of the analysis. For the student performance analyses, the prior achievement measures are outlined below:

- Grade 1: Mathematics and reading MAP assessment scores from spring 2018–19, when the students were in kindergarten
- Grade 2: Mathematics and reading MAP assessment scores from spring 2017–18, when the students were in kindergarten
- Grade 3: Mathematics and reading MAP assessment scores from spring 2017–18, when the students were in Grade 1
- Grade 4: Mathematics and reading MAP assessment scores from spring 2017–18, when the students were in Grade 2
- Grade 5: Mathematics and ELA SBAC assessment scores from spring 2017–18, when the students were in Grade 3
- Grade 6: Mathematics and ELA SBAC assessment scores and science CRT scores from spring 2018–19, when the students were in Grade 5
- Grade 7: Mathematics and ELA SBAC assessment scores and science CRT scores from spring 2017–18, when the students were in Grade 5
- Grade 8: Mathematics and ELA SBAC assessment scores from spring 2017–18, when the students were in Grade 6

As shown in Exhibit 4, the treatment and comparison students in Grades 1–5 included in the Gehring QED analyses were equivalent at baseline on the mathematics and reading/ELA assessments. All of the effect sizes were less than +/- 0.08 standard deviations (i.e., the effect size), and none of the differences were statistically significant (i.e., no *p* values were below .05).

Exhibit 4. Baseline Equivalence of the Pretest Achievement Measures for the Gehring Quasi-Experimental Design

Grade level/ Assessment	Treatment students mean	Treatment students SD	Treatment students n	Comparison students mean	Comparison students SD	Comparison students n	Treatment–Comparison difference	Effect size	*p* value
Grade 1: MAP mathematics	166.76	11.21	108	166.06	13.33	324	0.70	0.05	.62
Grade 1: MAP reading	162.91	11.84	108	162.92	14.05	324	-0.01	0.00	.99
Grade 2: MAP mathematics	166.97	11.00	102	165.94	12.97	306	1.03	0.08	.47
Grade 2: MAP reading	165.99	12.51	102	164.92	13.95	306	1.07	0.08	.49
Grade 3: MAP mathematics	187.57	11.57	91	187.56	12.05	273	0.01	0.00	.99
Grade 3: MAP reading	186.71	11.51	91	187.56	13.56	273	-0.85	-0.06	.59
Grade 4: MAP mathematics	190.56	8.73	100	190.96	10.76	300	-0.40	-0.04	.74
Grade 4: MAP reading	192.14	12.71	100	192.44	13.59	300	-0.30	-0.02	.84
Grade 5: SBAC mathematics	2,475.05	62.36	92	2,477.84	65.23	276	-2.79	-0.04	.72
Grade 5: SBAC ELA	2,478.85	57.43	92	2,478.74	69.49	276	0.11	0.00	.99

Note**. MAP = Measures of Academic Progress assessment. SBAC = Smarter Balanced Assessment Consortium. Rasch Unit (RIT) scores are presented for the MAP assessments. Scale scores are presented for the SBAC tests. The effect size is Hedges' *g*.**

As shown in Exhibit 5, the treatment and comparison students in Grades 6–8 included in the Lied QED analyses were equivalent at baseline on the mathematics SBAC, ELA SBAC, and science CRT assessments. All of the effect sizes were below +/- 0.04 standard deviations, and none of the differences were statistically significant.

Exhibit 5. Baseline Equivalence of the Pretest Achievement Measures for the Lied Quasi-Experimental Design

Grade Level/ Assessment	Treatment students mean	Treatment students SD	Treatment students n	Comparison students mean	Comparison students SD	Comparison students n	Treatment–Comparison difference	Effect size	*p* value
Grade 6: SBAC mathematics	2,531.15	81.89	532	2,529.64	84.75	1,596	1.52	0.02	.72
Grade 6: SBAC ELA	2,549.30	79.48	532	2,549.89	80.93	1,596	-0.59	-0.01	.88
Grade 6: CRT science	440.60	45.09	532	441.12	48.32	1,596	-0.53	-0.01	.83
Grade 7: SBAC mathematics	2,533.76	73.46	463	2,534.43	78.28	1389	-0.67	-0.01	.87
Grade 7: SBAC ELA	2,543.57	72.20	463	2,546.35	75.59	1389	-2.78	-0.04	.49
Grade 7: CRT science	434.08	46.58	463	435.40	46.20	1389	-1.31	-0.03	.60
Grade 8: SBAC mathematics	2,509.35	85.11	167	2,509.43	85.95	501	-0.07	0.00	.99
Grade 8: SBAC ELA	2,531.20	82.31	167	2,533.61	82.27	501	-2.41	-0.03	.74

***Note*. SBAC = Smarter Balanced Assessment Consortium. ELA = English language arts. CRT = Criterion Referenced Test. Scale scores are presented for the SBAC and CRT tests. The effect size is Hedges' *g*.**

Consistent with the comparisons for the other schools, the treatment and comparison students in Grades 6–8 included in the O'Callaghan QED analyses were equivalent at baseline on the mathematics SBAC, ELA SBAC, and science CRT assessments (see Exhibit 6). All of the effect sizes were below +/- 0.04 standard deviations, and none of the differences were statistically significant.

Exhibit 6. Baseline Equivalence of the Pretest Achievement Measures for the O'Callaghan Quasi-Experimental Design

Grade level/ Assessment	Treatment students mean	Treatment students SD	Treatment students n	Comparison students mean	Comparison students SD	Comparison students n	Treatment–Comparison difference	Effect size	*p* value
Grade 6: SBAC mathematics	2,503.42	80.38	479	2,504.22	83.40	1,437	-0.80	-0.01	.86
Grade 6: SBAC ELA	2,520.38	81.27	479	2,520.84	81.44	1,437	-0.46	-0.01	.92
Grade 6: CRT science	425.04	44.52	479	425.83	44.14	1,437	-0.79	-0.02	.74
Grade 7: SBAC mathematics	2,498.19	72.41	398	2,501.36	80.36	1,194	-3.17	-0.04	.49
Grade 7: SBAC ELA	2,509.04	77.14	398	2,511.97	79.69	1,194	-2.94	-0.04	.52
Grade 7: CRT science	416.25	42.54	398	417.50	42.93	1,194	-1.25	-0.03	.61
Grade 8: SBAC mathematics	2,445.29	65.56	280	2,445.33	78.92	840	-0.03	0.00	.99
Grade 8: SBAC ELA	2,472.47	67.26	280	2,470.19	74.28	840	2.28	0.03	.65

***Note*. SBAC = Smarter Balanced Assessment Consortium. ELA = English language arts. CRT = Criterion Referenced Test. Scale scores are presented for the SBAC and CRT tests. The effect size is Hedges' *g*.**

As shown in Exhibit 7, the treatment and control students included in the Lied RCT ITT analyses were equivalent at baseline on the mathematics SBAC, ELA SBAC, and science CRT assessments. All of the effect sizes were below +/- 0.11 standard deviations, and none of the differences were statistically significant. To meet WWC standards for RCTs, the impact analyses with a sample with this level of baseline equivalence (i.e., between +/- 0.05 and +/- 0.25 standard deviations) require statistical adjustment. That is, the pretest achievement measures need to be included as covariates in the regression models (U.S. Department of Education, 2020). Exhibit 7 also shows the baseline equivalence of treatment and comparison students included in the Lied RCT TOT analysis were equivalent at baseline on the mathematics SBAC, ELA SBAC, and science CRT assessments. The effect sizes for the mathematics SBAC, ELA SBAC, and science CRT assessments were below +/- 0.16 standard deviations, and none of the differences were statistically significant.

Exhibit 7. Baseline Equivalence of the Pretest Achievement Measures for the Lied Randomized Controlled Trial for Students in Grade 6

Sample/ Measure	Treatment students mean	Treatment students SD	Treatment students n	Control students mean	Control students SD	Control students n	Treatment–Control difference	Effect size	*p* value
ITT sample: SBAC mathematics	2,538.31	78.82	382	2,533.05	83.31	404	5.26	0.06	.36
ITT sample: SBAC ELA	2,556.86	75.90	382	2,548.71	84.35	404	8.15	0.10	.16
ITT Sample: CRT science	443.39	42.87	382	439.00	45.51	404	4.39	0.10	.16
TOT sample: SBAC mathematics	2,542.09	78.82	319	2,533.05	83.31	404	9.04	0.11	.14
TOT sample: SBAC ELA	2,560.61	74.75	319	2,548.71	84.35	404	11.91	0.15	.05*
TOT sample: CRT science	445.37	42.62	319	439.00	45.51	404	6.37	0.14	.05

***Note*. ITT = Intent-to-treat. TOT = Treatment-on-treated. SBAC = Smarter Balanced Assessment Consortium. ELA = English language arts. CRT = Criterion Referenced Test. Scale scores are presented for the SBAC and CRT tests. The effect size is Hedges' *g*. *p* values marked with an asterisk (*) were statistically significant.**

The demographic characteristics of the treatment and comparison students included in the Gehring, Lied, and O'Callaghan QED analyses are displayed in Exhibit 8. The percentages are aggregated across grades because the demographic patterns for the individual grades were consistent with the overall patterns shown in Exhibit 8. Consistent with the goal of propensity score matching, the demographic characteristics of the treatment and comparison students were very similar. For example, 51.11 percent of the treatment students at Gehring were eligible for FRL and 49.70 percent of the comparison students were eligible for FRL. Across the three samples, there was less than a 3-percentage-point difference between the treatment and comparison groups on each of the demographic characteristics. Additionally, none of the differences were statistically significant.

Exhibit 8. Demographic Characteristics of the MSAP and Comparison Students Included in the Quasi-Experimental Designs

Demographic characteristic	Gehring sample treatment students %	Gehring sample comparison students %	Gehring sample *p* value	Lied sample treatment students %	Lied sample comparison students %	Lied sample *p* value	O'Callaghan sample treatment students %	O'Callaghan sample comparison students %	O'Callaghan sample *p* value
Free/Reduced-price lunch	51.11	49.70	.59	41.22	41.80	.73	77.61	77.07	.70
Race/Ethnicity: African American/Black	12.78	12.85	.97	12.22	12.02	.86	6.14	6.83	.41
Race/Ethnicity: Asian	14.40	13.05	.45	6.11	6.43	.70	2.85	2.71	.80
Race/Ethnicity: Hispanic/Latino	32.25	30.63	.50	32.96	32.42	.73	75.97	75.54	.77
Race/Ethnicity: White	26.98	29.42	.30	34.85	34.83	.99	10.63	10.29	.74
Race/Ethnicity: Other	13.59	14.06	.79	13.86	14.31	.70	4.41	4.64	.75
Female	51.11	50.98	.96	52.32	51.46	.61	54.28	54.51	.89
Limited English Proficient	7.91	7.17	.58	2.15	1.64	.25	14.26	12.71	.18
Individualized Education Program	6.09	5.48	.61	6.45	5.94	.52	6.05	6.02	.97

***Note*. The Gehring sample included n = 493 treatment students and n = 1,479 comparison students across Grades 1–5. The Lied sample included n = 1,162 treatment students and n = 3,486 comparison students across Grades 6–8. The O'Callaghan sample included n = 1,157 treatment students and n = 3,471 comparison students across Grades 6–8. Other included Multiracial, Native American, and Native Hawaiian or Other Pacific Islander.**

The demographic characteristics of the treatment and control students included in the Lied RCT analyses are shown in Exhibit 9. In both samples, there was less than a 4-percentage-point difference between the treatment and control groups on each of the demographic characteristics with the exception of students eligible for FRL. For both the ITT and TOT samples, the percentage of students eligible for FRL was greater in the control group (i.e., the students who did not win the lottery) than in the treatment group, and these differences were statistically significant. It should be noted that these percentages are based on the analytic samples that had complete data and do not indicate that students eligible for FRL were less likely to win the lottery.

Exhibit 9. Demographic Characteristics of the MSAP and Control Students Included in the Lied Randomized Controlled Trial for Students in Grade 6

Demographic characteristic	Lied ITT sample treatment students %	Lied ITT sample control students %	Lied ITT sample *p* value	Lied TOT sample treatment students %	Lied TOT sample control students %	Lied TOT sample *p* value
Free/Reduced-price lunch	41.10	51.98	<.01*	37.30	51.98	<.001*
Race/Ethnicity: African American/Black	13.09	13.12	.99	11.91	13.12	.63
Race/Ethnicity: Asian	4.71	3.96	.60	4.70	3.96	.63
Race/Ethnicity: Hispanic/Latino	30.89	33.17	.49	29.78	33.17	.33
Race/Ethnicity: White	39.01	36.88	.54	40.44	36.88	.33
Race/Ethnicity: Other	12.30	12.88	.81	13.17	12.88	.91
Female	43.19	46.29	.38	43.89	46.29	.52
Limited English Proficient	1.83	1.49	.70	1.57	1.49	.93
Individualized Education Program	7.07	4.70	.16	5.64	4.70	.57

***Note*. ITT = Intent-to-treat. TOT = Treatment-on-treated. The Lied ITT sample included n = 382 treatment students and n = 404 control students in Grade 6. The Lied ITT sample included n = 319 treatment students and n = 404 control students in Grade 6. Other included Multiracial, Native American, and Native Hawaiian or Other Pacific Islander. *p* values marked with an asterisk (*) were statistically significant.**

Impact Analysis

WestEd conducted the impact analyses for the QEDs and RCT using ordinary least squares regression. The regression analyses were conducted separately by school and grade (for the QED). All of the variables used in the propensity score matching (e.g., mathematics and ELA SBAC scores and the demographic characteristics) to select the comparison groups were entered into the regression models as covariates (Rubin & Thomas, 2000). Utilizing these variables again as covariates in the regression models provided an added level of statistical control for the QEDs and was required to meet WWC standards for the RCT given the level of baseline equivalence on the pretest measures of achievement (U.S. Department of Education, 2020). A dummy-coded variable included in the models contrasting magnet and nonmagnet students provided the estimates of program impacts. The equation below illustrates the basic approach to estimating program effects:

$$Achievement_i = \beta 0 + \beta 1 T_i + \beta 2 PriorMathAch_i + \beta 3 PriorReadAch_i + \beta 4 FRL_i + \beta 5 Asian_i + \beta 6 Black_i + \beta 7 White_i + \beta 8 Other_i + \beta 9 Gender_i + \beta 10 LEP_i + \beta 11 IEP_i + e$$

Where: $Achievement_i$ represents the performance on the mathematics, reading, and science MAP assessment for student *i*; for the analysis of the STEM credit data, the number of credits was used in place of $Achievement_i$; T_i is a treatment indicator that equals one for students in the treatment group and zero otherwise; $PriorMathAch_i$ is the prior mathematics achievement measure (i.e., the test score from the year prior to entering the magnet); $PriorReadAch_i$ is the prior reading or ELA achievement measure (i.e., the test score from the year prior to entering the magnet); prior science achievement was also included as a covariate for Grades 6 and 7. FRL_i is a dichotomous indicator for student eligibility for FRL; $Asian_i$, $Black_i$, $White_i$, and $Other_i$ are each a dichotomous indicator for ethnicity and race; $Gender_i$ is a dichotomous indicator for whether a student is male or female; LEP_i is a dichotomous indicator for LEP and non-LEP students; and IEP_i is a dichotomous indicator for whether a student has an IEP. In this equation, $\beta 1$ represents the impact of attending one of the MSAP-funded magnet schools on student achievement. WestEd used the Benjamini-Hochberg correction for multiple comparisons (Benjamini & Hochberg, 1995) when examining the statistical significance of the various grade-level impact analyses for a single outcome domain (e.g., the five grade-level analyses that examined the impact of Gehring on students' performance on the mathematics assessments).

Data Results

Gehring Quasi-Experimental Design

The individual grade-level results for the mathematics, reading, and science MAP assessments are presented in Exhibit 10 based on the Gehring QED. There were three statistically significant positive impact estimates for mathematics, with effect sizes ranging from 0.24 to 0.34. There were also four statistically significant positive impact estimates for reading, with effect sizes ranging from 0.24 to 0.39. For the mathematics and reading outcomes that differed to a statistically significant extent, the Gehring students scored 5 to 11 percentile points higher than the comparison students. None of the impact estimates for science were statistically significant.

Exhibit 10. Impact Estimates From the Gehring Quasi-Experimental Design

Grade level/ assessment	Treatment students mean	Treatment students SD	Treatment students n	Comparison students mean	Comparison students SD	Comparison students n	Treatment–Comparison difference	Effect size	*p* value
Grade 1: Mathematics	67.23	22.86	108	60.58	28.70	324	6.66	0.24	.001*
Grade 2: Mathematics	60.77	20.07	102	53.22	26.18	306	7.55	0.30	<.001*
Grade 3: Mathematics	59.54	23.07	91	61.31	24.86	273	-1.77	-0.07	.38
Grade 4: Mathematics	51.63	21.48	100	52.15	24.84	300	-0.53	-0.02	.77
Grade 5: Mathematics	66.46	21.78	92	58.12	24.98	276	8.33	0.34	<.001*
Grade 1: Reading	71.81	23.51	108	60.69	29.60	324	11.12	0.39	<.001*
Grade 2: Reading	70.13	21.00	102	59.88	30.58	306	10.25	0.36	<.001*
Grade 3: Reading	66.29	26.50	91	65.52	25.33	273	0.77	0.03	.73
Grade 4: Reading	64.98	22.64	100	58.87	25.70	300	6.11	0.24	.002*
Grade 5: Reading	69.58	18.46	92	64.38	23.13	276	5.20	0.24	.01*
Grade 1: Science	68.90	23.82	91	70.13	24.77	273	-1.23	-0.05	.58
Grade 2: Science	63.36	25.40	100	62.02	25.95	300	1.35	0.05	.52
Grade 3: Science	70.66	24.06	92	68.03	25.31	276	2.64	0.11	.23

***Note*. Percentiles are presented for the Measures of Academic Progress (MAP) mathematics, reading, and science assessments. The effect size is Hedges' *g*. The means for the treatment group were calculated by adding the means for the comparison group (i.e., the unadjusted means) and the differences (i.e., the treatment–comparison contrasts from the regression models). *p* values marked with an asterisk (*) were statistically significant after applying the Benjamini-Hochberg correction.**

Lied Quasi-Experimental Design

The individual grade-level results for the mathematics, reading, and science MAP assessments are presented in Exhibit 11 based on the Lied QED. There were three statistically significant positive impact estimates for mathematics, with effect sizes ranging from 0.09 to 0.34. These impact estimates showed that Lied students scored approximately 3 to 9 percentile points higher than the comparison students. There was one statistically significant negative impact estimate for reading and no statistically significant impact estimates for science. Additionally, the treatment students in Grade 6 at Lied attempted and completed just over one third more STEM credits than the comparison students, and the treatment students in Grades 7 and 8 attempted and completed approximately one more STEM credit than the comparison students, all of which were statistically significant differences.

Exhibit 11. Impact Estimates From the Lied Quasi-Experimental Design

Outcome/Grade level	Treatment students mean	Treatment students SD	Treatment students n	Comparison students mean	Comparison students SD	Comparison students n	Treatment–Comparison difference	Effect size	*p* value
Grade 6: Mathematics	53.46	27.35	532	49.33	26.63	1,596	4.12	0.15	<.001*
Grade 7: Mathematics	55.76	27.86	463	53.26	27.05	1,389	2.50	0.09	.002*
Grade 8: Mathematics	53.45	28.15	167	44.48	26.25	501	8.97	0.34	<.001*
Grade 6: Reading	57.80	25.65	532	58.25	26.67	1,596	-0.45	-0.02	.56
Grade 7: Reading	55.29	27.64	463	59.74	26.09	1,389	-4.45	-0.17	<.001*
Grade 8: Reading	56.17	26.52	167	53.86	26.45	501	2.31	0.09	.13
Grade 6: Science	59.99	27.80	532	61.38	27.75	1,596	-1.38	-0.05	.09
Grade 7: Science	62.71	28.29	463	61.51	28.27	1,389	1.20	0.04	.21
Grade 8: Science	54.49	30.10	167	53.90	29.71	501	0.60	0.02	.73
Grade 6: STEM credits attempted	3.01	0.07	531	2.66	0.37	531	0.35	1.32	<.001*
Grade 7: STEM credits attempted	3.10	0.30	503	2.12	0.33	503	0.99	3.13	<.001*
Grade 8: STEM credits attempted	3.21	0.41	206	2.16	0.37	206	1.05	2.70	<.001*
Grade 6: STEM credits earned	2.90	0.34	531	2.53	0.49	531	0.37	0.88	<.001*
Grade 7: STEM credits earned	3.00	0.40	503	1.97	0.48	503	1.03	2.35	<.001*
Grade 8: STEM credits earned	3.11	0.48	206	2.00	0.52	206	1.11	2.22	<.001*

***Note*. STEM = Science, Technology, Engineering, and Mathematics. Percentiles are presented for the Measures of Academic Progress (MAP) mathematics, reading, and science assessments. The effect size is Hedges' *g*. The means for the treatment group were calculated by adding the means for the comparison group (i.e., the unadjusted means) and the differences (i.e., the treatment–comparison contrasts from the regression models). *p* values marked with an asterisk (*) were statistically significant after applying the Benjamini-Hochberg correction.**

Lied Randomized Controlled Trial

The results for the mathematics, reading, and science MAP assessments are presented in Exhibit 12 based on the Lied RCT ITT analysis for students in Grade 6. Consistent with the QED results for Grade 6, there was a significant positive impact estimate for mathematics with an effect size of 0.16 and no statistically significant impact estimates for reading and science. The mathematics impact estimate indicated that the students who won the lottery scored 4 percentile points higher than the control students who did not win the lottery. Additionally, the treatment students completed and attempted approximately one third more STEM credits compared to students who did not win the lottery to attend Lied, which were statistically significant differences. The results were similar for the TOT analyses and showed statistically significant impacts in mathematics and STEM credits (see Exhibit 12). For mathematics and STEM credits, the TOT impact estimates were slightly larger than the ITT impact estimates (e.g., 5.00 percentile points vs. 4.07 percentile points for mathematics). Under the assumption that attending Lied had a positive impact on the outcomes, the pattern is logical given that the students who won the lottery but did not attend Lied were excluded from the treatment group in the TOT analyses.

Exhibit 12. Impact Estimates From the Lied Randomized Controlled Trial for Students in Grade 6

Outcome/Grade Level	Treatment students mean	Treatment students SD	Treatment students n	Control students mean	Control students SD	Control students n	Treatment–Control difference	Effect size	*p* value
ITT sample: Mathematics	54.51	26.47	382	50.44	25.85	404	4.07	0.16	<.001*
ITT sample: Reading	58.47	24.82	382	58.52	26.18	404	-0.06	0.00	.96
ITT sample: Science	60.74	26.11	382	60.86	27.67	404	-0.12	0.00	.92
ITT sample: STEM credits attempted	2.95	0.21	382	2.62	0.33	404	0.33	1.20	<.001*
ITT sample: STEM credits earned	2.86	0.33	382	2.52	0.45	404	0.34	0.85	<.001*
TOT sample: Mathematics	55.44	26.23	319	50.44	25.85	404	5.00	0.19	<.001*
TOT sample: Reading	58.03	25.07	319	58.52	26.18	404	-0.50	-0.02	.67
TOT sample: Science	60.24	25.92	319	60.86	27.67	404	-0.62	-0.02	.60
TOT sample: STEM credits attempted	3.01	0.03	319	2.62	0.33	404	0.39	1.59	<.001*
TOT sample: STEM credits earned	2.93	0.23	319	2.52	0.45	404	0.41	1.09	<.001*

***Note*. ITT = Intent-to-treat. TOT = Treatment-on-treated. STEM = Science, Technology, Engineering, and Mathematics. Percentiles are presented for the Measures of Academic Progress (MAP) mathematics, reading, and science assessments. The effect size is Hedges' *g*. The**

means for the treatment group were calculated by adding the means for the control group (i.e., the unadjusted means) and the differences (i.e., the treatment–control contrasts from the regression models). *p* values marked with an asterisk (*) were statistically significant.

O'Callaghan Quasi-Experimental Design

The individual grade-level results for the mathematics, reading, and science MAP assessments are presented in Exhibit 13 based on the O'Callaghan QED. There were three statistically significant positive impact estimates for the mathematics assessment, with effect sizes ranging from 0.18 to 0.44. These differences were equivalent to impact estimates favoring the treatment group by 5 to 12 percentile points. Additionally, there were two statistically significant positive impact estimates for the reading assessment, with effect sizes of 0.08 and 0.13. The reading impact estimates were smaller than the mathematics impact estimates and were equivalent to advantages of 2 to 3 percentile points for the treatment group. For the science assessment, there was one statistically significant positive impact estimate (i.e., an effect size of 0.09 for Grade 7) and one statistically significant negative impact estimate (i.e., an effect size of -0.14 for Grade 6). Additionally, the treatment students in Grades 7 and 8 at O'Callaghan attempted and completed over one more STEM credit than the matched comparison students, which were statistically significant differences. Treatment students in Grade 6 also statistically significantly attempted and completed more STEM credits than the matched comparison students, though the impact estimates were smaller than in Grades 7 and 8.

Exhibit 13. Impact Estimates From the O'Callaghan Quasi-Experimental Design

Outcome/Grade level	Treatment students mean	Treatment students SD	Treatment students n	Comparison students mean	Comparison students SD	Comparison students n	Treatment–Comparison difference	Effect size	*p* value
Grade 6: Mathematics	45.77	25.38	479	41.02	25.82	1,437	4.75	0.18	<.001*
Grade 7: Mathematics	53.62	28.30	398	41.84	26.57	1,194	11.79	0.44	<.001*
Grade 8: Mathematics	31.83	19.93	280	26.77	20.18	840	5.06	0.25	<.001*
Grade 6: Reading	50.21	26.63	479	49.75	27.49	1,437	0.45	0.02	.59
Grade 7: Reading	50.17	26.34	398	47.92	27.69	1,194	2.26	0.08	.03*
Grade 8: Reading	39.21	23.20	280	35.92	25.32	840	3.29	0.13	.005*
Grade 6: Science	47.72	27.90	479	51.81	28.85	1,437	-4.09	-0.14	<.001*
Grade 7: Science	52.85	27.86	398	50.12	29.43	1,194	2.73	0.09	.01*
Grade 8: Science	34.23	23.72	280	34.37	26.02	840	-0.14	-0.01	.91
Grade 6: STEM credits attempted	3.04	0.45	470	2.72	0.43	470	0.32	0.73	<.001*
Grade 7: STEM credits attempted	3.20	0.40	435	2.15	0.39	435	1.04	2.64	<.001*
Grade 8: STEM credits attempted	3.23	0.43	364	2.19	0.41	364	1.03	2.46	<.001*
Grade 6: STEM credits earned	2.99	0.51	470	2.47	0.64	470	0.52	0.89	<.001*

Outcome/Grade level	Treatment students mean	Treatment students SD	Treatment students n	Comparison students mean	Comparison students SD	Comparison students n	Treatment–Comparison difference	Effect size	*p* value
Grade 7: STEM credits earned	3.13	0.47	435	1.97	0.54	435	1.16	2.28	<.001*
Grade 8: STEM credits earned	3.16	0.49	364	1.98	0.58	364	1.17	2.19	<.001*

Note. **STEM = Science, Technology, Engineering, and Mathematics. Percentiles are presented for the Measures of Academic Progress (MAP) mathematics, reading, and science assessments. The effect size is Hedges' *g*. The means for the treatment group were calculated by adding the means for the comparison group (i.e., the unadjusted means) and the differences (i.e., the treatment–comparison contrasts from the regression models). *p* values marked with an asterisk (*) were statistically significant after applying the Benjamini-Hochberg correction.**

Descriptive Findings Related to the Reduction, Elimination, or Prevention of Minority Group Isolation

As shown in Exhibit 14, Gehring and Lied achieved their goals for 2019–20 related to preventing the creation of environments with minority group isolation by stabilizing the percentage of White students. Both schools had declining proportions of White students for at least 3 years before the implementation of the MSAP-funded magnet schools. However, the trend was reversed for both of the schools in 2018–19 and 2019–20. O'Callaghan, which had been serving an increasingly Latino student body over the 5 years prior to the start of MSAP implementation, continued the trend in 2018–19 and 2019–20 and did not reach its goal for reducing the percentage of Latino students.

Exhibit 14. Racial/Ethnic Trends Over Time for the Three Magnet Schools Assistance Program (MSAP) Grant-Funded Schools

School year	Percentage of White students at Gehring	Percentage of White students at Lied	Percentage of Latino students at O'Callaghan
2013–14	26.68	38.20	68.74
2014–15	29.59	35.33	66.15
2015–16	30.67	32.13	70.38
2016–17	27.44	29.63	70.23
2017–18	25.86	27.95	72.60
2018–19 (Year 1 of implementation)	26.76	32.92	73.37
2019–20 (Year 2 of implementation)	25.90	35.37	74.31

Discussion of Results

This section of the report summarizes the findings for the three research questions. It also discusses limitations of the study and outlines future research options that WestEd will explore as part of the final evidence of promise report.

Research Question 1

- What impact does attendance at an MSAP-funded magnet school have on students' performance on mathematics, reading/ELA, and science assessments?

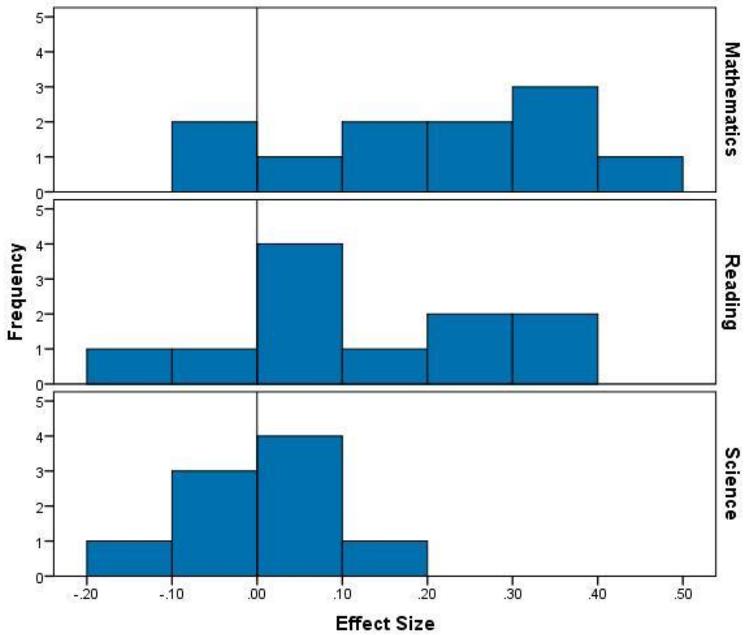
For each school, the analyses revealed positive, statistically significant impacts on mathematics and reading achievement across several grades. At Gehring, significant effects on mathematics ranged from 0.24 to 0.34 standard deviations, while significant effects on reading ranged from 0.24 to 0.39 standard deviations. The analysis did not show any effects, either positive or negative, on the science scores at Gehring.

At Lied, there were estimated positive effects from the QED on mathematics of 0.09 to 0.34 standard deviations across Grades 6, 7, and 8. There was one statistically significant negative effect for reading and no statistically significant effects on science achievement. The positive effect on mathematics in Grade 6 was consistent with the results from both the ITT and TOT analyses from the RCT that contrasted students who won the lottery with students who did not win the lottery. The RCT is a more rigorous design than the QED and provides stronger evidence regarding the impact of Lied on student achievement. Additionally, the finding that the RCT results at Lied were very consistent with the QED results for the same grade provides support for the validity of the QED findings from the other grades at Lied and the other schools.

At O'Callaghan, the estimated effects on mathematics achievement ranged from 0.18 to 0.44 standard deviations across Grades 6, 7, and 8. The effects on reading achievement in Grades 7 and 8 were 0.08 and 0.13 standard deviations, respectively, both of which were statistically significant. There was one statistically significant positive effect (i.e., an effect size of 0.09 in Grade 7) and one statistically significant negative effect (i.e., an effect size of -0.14 in Grade 6) on the science assessment.

The QED impacts of the MSAP-funded magnet schools on the students' performance in 2019–20 after 1.5 years of magnet implementation on the mathematics and reading assessments were generally positive (see Exhibit 15). Across the three schools, the average effect sizes were 0.20 for mathematics and 0.13 for reading. In contrast, the average effect size for the science assessment was 0.01, and the effect sizes were generally close to zero. The findings suggest that the MSAP-funded schools may be improving student performance on the mathematics and reading assessments but not on the science assessment.

Exhibit 15. Summary of the Impact Estimates on Student Performance From the Three Quasi-Experimental Designs



For the final evidence of promise report, WestEd will explore whether the effect sizes should be combined within or across schools using meta-analysis (Lipsey & Wilson, 2001) to determine whether the overall impacts are statistically significant. WestEd will also investigate the possibility of including student-level survey data (e.g., academic mindsets and behaviors from the Districtwide Climate survey) as covariates in the QEDs. Finally, WestEd will explore the implications for the conclusions that we can draw from the QEDs given the WWC's guidance that a single school with a unique organization and governance "is not a replicable intervention" (U.S. Department of Education, 2020, p. 86).

Research Question 2

- What impact does attendance at an MSAP-funded magnet school have on students' enrollment in and completion of STEM courses?

The QEDs showed that students at Lied and O'Callaghan attempted and earned more credits in STEM courses than did matched comparison students. For these analyses, one STEM credit is equivalent to one STEM course, such as a yearlong science course. Lied students in Grade 6 earned 0.37 more STEM credits than the matched comparison students, and Lied students in Grades 7 and 8 earned just over one more STEM credit than did matched comparison students. The results from the RCT analysis with students in Grade 6 at Lied replicated the findings from the QED. Similarly, O'Callaghan students in Grade 6 earned 0.52 more STEM credits than did matched comparison students, and students in Grades 7 and 8 earned 1.16 and 1.17 more STEM credits, respectively, than did the matched comparison students. These results show that attendance at the two magnet schools result in students earning more STEM credits than did comparison students who did not attend a magnet school. The difference for students in Grade 6 was lower than the differences in Grades 7 and 8 because one semester of computer literacy is required for students in Grade 6 in CCSD, and the STEM elective courses at Lied and O'Callaghan fulfilled this districtwide requirement.

Research Question 3

- Do MSAP-funded magnet schools reduce, eliminate, or prevent minority group isolation?

The school-level data indicated that Gehring and Lied reached their goals for 2019–20 related to preventing the creation of environments with minority group isolation by stabilizing the percentage of White students at the schools. In contrast, O'Callaghan did not achieve its goal for reducing the percentage of Latino students, the majority group at O'Callaghan, in 2019–20. A rigorous research design would be needed to determine whether the MSAP-funded magnet schools had a statistically significant impact on minority group isolation. For the final evidence of promise report, WestEd will explore how the research question related to minority group isolation could be addressed with student-level data so that the design could meet WWC Design Standards.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289−300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bifulco, R., Cobb, C. D., & Bell, C. (2009). Can interdistrict choice boost student achievement? The case of Connecticut's interdistrict magnet school program. *Educational Evaluation and Policy Analysis*, *31*(4), 323−345. https://doi.org/10.3102%2F0162373709340917

Bjorklund-Young, A., & Borkoski, C. (2016). *Do formative assessments influence student learning?: Research on i-Ready and MAP*. Johns Hopkins School of Education Institute for Education Policy. http://jhir.library.jhu.edu/handle/1774.2/63005

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, *84*(1), 151−161. https://doi.org/10.1162/003465302317331982

Dong, N., & Maynard, R. (2013). *PowerUp!*: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, *6*(1), 24−67. https://doi.org/10.1080/19345747.2012.673143

Leuven, E., & Sianesi, B. (2003). *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing* [Computer software]. Statistical Software Components S432001, Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s432001.html

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.

Nakamoto, J., Snipes, J., & Allender, S. (2020). *Clark County School District Magnet Schools Assistance Program (MSAP): Spring 2020 evaluation summary report*. WestEd.

Nakamoto, J., Snipes, J., Allender, S., & McMaken, J. (2021). *Clark County School District Magnet Schools Assistance Program (MSAP): Spring 2021 evaluation summary report*. WestEd.

Nevada Department of Education. (2019). *2019 Nevada School Performance Framework procedures manual: 2018–2019 school year*. http://nevadareportcard.nv.gov/DI/MoreDownload?filename=2019%20NSPF%20Procedures%20Manual.pdf

Resch, A., Berk, J., & Akers, L. (2014). *Recognizing and conducting opportunistic experiments in education: A guide for policymakers and researchers* (REL 2014–037). U.S. Department of

Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development. https://files.eric.ed.gov/fulltext/ED544801.pdf

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95*(450), 573−585. https://doi.org/10.1080/01621459.2000.10474233

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Smarter Balanced Assessment Consortium. (n.d.). *Smarter Balanced Assessment Consortium: 2017–18 summative technical report*. https://portal.smarterbalanced.org/library/en/2017-18-summative-assessment-technical-report.pdf

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1−21. https://doi.org/10.1214/09-STS313

Tuttle, C. C., Gleason, P., Knechtel, V., Nichols-Barrer, I., Booker, K., Chojnacki, G., Coen, T., & Goble, L. (2015). *Understanding the effect of KIPP as it scales: Volume I, impacts on achievement and other outcomes: Final report of KIPP's* Investing in Innovation *grant evaluation*. Mathematica Policy Research. https://files.eric.ed.gov/fulltext/ED560079.pdf

U.S. Department of Education. (2020). *What Works Clearinghouse: Standards handbook* (Version 4.1). https://ies.ed.gov/ncee/wwc/handbooks

Appendix

Survey Items Included in the Teacher Survey

Multiple Choice Answer Options:

Never, A Few Times per Year, Once per Month, Every Other Week, Once per Week, Daily

How frequently have you implemented the following project-based learning strategies this school year?

- Taught PLTW units
- Engaged in cross-curricular collaboration
- Integrated elective topics (for example, coding, robotics, and engineering) into core academic classes
- Assigned projects that connect to real-world issues
- Had students make some decisions about their projects
- Provided students the opportunity to reflect on their projects
- Provided students the opportunity to give, receive, and use feedback
- Had students display and present their projects
- Assigned projects that required students to think critically
- Assigned projects that required students to collaborate with one another

How frequently have you implemented the following blended learning strategies this school year?

- Chromebooks
- Google Classroom
- Canvas
- Utilized a mix of technology and face-to-face instruction
- Utilized digital and online content in a purposeful way

How frequently have you implemented the following personalized learning strategies this school year?

- Small group activities
- Flexible groupings

- Station rotations
- Had students engage in self-reflection
- Had students engage in goal setting
- Tailored learning plans to individual students
- Supported parent involvement in student learning
- Engaged in one-on-one interactions with students
- Allowed students to drive their learning paths

Survey Items Included in the Student Survey

Multiple Choice Answer Options:

Never, A Few Times per Year, Once per Month, Every Other Week, Once per Week, Daily

Gehring: About how often have you participated in each of these activities this year at your school?

- Robotics club
- Math club
- Tortoise habitat
- Gardening

Lied: About how often have you participated in each of these activities this year at your school?

- Robotics competition
- Computer programming competition
- Science Olympiad Team
- Girl Powered
- Girls Who Code

O'Callaghan: About how often have you participated in each of these activities this year at your school?

- Robotics competition
- Director's Cut
- Coding Club
- i3 Times