Final Impact Evaluation Report

Evaluation of ENpowered

U.S. Department of Education

Education Innovation and Research Program

August 2025

Submitted by:
The Policy & Research Group
www.policyandresearch.com

8434 Oak St.
New Orleans, LA 70118

107 Spring St.
Seattle, WA 98104

Katherine Lass, MPH, MSW; Sarah Walsh, PhD; Kelly Burgess, MPH; Eric Jenner, PhD. 2025. Evaluation of ENpowered: Final Evaluation Report. New Orleans: The Policy & Research Group (PRG).

ACKNOWLEDGMENTS

This study would not have been possible without the support of Kelli Easterly at Chicago Public Schools (CPS) as well as the *CPS External Research Office*, who we are grateful to have had the opportunity to work with. In addition, we are thankful for Project SYNCERE who developed the intervention and who worked diligently throughout the project to provide sites and staff with training and technical assistance in program implementation. We also thank Anne Wolf, PhD, who provided valuable guidance and technical assistance throughout the span of our evaluation. The following team members at PRG contributed valuable support throughout this five-year project: Alexis Mayfield, Rebekah Snider, Hilary Demby, and Gretchen Falk. Finally, we express deep appreciation to the study site partners, staff, and youth who participated in this study.

DISCLOSURE

No potential competing interest to report.

This publication was prepared under Grant Number U411C190108 awarded to Project SYNCERE, from the *Office of Elementary and Secondary Education* within the *U.S. Department of Education*. The views expressed in this report are those of the authors and do not necessarily represent the policies of the funders.

EXECUTIVE SUMMARY

Project SYNCERE (PS) received a *U.S. Department of Education*, Education Innovation and Research (EIR) Program early-phase grant in 2019 to implement and evaluate ENpowered, a project-based learning engineering curriculum for middle school students. The ENpowered program aims to create hands-on learning opportunities for underrepresented youth in Chicago Public Schools (CPS) to engage in science, technology, engineering, and math (STEM) programming with a goal to foster an enduring interest to pursue additional STEM education and careers. As part of the five-year grant, PS contracted with The Policy & Research Group (PRG) to evaluate the implementation and impact of the ENpowered program on middle school student outcomes. The purpose of this report is to present summative implementation and impact evaluation findings from the project.

The 10-week program engages students through a series of instructional lessons and small-group projects that focus on solving real-world challenges using engineering design principles. Through weekly lesson plans delivered in the classroom, the program introduces students to new STEM skills in computer science and engineering. Additionally, the program connects students with STEM professionals who attend classes to speak about their own experiences in their careers. This opportunity allows students to meet individuals from similar backgrounds working in different STEM fields as a means of developing an interest and sense of attainability to pursue similar paths. The program culminates with a cross-school competition called ENpowered Games where students convene to present their projects and showcase what they have learned throughout the program. PS designed the ENpowered program to increase interest and engagement in STEM topics for underrepresented youth in hopes of increasing achievement in math and science courses.

PRG conducted a rigorous impact and implementation evaluation of the ENpowered program's effect on middle school student outcomes. The impact study utilized a quasi-experimental design to examine the effects of one semester of exposure to ENpowered on two confirmatory student outcomes: end-of-course grades in math and science. We estimate program impact with a propensity score-weighted multilevel difference-in-difference regression model that contrasts the change in course grades for students in the ENpowered program with similar students that did not participate in the program. The treatment participants consisted of sixth-, seventh-, and eighth-grade students in CPS that participated in one semester of ENpowered. The selection for the comparison group depended on each school's implementation of ENpowered and consisted of either a within-school comparison group of students enrolled in the same school and grade as treatment participants, or an external-school comparison group comprised of students in the same grades as treatment participants. The implementation study (described in Appendix B) explores the extent to which the ENpowered program was implemented as intended at each study site during the three implementation school years (2021–22, 2022–23, and 2023–24).

IMPLEMENTATION OF ENPOWERED

There was some variation in the implementation of ENpowered across schools and cohorts, but overall, the grantee implemented the classroom programming to fidelity during the implementation period. The grantee implemented ENpowered in 19 middle schools in CPS during the spring 2022, 2023, and 2024 semesters. Motivating real-world problems that drove student projects focused on reducing residential energy consumption, creating more efficient e-commerce logistics systems, and improving autonomous vehicle safety.

Teachers and students provided both quantitative and qualitative feedback on their experiences participating in ENpowered at the end of each implementation period. Overall, teachers expressed satisfaction with classroom and ENpowered Games programming, materials, and student experiences. Teachers shared that they believed ENpowered had a positive impact on their students and their level of confidence in pursuing STEM activities in and outside the classroom. Additional feedback suggested that the program provides valuable exposure to the engineering field, particularly computing and robotics, and that students developed a strong connection with their co-instructors and the STEM professionals that visited their classrooms. Most students (77%) reported that they would participate in the program again and would recommend it to others (78%). Students' qualitative feedback suggested that their favorite parts about ENpowered were the competitive aspect of the ENpowered Games event, learning and being able to demonstrate new STEM skills, collaborating with their teams, and meeting STEM professionals.

Impact Study Findings

Benchmark statistical estimates for both research questions indicate that participating in the ENpowered program had no statistically significant effect on middle school students' end-of-course grades in mathematics (Research Question 1) or science (Research Question 2). Model estimates find that students' end-of-course grades were similar across both treatment and comparison students for both subjects. We did not observe any differences in program effects for subgroups broken down by grade level and gender.

Conclusion

Benchmark findings indicate that offering the ENpowered program to middle school students did not have a detectable impact on the confirmatory outcomes of students' final grades in math or science. Qualitative feedback from both teachers and students suggests that ENpowered appealed to both groups and has the potential to broaden students' perspectives on career pathways. The aim of this study was to produce empirical, causal responses to the posed research questions, and is just one part of the comprehensive evaluation PRG conducted on ENpowered. Future work should continue to examine how programs like ENpowered can have long-term effects on student motivation and interest in pursuing STEM-focused activities and careers beyond middle school.

TABLE OF CONTENTS

INTRODUCTION

Job growth in the fields of science, technology, engineering, and math (STEM) is expected to outpace that of non-STEM jobs in the near future (Pew Research Center, April 2021). Despite longstanding efforts to increase diversity in STEM, Black, Hispanic/Latino/a, and female workers remain underrepresented in the STEM workforce (National Center for Science and Engineering Statistics [NCSES], 2023). Gaps are particularly large in the fields of computing and engineering. For example, Black and Hispanic/Latino/a workers, who represent 11 and 17% of the workforce overall, make up between 5 and 9% of the workforce in computing and engineering (NCSES, 2023; Pew Research Center, April 2021). Additionally, women, who comprise roughly half of the STEM workforce overall, make up only 15 and 25% of computing and engineering workforces, respectively (Pew Research Center, April 2021).

Project SYNCERE (PS) is a Chicago-based nonprofit whose mission is to prepare the minds of underrepresented students and create pathways to pursue careers in STEM. Launched in 2009, PS has developed a series of in-school and after-school programming that aim to ensure youth from underserved communities have access to opportunities that will inspire and prepare them for future careers in STEM. The ENpowered program is a middle school, project-based learning engineering curriculum that provides students with an opportunity to address relevant, real-world problems using engineering design principles. The ENpowered program bridges informal and formal learning into the school day through its weekly lesson plans and connects underrepresented youth with diverse STEM professionals to help illuminate pathways to the STEM workforce. The program culminates with a cross-school competition, ENpowered Games, where students present their projects and showcase what they have learned throughout the 10-week course. The program is based on research supporting the value of project-based learning, mentor engagement, competition to increase student engagement, and curriculum taught by educators trained in content knowledge.

Research has shown that early exposure to STEM subjects, including engineering, by trained teachers or educators with expertise in the particular field being taught can promote interest and motivation to pursue subsequent learning and degrees (Sithole et al., 2017; Tai et al., 2006). Moreover, a strong sense of self-identity in STEM is an important building block to growing participation in related fields, particularly for Black, Hispanic/Latino/a, and female students (Carlone & Johnson, 2007; Schembri & Rampersad-Ammons, 2021). A strong sense of science identity is demonstrated by someone who has scientific knowledge, motivation to understand processes in a scientific way, has the skills to perform scientific practices, and recognizes themselves as a “science person” (Carlone & Johnson, 2007). This identity can be either strengthened or hampered depending on the degree to which students receive recognition from peers, family, and respected professionals from similar backgrounds. Evidence suggests that exposure to engineering education and principles in middle school can support acquisition of widely applicable knowledge and skills needed to enter the STEM workforce, such as creativity, problem solving, teamwork, and critical thinking (Brophy et al., 2008; Samuels & Seymour, 2015).

The purpose of this report is to present summative findings from a five-year project that implemented and evaluated the ENpowered program. Through its Education Innovation and Research (EIR) Program, the *U.S. Department of Education* (ED) provided competitive grants to expand the implementation of, and investment in, innovative practices that are demonstrated to have an impact on improving student achievement or student growth, closing achievement gaps, decreasing dropout rates, increasing high school graduation rates, or increasing college enrollment and completion rates.[1] The program provides funding to applicants who propose to create, develop, implement, replicate, and/or take to scale evidence-based innovations and rigorously evaluate those innovations.

Funded through a 2019 EIR early-phase grant, this project was a collaborative effort between PS, the grantee and program developer, and The Policy & Research Group (PRG), the independent evaluator. PRG conducted a multisite, multiyear quasi-experimental impact study (QED) designed to assess ENpowered's impact on student achievement in math and science. PRG also conducted a concurrent implementation study aimed at understanding the extent to which ENpowered was conducted with fidelity each year. The impact study consists of 475 students who participated in ENpowered during the spring 2022, 2023, and 2024 semesters across 8 middle schools in Chicago Public Schools (CPS).

This report describes impact and implementation findings from the three-year evaluation of the ENpowered program. We first present an overview of the ENpowered program model, followed by an overview of the impact study design, including assignment procedures, outcome measures, and data collection and analytic methods. We then describe the analytic samples, including a discussion of baseline equivalence testing, and present findings and discussion from the implementation evaluation and impact analyses. Supplemental details are provided in a series of appendices that present a graphical representation of the ENpowered logic model (Appendix A), fidelity of implementation overview and findings (Appendix B), detailed variable operationalization and analytic methods (Appendix C), detailed analytic results (Appendix D), and comparative results from a series of sensitivity analyses (Appendix E).

ENPOWERED MODEL

The ENpowered program is designed to increase student interest and engagement in STEM topics by exposing them to engineering education in an engaging and interesting way that provides hands-on experience and opportunity for mastery. Given that research suggests that early exposure to STEM topics can lead to increased motivation to spend more time devoted to those subjects, PS hypothesizes that students who participate and succeed in the ENpowered program will become more interested in STEM and gain the confidence needed to pursue difficult challenges. ENpowered gives students the opportunity to apply computer science and engineering principles using technology and math concepts to solve real-world problems, which PS hypothesizes will yield comparatively higher achievement in mathematics and science.

PS has developed frameworks to implement ENpowered as either an in- or after-school program with students across Grades 3 through 12. This EIR grant funded the implementation and evaluation of the ENpowered Games for Middle School Model, as offered to middle school students (Grades 6 through 8) during the regular school day, which includes a structured curriculum of instructional lessons and the culminating competition experience, ENpowered Games.[2]

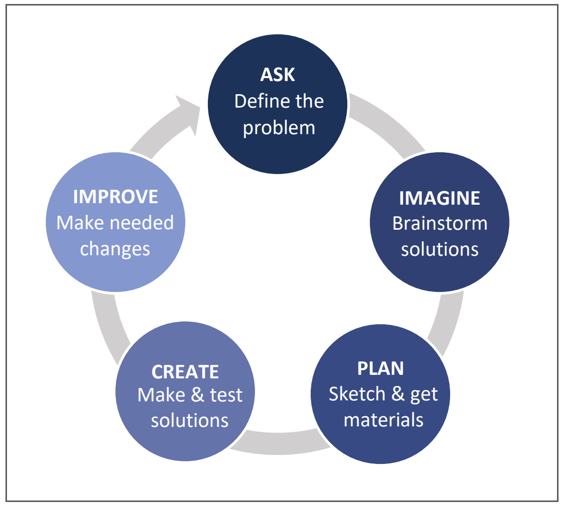
[1] For more information on the EIR program, see https://www.ed.gov/grants-and-programs/grants-special-populations/economically-disadvantaged-students/education-innovation-and-research

[2] For more information about Project SYNCERE and the ENpowered program, see https://projectsyncere.org/our-programs

The program was implemented in CPS middle schools each spring academic semester (January through May) during the 2021–22, 2022–23, and 2023–24 school years.[3] At each school students in the program met for a total of 20 hours over the course of 10 weeks (2 hours each week) to complete program activities organized within the structured curriculum. ENpowered lessons took place in science and STEM elective periods (e.g., STEM lab).

The classroom curriculum lessons are taught by two PS co-instructors who lead students through the Engineering Design Process (EDP; see Box 1), core engineering principles, and hands-on, problem-solving activities. All lessons incorporate problem-based learning and focus on solving a real-world technical challenge. Through this model, students are first presented with a real-world challenge, such as the need to automate residential energy usage or reduce accidents in autonomous vehicles, and then learn the applicable knowledge and skills necessary to respond to and solve the problem. During project-based learning, students use the EDP framework to actively question, create and research ideas, and develop and test solutions. Students work to improve their initial solution and prepare to present their final project to their peers and a network of professionals at the ENpowered Games event.

Box 1. The Engineering Design Process



During the 10-week class, students gain skills such as problem solving, project management, collaboration, and leadership to complete their projects. The program is designed to spark students' curiosity in STEM topics and demonstrate how classroom activities relate to real-life problems that can be solved through technical thinking, technology, and computer science skills. By solving real-world problems themselves, students gain confidence in their abilities.

At least once during the semester, professionals from the Chicago STEM business community accompany the PS co-instructors to classrooms where they talk to the students about their careers, including the pathway they took from middle school onward. The purpose of having STEM professionals from the community discuss their careers is to increase exposure to and familiarity with different types of careers in STEM available in their community. Additionally, recruiting STEM professionals who have similar backgrounds to the students in the program allows students to visualize themselves becoming successful in a technical field they might not have previously considered, increasing their motivation to engage with topics.

At the end of the semester, students from each school who participate throughout the 10-week program come together to compete in the ENpowered Games competition. The ENpowered Games event offers participants the opportunity to showcase what they learned throughout the semester by

[3] PS implemented a virtual pilot implementation during the 2020–21 school year. Schools in CPS were fully remote during this academic year due to the COVID-19 pandemic. This pilot year was an intentional deviation from the traditional program model to reach students before the start of the impact study and finalize lesson plans. As the program is designed to be implemented in person to promote the building of connections with students, STEM professionals, and instructors, this pilot implementation year is not considered part of the summative impact evaluation.

completing a series of timed engineering design challenges. Leading up to the ENpowered Games, students work in small groups to complete their design solution to the motivating problem and complete an *Engineering Notebook*, which lays out a design plan for the final design challenge that will be executed at the event. At the ENpowered Games, the student groups present the content of their *Engineering Notebook* to a panel of judges (volunteer STEM professionals from the community). Judges give each group two scores: one for the quality of the presentation and one for the content of the *Engineering Notebook*. After presenting their notebook, students then complete a series of timed design challenges where they are presented with a prompt related to but modified from the project they just presented to the judges to test their knowledge and application of the skills learned in the classroom. The purpose of the ENpowered Games competition is to give students the opportunity to execute their design plan in a competitive, fun environment as well as engage with other students across the city of Chicago who are excelling in STEM topics. The competitive aspect of demonstrating their learned skills is believed to increase motivation and engagement in the project. Giving students the opportunity to see other students from their community and from similar backgrounds be curious and demonstrate mastery of STEM skills aims to help students see themselves as having the potential to be successful in the field.

IMPACT STUDY OVERVIEW

To assess whether the program had an effect on math and science achievement, PRG conducted a QED that employed propensity score weighting and a difference-in-differences (DID) approach to assess the impact of one semester of participation in ENpowered on two confirmatory outcomes: student academic achievement in math (defined as final course grade in math) and student academic achievement in science (defined as final course grade in science). Intervention or “treated” participants consist of a weighted sample of sixth- through eighth-grade students who were enrolled in the ENpowered program at study implementation schools; comparison participants are a weighted sample of students at either study implementation schools (if available) or other, similar schools within CPS who do not implement ENpowered. For each study implementation school, we selected a pool of comparison students in one of two ways, depending on how ENpowered was implemented at that school: (1) a *within-school* comparison group comprised of students at the same school in the same grade(s) as treatment participants; or (2) if there were no additional students in the same grades available at the implementation school, an *external-school* comparison group comprised of students in the same grade(s) as treatment participants at comparable schools in CPS.

PS recruited schools from CPS to participate in the grant-funded project each fall prior to implementation. Primary coordinators at each implementation school decided which grade level and, if applicable, the classroom that would receive the ENpowered program. The program is designed to be accessible by students in sixth through eighth grade and is therefore flexible to be implemented across grade bands.

We estimate the treatment effect by way of a propensity score-weighted multilevel DID regression model that contrasts the change in course grades for students in the treatment group with those of the comparison group. We compare change in science and math outcomes from baseline to post program for students who participated in the ENpowered program (treatment) to an equivalent group of students who did not (comparison).[4] Baseline measures of the outcome are the Quarter 1 report card grades, recorded in November of each school year. Post-program outcomes are the final, end-of-course

[4] Comparison students did not receive the ENpowered program during the academic year under examination nor during a previous year.

grades, recorded in June. We use inverse probability of treatment weights calculated from propensity scores to improve the equivalence of the two groups on an array of observable baseline features.[5]

Because students from within the same school are most likely to be similar to one another in terms of important and unobservable baseline characteristics, we prioritize and use within-school comparisons when possible (Cook et al., 2008). PRG used academic records obtained from CPS to create the final analytic samples of students (propensity score weighting) and assess impact on the confirmatory outcomes (final grades in math and science courses). Propensity score weighting was conducted with baseline data before any outcome (post-program) data were merged into the dataset.

Research Questions

The impact evaluation answers two confirmatory research questions concerned with ENpowered's effect on outcomes identified by the program's theory of change and logic model, presented in Appendix A. As listed in this section, the research questions for this study focused on students' performance in their math and science courses. Operational definitions for these outcomes, all inclusion criteria, and the analytic framework and procedures to estimate the efficacy of ENpowered were prespecified by the evaluation team prior to the collection of any outcome data.[6]

The research questions are as follows:

- **Research Question 1:** As compared to middle school students who are similar but not exposed to the program (comparison), what is the impact of participating in ENpowered (treatment) on math achievement at the end of the academic school year (as measured by final course grade in math)?
- **Research Question 2:** As compared to middle school students who are similar but not exposed to the program (comparison), what is the impact of participating in ENpowered (treatment) on science achievement at the end of the academic school year (as measured by final course grade in science)?

Comparison Experience – Class as Usual

The comparison sample is comprised of students in CPS in Grades 6 through 8 who did not participate in ENpowered during the observation period. These students either attended a school that implemented ENpowered or a school that did not implement ENpowered.[7] Instead, comparison students received the standard STEM classroom instruction typically offered by the school to meet CPS and Common Core standards for their grade. This is considered a "class-as-usual" comparison condition. This study contrasts outcomes achieved by students who participate in the treatment program with outcomes for students who participate in what they would have otherwise received. Assuming internal validity of our study, the estimated difference between the two groups is an externally valid estimate of what the real-world impact of the program would be for the sample under study. When treatment students were attending and participating in the ENpowered Games, comparison students were assumed to be in their normal classes, unless otherwise absent from school. This class-as-usual comparison condition naturally varies across schools, depending on the differences in schedule and curriculum at each school.

---

[5] Propensity scores were estimated with covariates deemed predictive of selection and the outcome, including the baseline measure of the outcome (Quarter 1 grades) and demographic characteristics reported by CPS.

[6] The *Evaluation Plan* was registered on the Registry of Efficacy and Effectiveness Studies (REES; Registry ID 7341) in 2022.

[7] While comparison students did not participate in any ENpowered program activities during the given year of implementation or any year prior, they may have been offered the program during a subsequent year and may have been a part of multiple cohorts of the study, acting as a comparison student first and then a treatment student later.

Therefore, there is not a single definition of what the comparison experience was for those students beyond standard sixth- through eighth-grade instruction.

IMPACT STUDY DESIGN

This impact study investigates the effect of ENpowered on participating students' course performance in middle school math and science. We do this by employing a propensity score-weighted multilevel DID model that assesses growth in course grades from the first quarter report card (baseline) to the end-of-course final grade (post program) for students in the treatment and comparison groups. Outcome data were collected from the *CPS External Research Office*, *Department of School Quality Measurement and Research* (SQMR) who provided individual-level student administrative records. Data collection procedures were the same for students enrolled in both the treatment and comparison conditions.

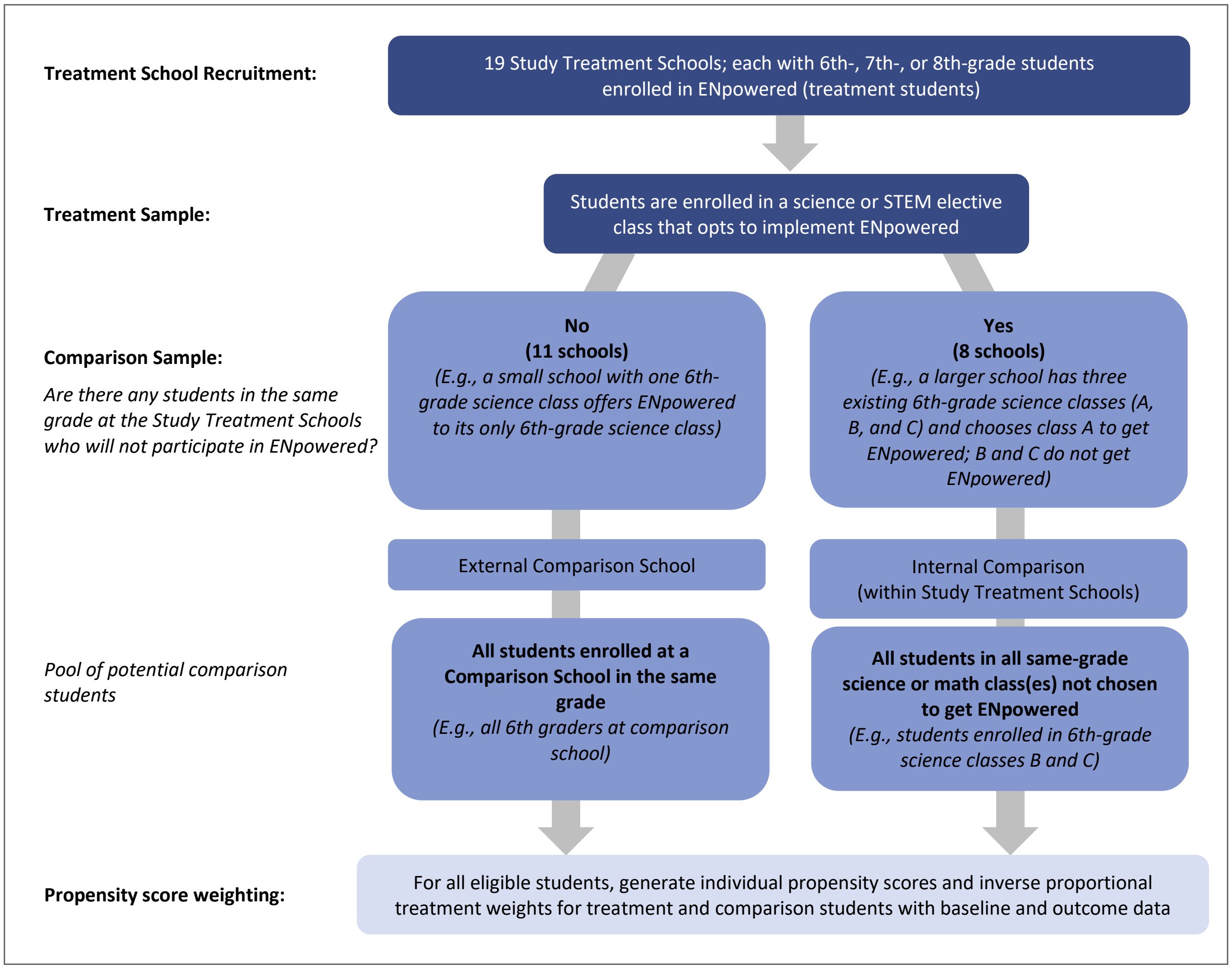
For each student, we use a single pre- and single post-intervention observation and measure the impact of the intervention as the average "difference in differences" between the treatment and comparison groups. Two separate analytic models were run for each outcome – one for the within-school assignment of participants (where students were assigned to treatment or comparison via their classroom enrollment), and one for the external assignment of participants (where students were assigned via their enrollment at an implementation or external comparison school). Impact estimates were then averaged together with a weighted treatment effect and its standard error.[8] We provide an overview of assignment procedures, outcome measures, and analytic methods in the subsequent sections. Additional technical details, including operationalization of outcome measures and analytic methods are provided in Appendix C.

ASSIGNMENT PROCEDURES

The comparison student selection process varied depending on the availability of students in the same grade as ENpowered participants at study implementation schools. As a result, we have two distinct (nonoverlapping) analytic samples in this study: (1) one where students were assigned at the *classroom level* (within-school contrast); and (2) one where students were assigned at the *school level* (external contrast). A visual representation of the comparison group selection process that was conducted for each implementation site is shown in Box 2.

[8] We calculate the precision weighted average of the program's effect and the standard error of that estimate using the formula for independent, nonoverlapping subsamples developed by Price and Wolf (2024) and described in Appendix C.

Box 2. Analytic Sample Prescreening and Selection



As shown in Box 2, the process by which students are selected in the study sample begins with the selection of study treatment schools and the treatment classrooms. When nonprogram students in the same grade as program students were available, we selected comparison students from the study implementation site (within-school comparison). When unavailable, we identified, for each implementation school, a comparison school and selected students from this alternate site who were in the same grade as the students in ENpowered. This process repeated for each study implementation site during each of the three study years. Regardless of whether students are selected into the treatment or comparison pool at the classroom- or school level, final sample selection was done at the individual level by estimating individual propensity scores and subsequent inverse proportional treatment weights (IPTWs) as represented in the bottom row in Box 2.

Treatment Students

PS was responsible for recruiting, selecting, and retaining the schools that would offer ENpowered during a given school year. Grantee staff worked with the school to figure out how the program would be implemented at the school including what grade(s) to offer the program to, in what science or STEM elective class to implement the lessons, days/times of instruction, and coordination of transportation to the ENpowered Games event.

To be included in the evaluation, schools had to confirm that they would be offering the program to students who had not participated in ENpowered during a previous academic year, and that they would recruit students for participation without regard to underlying interest in STEM topics – in other words, students were not handpicked for participation based on their STEM interests. The purpose of this requirement was to minimize the potential threat of confounding factors that could not be assessed in the comparison group using administrative data.

Students were considered part of the treatment condition if they were enrolled in the class that was selected by school coordinators to receive the ENpowered program. Class rosters were finalized at the beginning of the school year, prior to designation to receive the program. Treatment status was confirmed using program attendance rosters collected during the 10-week class. As this is a QED, we exclude a student who was listed on the classroom roster but who did not attend the ENpowered program classes from the treatment sample.

COMPARISON STUDENTS

CLASSROOM-LEVEL ASSIGNMENT

As shown in Box 2, out of the 19 implementation blocks (6 to 7 schools across three cohorts), 8 had at least 2 classes of students in the same grade, 1 of which participated in ENpowered during a given school year and 1 that did not. For these schools, we selected the nonprogram classroom as the pool of comparison students for the respective treatment class. These 16 classes of 361 students comprise the classroom-level sample (176 treatment and 185 comparison).

SCHOOL-LEVEL ASSIGNMENT

Box 2 shows that out of the 19 implementation blocks, 11 did not have a nonprogram class of students in the same grade, and therefore we identified a similar comparison school in CPS to construct the pool of comparison students at the school level. For any implementation school that did not have a comparison classroom available, PRG selected an alternate comparison school by identifying a school within CPS that was comparable to the study implementation site on relevant pre-intervention school-level characteristics.[9] Research indicates that observational methods are more successful at reducing the effects of selection bias when the comparison sites are as similar as possible (Cook et al., 2008).

When the study treatment school did not have a comparison pool of students, we examined school-level criteria for all CPS middle schools (i.e., schools that serve Grades 6 through 8) to identify a potential matched comparison school for a given study implementation school.[10] We used the following criteria to conduct a two-stage "matching" process where we began with the full list of all CPS-run elementary/middle schools and first narrowed the pool of potential comparison schools using four characteristics that we believed to be the most important (denoted with an asterisk):

- Overall Illinois Assessment of Readiness (IAR) proficiency in mathematics (percentage who met or exceeded grade-level expectations)*
- CPS administrative network[11]*

[9] This is consistent with the guidance offered by Song and Herman (2010).

[10] School-level characteristics were obtained from the CPS School Data Tools: https://cps.edu/SchoolData/Pages/SchoolData.aspx

[11] District-run CPS schools are organized into 17 networks, which provide administrative support, strategic direction, and leadership development to the schools within each network.

- Racial/ethnic composition of student body*
- Percentage of students who qualify for free and reduced price meals at school
- Percentage of students who are English Language Learners (ELLs)
- Percentage of students who are enrolled in an Individualized Education Plan (IEP)
- Number of students who attend the school
- Academic program group (e.g., general education, STEM, scholastic academy)*

The first stage aimed to narrow the selection of potential comparison schools by identifying any CPS school that met the following criteria:

- Average IAR proficiency in mathematics: A school's percentage of students who met or exceeded grade-level expectations must be within 10 percentile points (+/–) of the study treatment school percentile.
- CPS administrative network: A school must be located within the study treatment school's network or a geographically adjacent network.
- Racial/ethnic student body composition: The percentage of students who are Black and/or Hispanic/Latino/a must be within 10 points (+/–) of the study treatment school's composition.
- Academic program group: A school must be categorized as a similar academic program group as the study treatment school.

If this first round of selection criteria did not yield any potential comparison schools (i.e., we did not achieve a match on the four priority characteristics), we then widened the margin for matching on the IAR attainment scores and racial composition such that the score had to be within a 15-point (+/–) difference from the study treatment school and omitted the criteria related to administrative network.

Once we obtained at least one, but ideally multiple, potential matches, we reviewed the second set of criteria to make the final comparison school selection. In the second stage of the school selection process, we examined additional characteristics to further refine the rank order of potential comparison schools. These additional characteristics are: percentage of students who qualify for free and reduced price meals at school, percentage of students who are bilingual or ELLs, percentage of students who are enrolled in an IEP, number of students who attend the school, and average attendance rate. Once an external comparison school was selected, all students from the specified grade level (the grade level that ENpowered students were enrolled in) were selected into the comparison pool for weighting procedures.

FINAL SAMPLE FORMATION

Because this study is observational (i.e., not a randomized controlled trial), we employ a quasi-experimental method to mitigate the confounding effects of selection that have the potential to bias estimates of program impact. Specifically, we conducted a propensity score weighting procedure, where weights assigned to participants serve as a way to proportionately select participants into the treatment or comparison groups for analyses.[12] Conceptually, we generate an empirical score that quantifies the conditional probability that an individual would select into the treatment group – or alternatively the comparison group – this is the propensity score. We then use this score to upweight those cases that are

[12] The second method used to mitigate confounding effects of selection is the use of a DID analytic model, described in more detail in the Analytic Methods section. The DID analytic approach aims to reduce the influence of unobserved and confounding features on our impact estimates. The use of these two methods in tandem – propensity score weighting and DID – is a "doubly robust" method because we use both IPT weights in the impact regression equation and a DID design to mitigate selection bias. The first reduces bias with observed covariates – by balancing on them; the second uses fixed effects to estimate an unbiased estimate of treatment effect when important explanatory variables are unobserved.

more alike (i.e., are in the middle of the likelihood distribution) and downweight those that are less alike (i.e., are on the extremes of the likelihood distribution). The weighting procedure effectively serves to select proportions of eligible students into the final treatment or comparison sample for analyses.

Propensity score weighting aims to make treatment assignment ignorable, conditional on the set of observable baseline characteristics (Guo & Fraser, 2010). Individual-level propensity scores were estimated separately for the students in the classroom-level and school-level assignment pathways using a set of baseline covariates. We generated propensity scores with a logistic regression model that calculates the likelihood of treatment assignment using the following pre-intervention student-level variables, obtained from CPS:

- Cohort: A series of dummy variables indicating whether or not the student was enrolled in the study during a given school year
- Two variables of baseline achievement were used:
    - First quarter report card grade for math at the beginning of the implementation year
    - First quarter report card grade for science at the beginning of the implementation year
- Race/ethnicity: A series of dummy variables indicating student as Hispanic/Latino/a, Black, White, Multiracial, or Other[13]
- Gender: A dummy variable indicating whether the student is identified as female or not
- Age in years at study entry
- Socioeconomic status: A dummy variable indicating whether or not a student was eligible for free and/or reduced-price lunch
- Academic disadvantage status: A dummy variable indicating whether or not a student was designated as receiving special education/IEP and/or has an ELL designation
- Home language: A dummy variable indicating whether or not CPS reported the student's home language was something other than English
- Grade level: A series of dummy variables indicating whether or not the student was enrolled in sixth, seventh, or eighth grade during a given school year
- School site: A series of dummy variables indicating whether or not the student was enrolled in a given school during a given school year[14]

After generating the individual-level propensity scores, we then "trimmed" the scores such that any observations that fell outside the region of common support, we re-coded propensity scores that transcended this region at the threshold value (Austin & Stuart, 2015; Imbens & Rubin, 2015).[15]

In the final step, we calculated weights for each student according to their propensity score so that the regression analysis was conducted on the full sample of participants with baseline and outcome data (as opposed to conducting individual matching). Cases were upweighted if they are more alike (according to observed covariates) and downweighted if they are less alike. We use IPTW to balance the treatment

[13] The final logistic regression model only included two of these dummy variables representing Black and Hispanic/Latino/a. For the most part, study schools had a highly homogenous population of students, with few identifying as White, Multiracial, or some other race/ethnicity. This is a reflection of the target population of the program and evaluation. Due to rare occurrences, we omitted the dummy variables for White, Multiracial, and Other race/ethnicity from the logistic regression model.

[14] This set of dummy variables was only included for the classroom-assignment sample, given that school is synonymous with treatment assignment in the school-assignment sample.

[15] Specifically, we first determined the maximum propensity score for the comparison group and the minimum score for the treatment group. For any participants in the treatment group whose propensity score was more than this (comparison) maximum value, we re-coded their propensity score to the threshold value. Similarly, any individuals in the comparison group whose propensity score is less than the (treatment) minimum was re-coded to the minimum threshold value.

and comparison groups and estimate the average treatment effect (ATE). The formula for the weighting procedure is:

$$w(D,x) = \frac{D}{e^{\wedge}(x)} + \frac{1-D}{1-e^{\wedge}(x)}$$

where w equals the ATE weight, conditional on treatment status *D* and conditioning set *x*, and $e^{\wedge}(x)$ equals the estimated propensity score.

OUTCOME MEASURES

Confirmatory outcomes are operationalized as follows: (1) end-of-course grade in math – a count variable that represents the final letter grade in their standard math course; and (2) end-of-course grade in science – a count variable that represents the final letter grade in their standard science course. Outcomes are measured and analyzed at the individual student level. Outcome data were collected uniformly from CPS for all study participants.

MATH ACHIEVEMENT

Achievement in math is defined as the final (end-of-course) letter grade recorded for the CPS core mathematics course for a given grade level. Letter grades were provided by CPS and were re-coded to a numeric score where F = 0, D = 1, C = 2, B = 3, A = 4. Final course grades were recorded in June of each school year.

The baseline measure of the outcome is defined as the first quarter report card grade for the same mathematics course. First quarter grades are again reported by CPS as letter grades and were re-coded to a numeric score using the same scale noted above. First quarter grades were recorded in November of each school year, prior to the beginning of programming, which began in January of each year.

SCIENCE ACHIEVEMENT

Achievement in science is defined as the final (end-of-course) letter grade recorded for the CPS core science course for a given grade level. Letter grades were provided by CPS and were re-coded to a numeric score where F = 0, D = 1, C = 2, B = 3, A = 4. Final course grades were recorded in June of each school year.

The baseline measure of the outcome is defined as the first quarter report card grade for the same science course. First quarter grades are again reported by CPS as letter grades and were re-coded to a numeric score using the same scale noted above. First quarter grades were recorded in November of each school year, prior to the beginning of programming, which began in January of each year.

COVARIATES

In addition to course grades, we requested the following student-level covariates from CPS that were included in our analytic models: date of birth, race/ethnicity, gender, ELL status, IEP status, free and/or reduced-price lunch status, primary language used at home, grade level, and school of enrollment. We also included a series of dummy variables indicating study cohort of enrollment in our models. Additional details of covariate operationalization can be found in Table C.1 in Appendix C.

DATA SOURCES AND COLLECTION

All baseline and outcome data were collected from CPS' *External Research Office*, *SQMR*. PRG entered into a formal data sharing agreement with CPS in 2021, which directed all data sharing procedures. PRG requested student-level data according to the procedures outlined in the data sharing agreement annually, at the conclusion of each school year. CPS sent data to PRG using a secure file transfer protocol and data were de-identified but included a student-specific identifier so that student records could be linked across data files and school years. Data collection procedures were identical for both the treatment and comparison samples.

ANALYTIC METHODS

As detailed in our research questions, our proposed impact study investigates whether participating in the ENpowered intervention impacts achievement in math and science for students in middle school at the end of the first academic year of exposure to the program. In other words, we examined impact on the sixth-grade math and science grades for students first enrolled in the study while in sixth grade, the seventh-grade math and science grades for students first enrolled during their seventh-grade year, and similarly for eighth-grade final grades/participants.

The benchmark approach fits two separate impact analysis models, one for each of the mechanisms of assignment into treatment and comparison samples (school- or classroom level). Specifically, we separate the study sample into two datasets, defined by level of assignment, and estimate the effects and standard errors from each multilevel model. These model estimates are then used to calculate a precision weighted average of the program's effect using procedures outlined by Price and Wolf (2024) and detailed in Appendix C. Finally, a statistical test for the average effect is estimated to determine the *p*-value of the average impact estimate. This approach is thought to produce more accurate standard errors than pooling the samples and constructing a single analytic model, which could overestimate the standard error of the impact estimate and reduce statistical power for detecting a statistically significant treatment effect (Price, 2017).[16]

The multilevel impact models are structured similarly in that time (observations) is nested within students who are nested within either the classroom or school clusters, depending on their assignment pathway. We include school site dummies as fixed effects for the model where students are nested within classrooms. In both models, we include a set of covariates for cohort, grade level, age at baseline, race/ethnicity, gender, socioeconomic status, academic disadvantage status, as well as the IPT weights which are included at the student level.

STUDY SAMPLES

In this section, we first describe the overall sample of students selected in the school-level assignment pathway, followed by the sample of students selected into the classroom-assigned sample.

SCHOOL-LEVEL SAMPLE

Table 1 presents the aggregate (school-level) descriptive characteristics of the middle schools included in the school-level sample. We present this information as a demonstration of the first step in the sample selection process where we identify, for each implementation school, a comparison school that

[16] Additional details on the analytic procedures, including model specification and the formula for calculating a weighted average impact estimate are provided in Appendix C.

is similar on a set of school-level characteristics and from which we will identify the pool of comparison students.

Table 1. Aggregate Characteristics of Study Schools – School-Level Sample

Characteristic	ENpowered (*n* = 11)	Comparison (*n* = 11)
Average number of students	**221.9**	**317.3**
Black or African American	95.5%	93.4%
Hispanic/Latino/a	1.7%	4.1%
Eligible for free and/or reduced price lunch	69.5%	68.1%
Has an Individualized Education Plan (IEP)	11.3%	12.4%
English Language Learner (ELL)	0.3%	1.1%
Met or exceeded expectations in math[17]	20.1%	18.0%

As shown in Table 1, the schools selected to be comparison sites are similar to the implementation (ENpowered) schools in terms of student profile. A majority of students (93 to 96%) at the schools are Black, and more than two thirds are eligible for free and/or reduced price lunch (68 to 69%). About one fifth of students met or exceeded expectations in math (18 to 20%) and a small proportion have an IEP (11 to 12%). Table 2 presents the *unweighted*, descriptive characteristics of students in the treatment and comparison analytic samples for the subset of students enrolled in the study through school-level cluster assignment.

[17] We present the percentage of students who met or exceeded proficiency in math for the specific grade level that ENpowered was implemented in at the school.

Table 2. Descriptive Characteristics of Study Participants – School-Level Sample

	ENpowered (*n* = 248)		Comparison (*n* = 407)	
Characteristic	**Number Reporting**	**Statistic**	**Number Reporting**	**Statistic**
Age				
Mean age in years (at baseline)	248	12.8	407	13.0
Race/ethnicity				
Black or African American	239	96.4%	390	95.8%
Hispanic or Latino/a	2	0.8%	11	2.7%
White	1	0.4%	3	0.7%
Multiracial	5	2.0%	2	0.5%
Other[18]	1	0.4%	1	0.3%
Gender				
Female	133	53.6%	203	49.9%
Male	115	46.4%	203	49.9%
Nonbinary	0	0.0%	1	0.3%
Academic disadvantage				
Yes	31	12.5%	62	15.2%
Free and/or reduced price lunch eligible				
Yes	182	73.4%	316	77.6%
Home language				
English	242	97.6%	397	97.5%
Other[19]	6	2.4%	10	2.5%
Grade level				
6th grade	83	33.5%	89	21.9%
7th grade	135	54.4%	248	60.9%
8th grade	30	12.1%	70	17.2%

As shown in Table 2, 655 students are included in the analytic sample by way of school-level cluster assignment. The treatment group includes 248 middle school students who attended implementation schools and participated in ENpowered and the comparison group includes 407 students enrolled in similar schools that did not implement ENpowered during the given school year. Across the treatment and comparison groups, the vast majority (96%) of students identify as Black and report English as their primary language spoken at home (98%). Three quarters of the sample are eligible for free and/or reduced price lunch (76%) whereas a smaller proportion were designated as having a learning disadvantage (14%), either through an IEP or ELL plan. More than half (58%) of the treatment sample

[18] *Other* includes one student who identifies as Asian and one student who identifies as Native Hawaiian/Pacific Islander.

[19] *Other* languages reported include Spanish (*n* = 4), French (*n* = 3), Yoruba (*n* = 2), Akan (*n* = 1), Arabic (*n* = 1), Haitian-Creole (*n* = 1), Ibo/Igbo (*n* = 1), Oulof (*n* = 1), Swahili (*n* = 1), and Taiwanese (*n* = 1).

were enrolled in seventh grade at the time of program enrollment, a quarter (26%) were in sixth grade, and the remaining 15% were in eighth grade. Seventh graders make up a slightly higher proportion of the comparison sample (61%).

CLASSROOM-ASSIGNMENT SAMPLE

Table 3 presents the unweighted, descriptive characteristics of students in the treatment and comparison analytic samples for the subset of students enrolled in the study through classroom-level cluster assignment.

Table 3. Descriptive Characteristics of Study Participants – Classroom-Level Sample

	ENpowered (*n* = 177)		Comparison (*n* = 184)	
Characteristic	**Number Reporting**	**Statistic**	**Number Reporting**	**Statistic**
Age				
Mean age in years (at baseline)	177	13.1	184	13.0
Race/ethnicity				
Black or African American	72	40.7%	82	44.6%
Hispanic or Latino/a	102	57.6%	95	51.6%
White	3	1.7%	6	3.3%
Multiracial	0	0.0%	1	0.5%
Gender				
Female	83	46.9%	92	50.0%
Male	94	53.1%	92	50.0%
Learning disadvantage				
Yes	49	27.7%	88	47.8%
Free and/or reduced price lunch eligible				
Yes	160	90.4%	167	90.8%
Home language				
English	100	56.5%	106	57.6%
Spanish	71	40.1%	72	39.1%
Other[20]	6	3.4%	6	3.3%
Grade level				
6th grade	82	46.3%	84	45.7%
7th grade	0	0.0%	0	0.0%
8th grade	95	53.7%	100	54.4%

[20] *Other* languages reported include Yoruba (*n* = 5), French (*n* = 2), Akan (*n* = 1), Arabic (*n* = 1), Ibo/Igbo (*n* = 1), Philipino (*n* = 1), and Polish (*n* = 1).

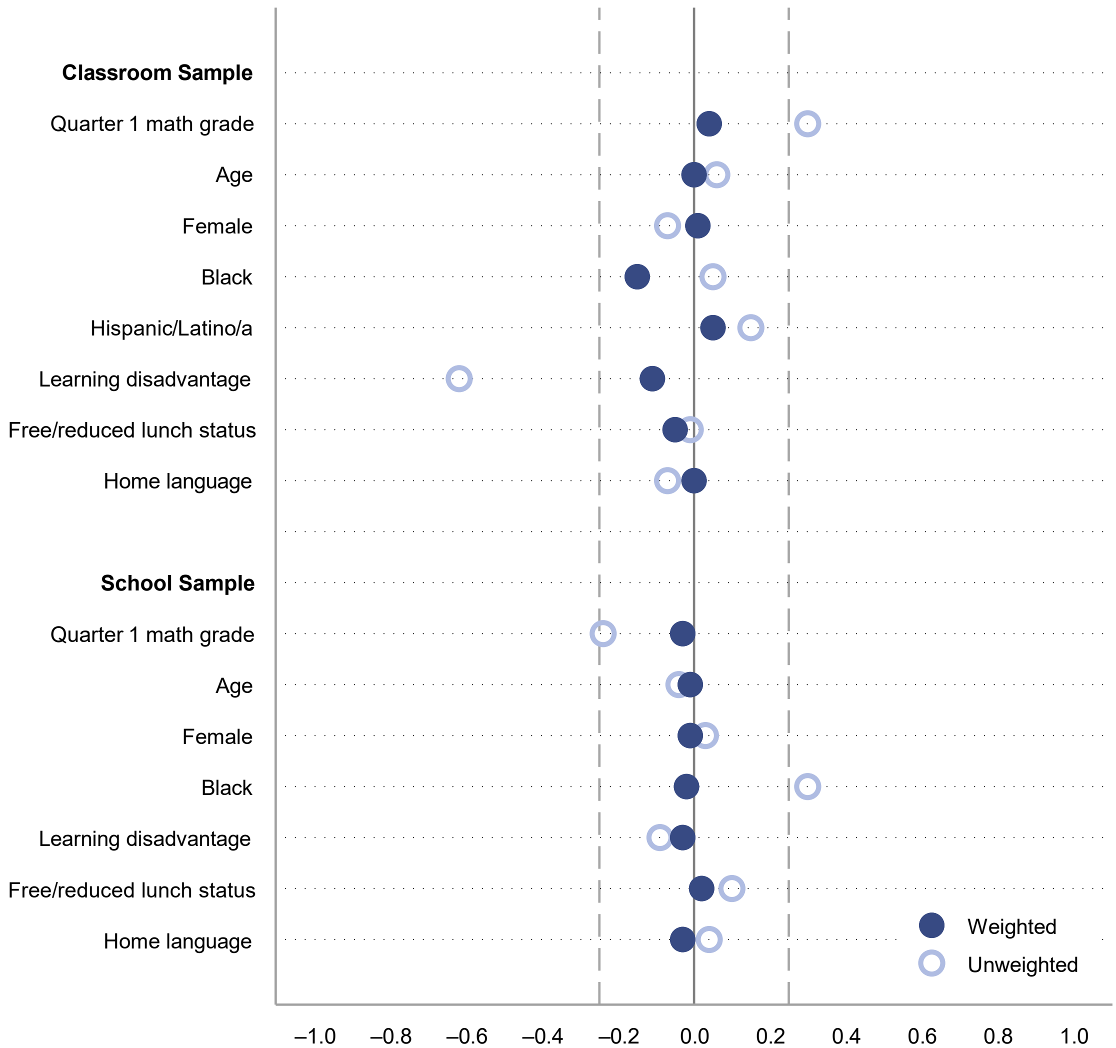
A total of 361 students are included in the impact study analytic sample by way of classroom-level cluster assignment. The treatment group includes 177 middle school students who attended implementation schools and participated in ENpowered and the comparison group includes 184 students enrolled in the same schools and grades, but who were enrolled in classrooms that did not participate in ENpowered. Across the treatment and comparison groups, more than half (55%) of students identify as Hispanic/Latino/a and 43% identify as Black. Fifty seven percent report English as their primary language spoken at home, with 43% reporting they speak a different language. A large majority (91%) were eligible for free and/or reduced price lunch and 38% were designated as having a learning disadvantage status. A larger proportion of students in the comparison group were reported to have either an IEP or ELL status compared with treatment students (48% versus 28%, respectively). About half of students were enrolled in either sixth grade (46%) or eighth grade (54%). The classroom-assigned sample did not include any seventh-grade students.

BASELINE EQUIVALENCE

As discussed in the Assignment Procedures section of this report, each of our two research questions has two independent (nonoverlapping) analytic samples, depending on whether implementation schools had a treatment and comparison class of students in the same grade, in which case students are clustered within classrooms, or only had one classroom of students in a given grade, in which case students are clustered within schools.

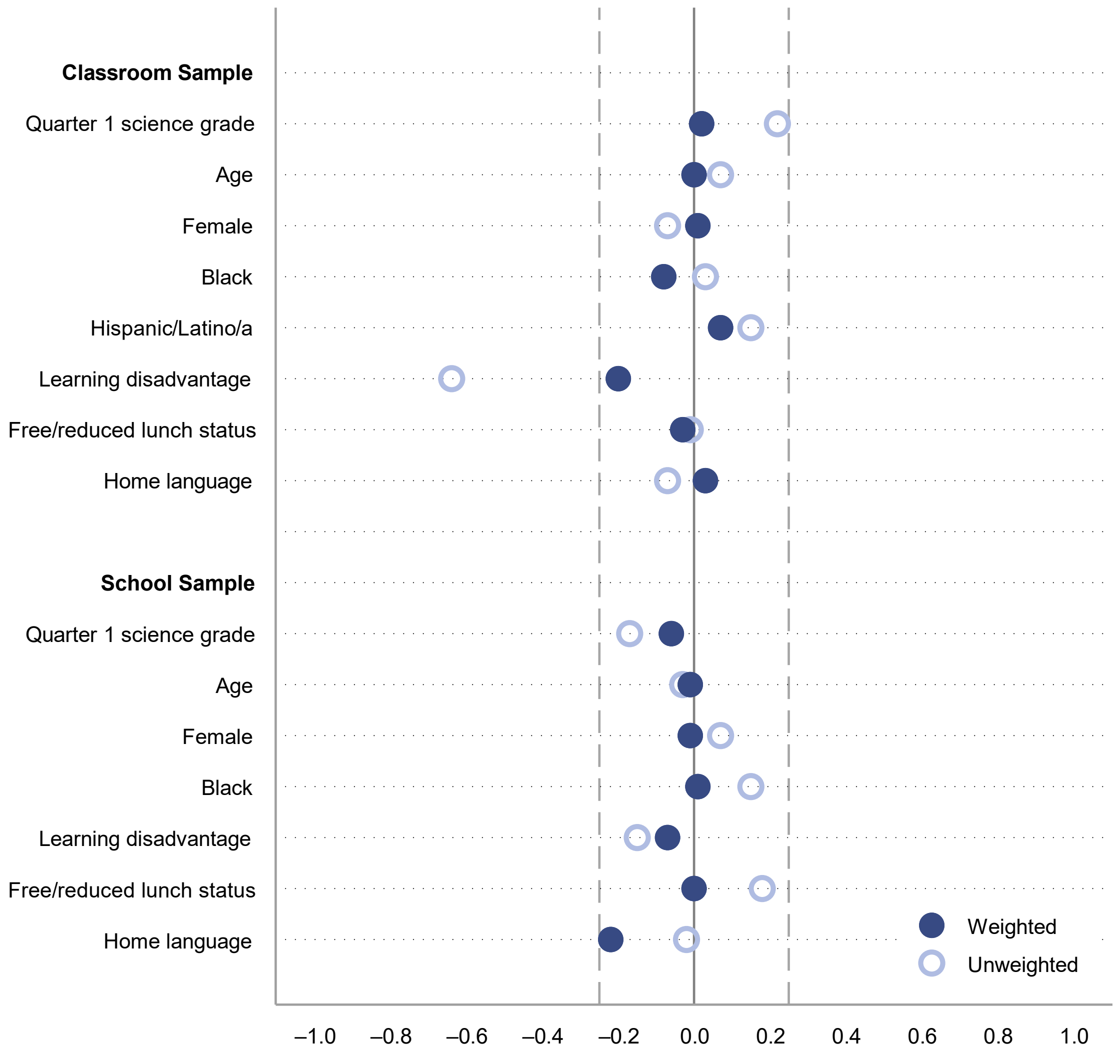
Figures 1 and 2 present the baseline balance statistics (either the standardized mean difference or difference in probability) for the baseline measure of the outcome (Quarter 1 grade) and demographic characteristics for the two analytic samples for Research Questions 1 and 2, respectively. In each figure, we present both the unweighted and weighted differences to demonstrate the degree to which the balance improved after applying the IPT weights for each sample. Vertical dashed lines indicate the window of adjustment (±0.25) for satisfying baseline equivalence standards described by the What Works Clearinghouse (WWC). Dots that are closer to the vertical solid line at zero on the *x*-axis indicate smaller differences between the treatment and comparison groups. We present details of the descriptive and standardized differences for each sample within each research question in Table C.3 in Appendix C.

Figure 1. *Baseline Equivalence of Analytic Samples – Research Question 1*



As shown in Figure 1, baseline balance statistics largely improved after propensity score weighting procedures for both the classroom- and school-assigned samples for Research Question 1 (math achievement). This is demonstrated by the dark blue dots, representing the weighted differences, aligning closer to the vertical line at zero compared with the lighter blue circles, which represent the unweighted differences. In particular, the mean difference between the treatment and comparison groups decreased to less than 0.05 for the first quarter math grades for both samples, indicating equivalence was achieved for the baseline measure of the outcome. All demographic characteristics are within the adjustment range after weighting. Since we include these measures as covariates in the impact analytic models, we satisfy the baseline equivalence standards for a QED as outlined by the WWC.

Figure 2. Baseline Equivalence of Analytic Samples – Research Question 2



As with Research Question 1, baseline balance statistics for both the classroom- and school-assigned samples for Research Question 2 (science achievement) largely improve after weighting procedures. The mean difference between the treatment and comparison groups decreases to less than 0.06 for the first quarter science grades for both samples. All demographic characteristics are again within the adjustment range, satisfying WWC criteria for baseline equivalence for QEDs.

IMPLEMENTATION OF ENPOWERED

In this section, we provide a brief overview of the implementation of the ENpowered program each study year followed by some descriptive and qualitative perspectives from the school-based coordinators and students who participated in the program during the study. Table 4 presents the ENpowered program details for each year of implementation, including the field of engineering at the center of a given year's program, the motivating "real-world" problem and project students worked together to address during the 10-week class, as well as the additional challenge(s) completed during the ENpowered Games event and the relevant technology used throughout the program.

Table 4. ENpowered Program Details

Program Details	Spring 2022	Spring 2023	Spring 2024
Number of schools[21]	9 total; 6 in study	9 total; 7 in study	13 total; 6 in study
Engineering field	Computer Science	Mechanical Engineering Computer Science	Mechanical Engineering Computer Science
Motivating problem	Rising concerns over energy consumption and environmental sustainability have created a pressing need for green smart homes that minimize ecological impact. By integrating energy-efficient technologies and automations, green smart homes can optimize resource usage, reduce carbon footprints, and promote sustainable living.	The growing demand for faster and more reliable logistics driven by e-commerce has highlighted the inefficiency of manual package sorting, which is time-consuming and prone to errors. Machines capable of efficiently sorting packages can address these challenges by increasing speed, accuracy, and scalability in supply chain operations.	As autonomous vehicles become more prevalent, ensuring their safety is critical to gaining public trust and preventing accidents. Advanced safety systems are needed to address challenges such as real-time decision making, handling complex environments, and mitigating obstacles in unpredictable scenarios.
Small-group project prompt	Use the EDP to automate an aspect of a home to make it more environmentally friendly, making sure to identify the environmental challenge the solution addresses.	Use the EDP to create a machine that automatically identifies and packages a product. The challenge will test students' knowledge of mechanisms, design thinking, and sensors.	Use the EDP to design an autonomous car that self-drives and navigates obstacles while on the road. The challenge will test students' knowledge of mechanisms, design thinking, and sensors.
Relevant technology/ tools	Tablets and *Lego* SPIKE Prime Kits	Tablets and *Lego* SPIKE Prime Kits	Tablets and *Lego* SPIKE Prime Kits
ENpowered Games challenge(s)	Design, build, and test a robot that automatically finds cups on a playing field and pushes them into a series of zones to gain the most points.	Student groups receive a set of code that contains errors that prevent their machines from working. Within 30 minutes, each team must debug the code and demonstrate that it works on their machine. This challenge will test students' coding knowledge and troubleshooting skills.	Student groups collaborate to program their autonomous vehicles to follow specific instructions using the minimum number of coding blocks. Working together, they strategize and determine the optimal code sequence, create a detailed flowchart depicting their chosen code structure, and demonstrate the code's functionality.

The ENpowered program aims to help middle school students understand how technology and engineering, and in particular, the EDP, can be used to solve real-world problems they may encounter. The focus of the ENpowered program changes each year, to ensure students are engaged in relevant, timely problem solving that addresses current engineering challenges. At the start of the 10 weeks, students are oriented to the program through the presentation of a real-world issue or problem that their work will aim to address. Over the course of the 10 weeks, they learn about principles of the EDP,

[21] Additional program-only schools that did not participate in the grant-funded study included those that implemented as an after-school program or who declined to participate in the evaluation (and therefore did not receive grant-funded stipends to cover the cost of programming).

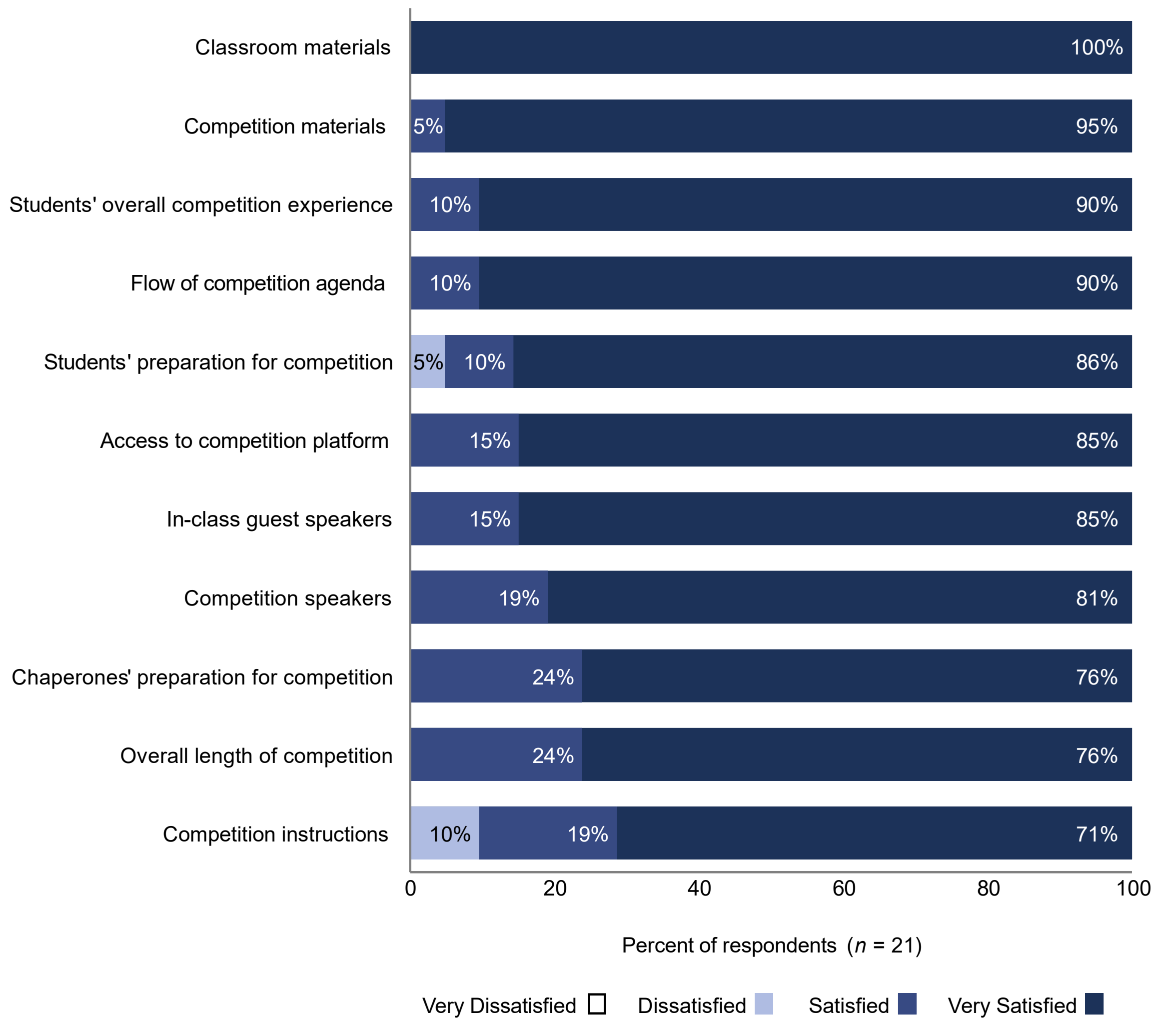
acquire and demonstrate computer science skills, and work within a small group to design, test, and improve their solution to the problem using scientific thinking. Students work together during the class to document their progress on completing their projects and then showcase their work to a panel of judges at the ENpowered Games event.

For Cohort 1 (spring 2022), student projects were grounded in computer science and focused on incorporating automation and efficiency into residential homes to reduce ecological impacts of energy usage. Students worked in small groups to automate an aspect of a home, either a particular room or the entire house. At the culminating Games, students had to work in their groups to build and test a robot that could automatically identify objects on a playing field and move them into specific zones. During Cohort 2 (spring 2023) student projects were grounded in mechanical engineering and computer science and focused on designing and building a machine that could automatically sort different items, as a means of increasing e-commerce efficiency. While at the Games event, they had to work together to identify and correct a bug in their final project's coding to get their machine working again before time ran out. In the final cohort (spring 2024) students again worked within mechanical engineering and computer science fields to design an autonomous vehicle that could navigate different obstacles on the road. At the Games, students were challenged to program their vehicle to navigate a previously unseen course with the fewest coding blocks necessary to complete the track. In all three years, students used *Lego* SPIKE Prime kits to build their devices or machines.

TEACHER PERSPECTIVES

At each school, the teachers provided feedback to PS at the end of the program, including their satisfaction with different aspects of the classroom lessons and the ENpowered Games competition. Figure 3 presents the responses from the *Teacher Feedback Survey*, a tool developed and administered by grantee staff. Teachers were asked to report their level of satisfaction across 11 items relating to classroom materials and preparation for the competition, as well as aspects of the competition (e.g., materials, instructions, flow) and students' overall experience.

Figure 3. *Teacher-Reported Satisfaction*[22]



Overall, teachers reported being satisfied or very satisfied with the different aspects of classroom and competition programming. All teachers (n = 21) reported being very satisfied with the materials provided to students for classroom instruction and reported satisfaction with 8 of the 10 remaining items addressing competition materials, flow, speakers, technology, chaperone preparation, and students' overall experience. A small proportion of teachers reported being dissatisfied with students' preparation for the competition (5%) and the clarity of instructions given to students during the competition (10%); otherwise teachers reported being satisfied or very satisfied with these components. No teachers reported being very dissatisfied with any of the aspects of programming reflected in the survey.

Teachers provided more insight into their perspectives on the program through open-ended questions on the *Teacher Feedback Survey* and a select few participated in semi-structured interviews with PRG

[22] For the following items, n = 20: *In-class guest speakers; Access to competition platform:* and *Classroom materials*.

staff at the end of the grant.[23] On the survey, teachers were prompted to provide their perspective on what they think students get out of participating in a program like ENpowered. When asked to describe what part of the competition experience they think students most resonated with, several indicated that students enjoyed interacting with and presenting to the judges panel of STEM professionals and receiving positive feedback. One teacher reflected, “my students were extremely nervous and that encouragement from the judges and staff was extremely inspiring.” Other teachers reported that students enjoyed being able to demonstrate their skills and succeed in completing the challenges on time, with one teacher noting that although their students did not place in the top teams that day, “the students felt really good about their work.” Finally, two teachers noted that their students really enjoyed “seeing other students like them” competing and succeeding at the competition. The teachers who participated in interviews with evaluation staff largely echoed these points and offered that their students developed a strong connection with their co-instructors over the 10 weeks.

Teachers also provided their thoughts on how a program like ENpowered could be impactful for middle school students. About half of respondents reported that the program provides valuable exposure to the STEM field, particularly computing and robotics, and broadened students’ horizons for what type of career they might be interested in pursuing. One teacher noted that the program gives students new perspective of the field and “students who may not have been interested in coding get to see how interesting this world is,” while another noted that the program “builds their confidence that they are competent to consider engineering careers.” Other teachers suggested that the program helps students strengthen critical skills such as problem-solving, collaboration, and critical thinking, while others noted that it reinforces the need for persistence, with one explaining, “it gives them hope to know what seems impossible is in fact possible.”

STUDENT PERSPECTIVES

At the end of each year of programming, students reported their general satisfaction with the program on a brief *Student Feedback Survey*, developed and administered by PS staff at the conclusion of the ENpowered Games event. Students shared whether they felt prepared going into the competition, and if they would participate again and/or recommend the program to others. Figure 4 presents the proportion of participants who reported they felt prepared to go into the ENpowered Games competition at the end of their classroom curriculum, by grade. Figure 5 then presents the proportion of participants who reported they would participate in the program again if offered and who would recommend the program to friends or family, also by grade level.

[23] Three teachers from Cohort 3 agreed to participate in an interview with PRG staff at the end of programming. Two of these teachers had also participated in earlier cohorts of the study.

Figure 4. Perceived Competition Preparedness, by Grade Level

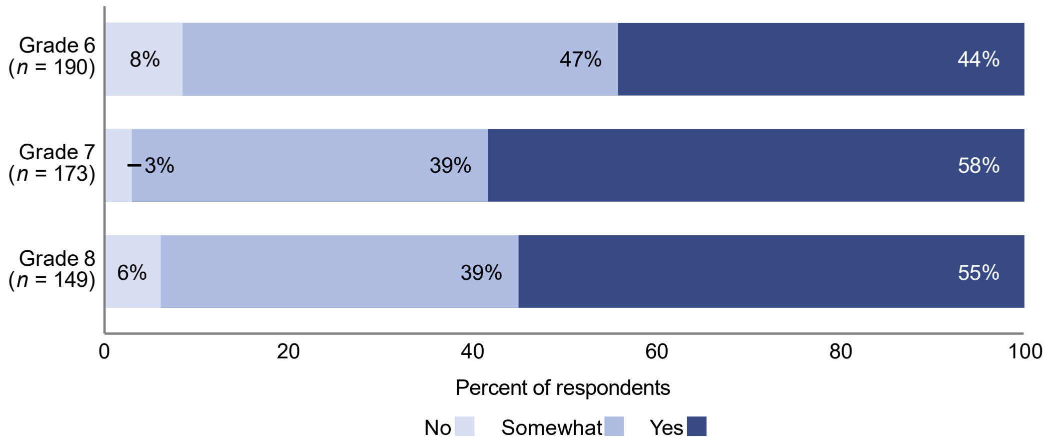
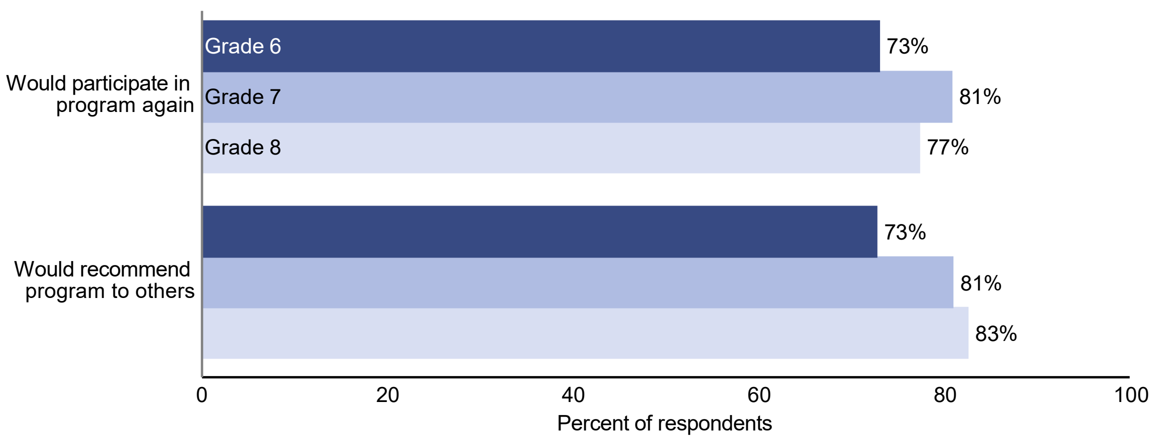


Figure 5. Student Satisfaction, by Grade Level[24]



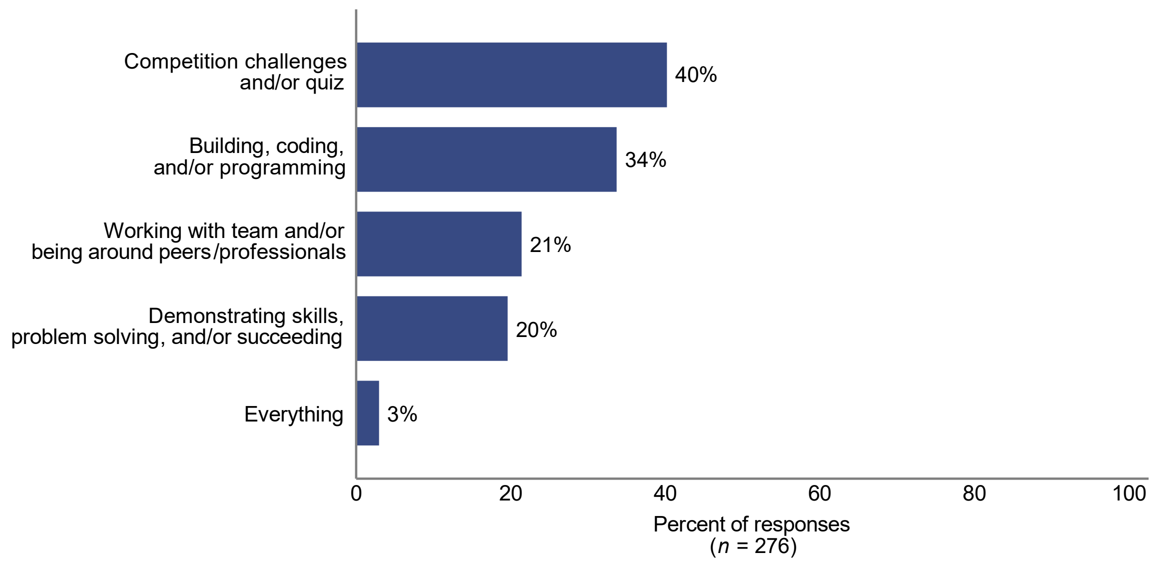
Across grade levels, more than half (52%) reported feeling prepared going into the ENpowered Games competition, with an additional 42% reporting they felt *somewhat* prepared. Only a small proportion arrived at the event not feeling prepared for the competition. Across grades, seventh graders were the most likely to report feeling prepared (58%), with sixth graders being the least likely (44%).

Overall, students reported a high degree of satisfaction after participating in the ENpowered program. More than three quarters of participants reported that they would participate in the program again (77%) and recommend the program to others (78%). As shown in Figure 5, seventh and eighth graders were the most likely to report being satisfied with the program, with sixth graders reporting comparatively lower, yet still high, rates of satisfaction (73%).

[24] Sample sizes for participation are as follows: Grade 6 (*n* = 189), Grade 7 (*n* = 172), Grade 8 (*n* = 150). For recommending the program, sample sizes are as follows: Grade 6 (*n* = 187), Grade 7 (*n* = 173), Grade 8 (*n* = 149).

On the same feedback instrument, students were asked to report their favorite part of the ENpowered program using an open-ended question. Students' responses were reviewed and grouped into the following categories describing program features or activities: (1) activities involving building, coding, or programming their devices; (2) working with their team or interacting with others at the competition (e.g., judges, volunteers, peers, etc.); (3) one or more of the competition challenges; (4) being able to present or demonstrate learned skills, including succeeding in a challenge or problem solving; and (5) responses that indicated the student liked "everything" about the program. Figure 6 presents a summary of responses from students' open-ended responses.

Figure 6. Student-Reported Favorite Program Activities[25]



Students were most likely to report that their favorite part of the ENpowered program was a specific challenge they completed at the ENpowered Games event, including a Kahoot learning quiz conducted at the end of the event (40%).[26] Students also reported enjoying the process of building, coding, and/or programming their devices (34%; e.g., "building the robot" or "coding the car"). About one fifth (21%) of students noted enjoying being able to work with or contribute to their teams (e.g., "doing my best with my team" or "helping to do my part") and/or interacting with others at the ENpowered Games event (e.g., "looking at everyone's builds and talking"). A similar proportion (20%) also noted being able to successfully demonstrate skills they learned in the program as their favorite part (e.g., "I loved seeing the code work because it felt rewarding" or "being able to compete and put all our hard work together").

[25] Student responses could be coded to reflect more than one feature or activity. As a result, proportions presented in Figure 6 do not sum to 100%.

[26] For more info see https://kahoot.com/

RESULTS

In this section, we present the findings from the benchmark analyses for each of the two confirmatory research questions, as well as a discussion of additional subgroup analyses. Table 5 presents select estimates from the weighted, multilevel DID regression model for the two analytic samples (classroom and school assignment pathways), as well as the weighted average of the impact estimates across the two samples for both Research Questions 1 and 2. For each sample, we present the model-adjusted mean grade in math and science at the end of the first quarter (baseline) and at the end of the spring term (post program) for participants in the ENpowered and comparison conditions, as well as the difference between the difference in baseline and post grades across conditions (the impact estimate), its standard error, *p*-value, and standardized mean difference (effect size). We then report the weighted average impact estimate and effect size across the two analytic samples.[27] Figure 7 presents the change in mean grades in math and science for treatment and comparison groups at each time point graphically.

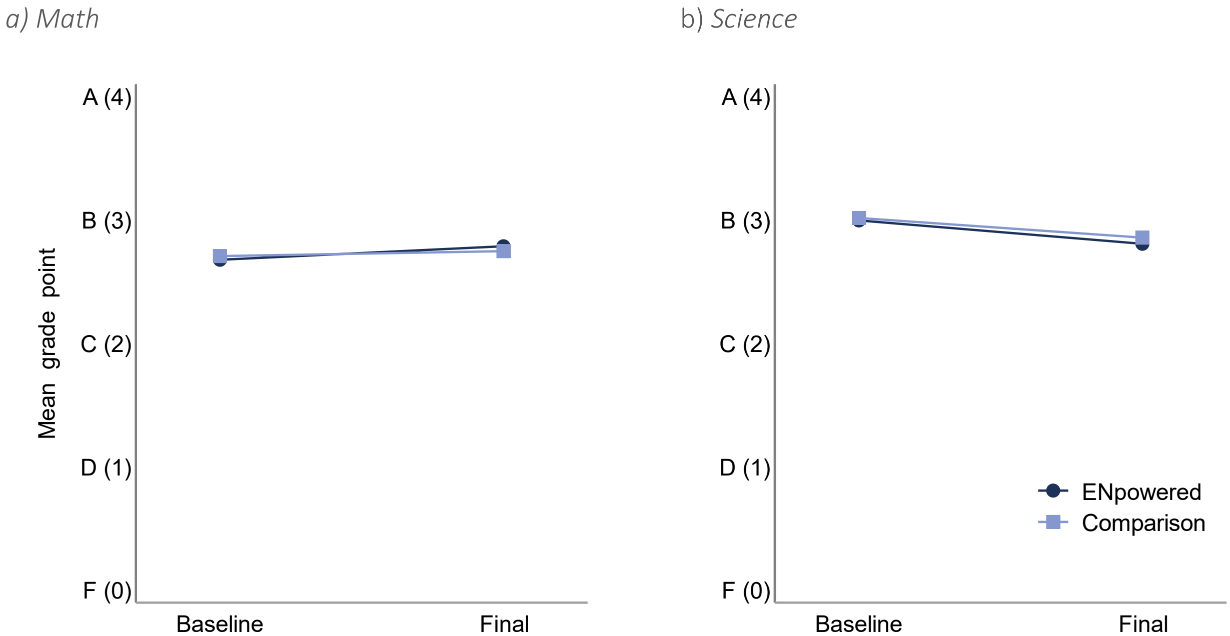
Table 5. Research Questions 1 and 2 – Statistical Estimates

		ENpowered Means		Comparison Means				
Analytic Sample	**Number Reporting**	**Baseline**	**Post**	**Baseline**	**Post**	**DID Impact Estimate (SE)**	***p*-value**	**Effect Size**
Research Question 1								
Classroom sample	359	2.71	2.70	2.69	2.68	0.01 (0.17)	0.976	0.01
School sample	623	2.68	2.79	2.71	2.75	0.07 (0.11)	0.513	0.08
Weighted average						*0.05 (0.09)*	*0.571*	*0.06*
Research Question 2								
Classroom sample	361	3.00	2.81	3.02	2.86	–0.03 (0.10)	0.735	–0.04
School sample	532	3.01	2.98	3.05	3.08	–0.06 (0.21)	0.766	–0.08
Weighted average						*–0.04 (0.09)*	*0.664*	*–0.05*

Note: ~ $p < 0.10$; *$p < 0.05$; **$p < 0.01$.

[27] We provide the full model output in Tables D.1 and D.2 in Appendix D.

Figure 7. Quarter 1 and Final Grades, by Treatment Group



RESEARCH QUESTION 1: MATH ACHIEVEMENT

Benchmark statistical estimates for Research Question 1 indicate that students who participated in the ENpowered program had slightly greater growth in mathematics grades than comparison students, but this observed difference was not statistically significant. Estimates presented in Figure 7 show that students in both treatment and comparison improved their math grades from baseline to post program, and that treatment students' growth was slightly greater than comparison students, but the difference was very small. The weighted average impact estimate is 0.05 (p = 0.571), indicating treatment students' grades improved from baseline to post less than one tenth of a grade point more than comparison students. The standardized magnitude of effect for this difference is 0.06, indicating a small, but not statistically significant, effect in the desired direction. We examined the program's impact at the grade level and for male and female students separately but did not find any evidence of subgroup effects (impact estimates were not significant).

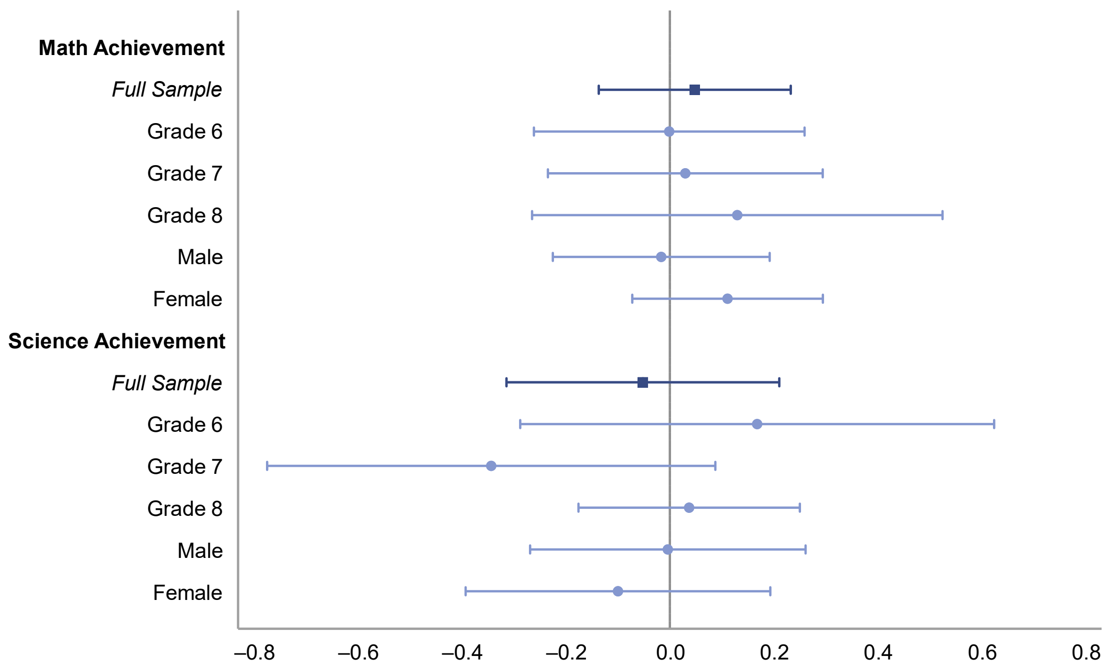
RESEARCH QUESTION 2: SCIENCE ACHIEVEMENT

Benchmark statistical estimates for Research Question 2 indicate that students who participated In the ENpowered program had slightly lower growth in science grades than comparison students, but this observed difference was not statistically significant. Estimates presented in Figure 7 show that students in both treatment and comparison groups declined slightly from baseline to post program in their science grades, with treatment students declining slightly more, but this difference is very small. The weighted average impact estimate, or the difference between the group-specific pre-post trend, is –0.04 (p = 0.664), indicating treatment students grades declined from baseline to post less than one tenth of a grade point more than comparison students. The standardized magnitude of effect for this difference is –0.05. We again examined the program's impact at the grade level and for male and female students separately but did not find evidence of heterogeneous effects (impact estimates were not significant).

DISCUSSION

Benchmark findings indicate that participating in the ENpowered program did not have a detectable impact on the confirmatory outcomes of final grades in math and science for middle school students. For both outcomes, we observed no difference in the end-of-course grades in math and science between the students who participated in ENpowered and a similar group of students who did not participate. Subgroup analyses that examine ENpowered's impact at each grade level and by gender corroborated the benchmark results. Figure 8 presents the DID impact estimates and 95% confidence intervals for each subgroup examined for both outcomes.

Figure 8. Subgroup Impact Estimates



As shown in Figure 8, we do not observe any statistically significant heterogeneous program effects for a specific grade level or gender (as denoted by the fact that all confidence intervals overlap with the vertical line at zero). Although treatment students in eighth grade appear to achieve comparatively higher math grades (effect size of 0.14) and sixth-grade treatment students appear to achieve higher science grades (effect size of 0.19), these differences do not reach statistical significance.

Although null results are not desirable in the context of promising educational programming, they are not uncommon in applied research. A cross-project summary of 67 Investing in Innovation (i3) grantees, a predecessor to the EIR program, found that less than one fifth (18%) of studies identified positive statistically significant impacts on student outcomes (Boulay et al., 2018). Among the subset of 30 evaluations that examine impact on science or math outcomes, 6 (20%) identified positive effects.

We believe we have produced the most rigorous nonexperimental analysis possible; however, any nonexperimental design has its constraints. The primary constraint of a nonexperimental design is the threat of unobserved bias in our treatment and comparison groups. Unweighted balance statistics presented in Figures 1 and 2 indicate that the treatment and comparison groups were broadly similar on observed baseline characteristics, including the baseline measure of the outcome, and weighting procedures made the two groups even more aligned. Although there is no way to be certain that our treatment and comparison groups are truly equivalent, we believe that the two-stage process we employed to first identify similar external comparison schools and then weight students individually provided an internally valid contrast.

The ENpowered program aims to introduce a new avenue for applying STEM skills to tackle real-world problems and in so doing spark an interest and passion for pursuing STEM opportunities for students who are historically underrepresented in the field. Participant and teacher qualitative feedback indicate that students enjoyed the experience and specifically the opportunity to learn and demonstrate a new skill through the competitive event at the end of the semester. The program is not necessarily designed to be a one-time opportunity and schools can offer students the opportunity to participate across multiple years in middle school. Our study examined the potential for promising evidence of participating for one semester on pragmatic outcomes of math and science grades at the end of the implementation semester. Future research should examine the full scope of the program's theory of change across the long-term outcomes of pursuing additional STEM electives in high school and pursuing and completing STEM postsecondary degrees to more fully understand the potential impact a program like ENpowered can have on middle school students.

REFERENCES

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, *34*(28), 3661–3679.

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., Gan, K., Harvill, E., & Sarna, M. (2018, June). *The investing in innovation fund: Summary of 67 evaluations: Final report* (NCEE 2018-4013). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Brophy, S., Klein, S., Portsmore, M., & Rogers, C. (2008). Advancing engineering education in P-12 classrooms. *Journal of Engineering Education*, *97*(3), 369–387. https://doi.org/10.1002/j.2168-9830.2008.tb00985.x

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187–1218. https://doi.org/10.1002/tea.20237

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*(4), 724–750.

Guo, S., & Fraser, M.W. (2010). *Propensity score analysis: Statistical methods and applications*. Sage.

Hilbe, J. M. (2014). *Modeling count data*. Cambridge University Press.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics*, *7*(1), 295–318. https://doi.org/10.1214/12-AOAS583

McNeish, D. Stapleton, L. M., & Silverman, R. D. On the unnecessary ubiquity of hierarchical linear modeling. (2017). *Psychological Methods*, *22*(1), 114–140.

National Center for Science and Engineering Statistics. (2023, January). *Diversity and STEM: Women, minorities, and persons with disabilities* (Special Report NSF 23-315). National Science Foundation. Retrieved April 20, 2025, from https://nsf-gov-resources.nsf.gov/doc_library/nsf23315-report.pdf?VersionId=OfErRcu.MTEp_KCzcuAfyfIkh0cU_XQO

Pew Research Center. (2021, April). STEM jobs see uneven progress in increasing gender, racial and ethnic diversity. Retrieved April 20, 2025, from https://www.pewresearch.org/wp-content/uploads/sites/20/2021/03/PS_2021.04.01_diversity-in-STEM_REPORT.pdf

Price, C. S. (2017, February). *Analysis approaches for designs where both clusters and individuals are randomly assigned to treatment and control conditions*. Abt Associates.

Price, C., & Wolf, A. (March 2024). *Calculating average treatment effects for two scenarios: Non-overlapping subsamples and multiple outcome measures in the same domain.* Prepared for the U.S. Department of Education, Institute of Education Sciences. Abt Associates.

Samuels, K., & Seymour, R. (2015) The middle school curriculum: Engineering anyone? *Technology and Engineering Teacher*, *74*(6), 8–12.

Schembri, S., & Rampersad-Ammons, J. (2021). Science identity and Hispanic students in science, technology, engineering, and mathematics (STEM). *Journal of Women and Minorities in Science and Engineering*, *27*(5), 1–24. https://doi.org/10.1615/JWomenMinorScienEng.2021035960

Sithole, A., Chiyaka, E., McCarthy, P., Mupinga, D., Bucklein, B., & Kibirige, J. (2017). Student attraction, persistence, and retention in STEM programs: Successes and continuing challenges. *Higher Education Studies*, *7*(1), 46–59. https://doi.org/10.5539/hes.v7n1p46

Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the What Works Clearinghouse (Phase I). *Educational Evaluation and Policy Analysis*, *32*, 351–371.

Tai, R. H., Qi-Liu, C., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science* (New York, N.Y.), *312*(5777), 1143–1144.

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook* (Version 5.0). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks

APPENDIX A. ENPOWERED LOGIC MODEL

Inputs

- Board of directors
- Staff (instructors and admin)
- Engineering-focused project-based learning curriculum
- *Lego* SPIKE Kits and other materials
- Computers
- Host site for competition
- Partner schools
- Independent contractors
- Corporate partners
- Mentors/volunteers
- College/career readiness partners
- Evaluation tools
- Funding
- Student data
- Curriculum

Key Components and Indicators

Professional Development

- Train instructors on curriculum each year
- Conduct observations to ensure program fidelity
- Conduct biweekly program meetings to provide continuous feedback

Classroom Programing

- 20 hours of instruction offered
- Engineering professional partner visits ENpowered classroom at least once

ENpowered Games

- ENpowered Games held
- Corporate partners engage in ENpowered Games
- Students attend ENpowered Games
- Student groups complete engineering notebook
- Students present design plan to judges

Mediators

Instructors:

- Improved instructional skills

Students:

Increased:

- Awareness of STEM career opportunities
- Self-efficacy in STEM
- STEM engagement
- Ability to explain the connection between engineering and real-world applications
- Understanding of engineering principles and practices
- Self-concept in STEM
- Interest in STEM Careers

Outcomes

Short

- Increased achievement in math
- Increased achievement in science

Medium

- Improved peer relationships
- Increased participation in academic or extracurricular STEM programs
- Increased enrollment in advanced math and science classes in high school
- Improved grades in school
- Increased proficiency in engineering skills and practices

Long

- Increased knowledge about STEM careers
- Increased understanding and knowledge of STEM content
- Decreased gap between aspiration and expectation for students' college aspirations
- Increased diversity in STEM fields
- More equitable access to STEM education
- Pathway of opportunities for students to stay engaged in STEM fields
- More students pursue STEM degrees in college

APPENDIX B. FIDELITY OF IMPLEMENTATION

The purpose of this appendix is to present results of The Policy & Research Group's (PRG) fidelity of implementation study of the ENpowered program. In the below matrix, schools are given a numeric score for their level of performance on indicators of program implementation. For all indicators listed below each key component, the unit of implementation is either the program or implementing school. We implemented this study in 19 schools over three academic years. We first present a description of the ENpowered program and its key components, followed by a description of the data sources used in the implementation study. The unit-level implementation fidelity scores are presented in the fidelity matrix table and we then aggregate the implementation findings to the cohort (school year) level. A depiction of the ENpowered logic model can be found in Appendix A.

KEY COMPONENTS

PROFESSIONAL DEVELOPMENT

The ENpowered curriculum is taught to students in a classroom-based environment by two Project SYNCERE (PS) instructors. Each instructor receives up to 30 hours of training before they begin teaching the program in schools. Some of this training curriculum is self-directed, with instructors reading training materials on their own, and some is led by PS program managers. Instructors are assigned to teach in pairs at each school.

Each school is also assigned a program manager who monitors that the school's two instructors are following the prescribed pace of the program curriculum. Program managers conduct visits to their assigned schools monthly and, while there, they observe instructor performance and complete an observation form that assesses the degree to which the instructors facilitated the classroom session appropriately. Feedback from the observation forms is reviewed with instructors at bimonthly reoccurring meetings, unless an issue is identified in which case it is addressed sooner.

CLASSROOM PROGRAMMING

Each spring, PS offers 20 hours of classroom instruction (usually two 1-hour lessons each week for 10 weeks), using the problem-based learning curriculum, to students during the regular school day. During the 20 hours of classroom programming, PS instructors present information on basic engineering principles before moving into the specific engineering field of focus for that program year (e.g., electrical engineering, biomedical engineering, computer engineering, chemical engineering). Throughout the semester, students apply what they learn through small design challenges, tracking their progress and notes in an engineering notebook.

Each year, PS engages several corporate partners located in the Chicago community who designate one or more of their employees to act as volunteers for the ENpowered program. Corporate partners can be any business in Chicago whose employees participate in the kind of science, technology, engineering, and math (STEM) work that students in the ENpowered program are learning about during a particular year. During at least one of the classroom lessons at each school, a volunteer from one of the corporate partners visits the classroom to discuss their career and answer student questions.

ENPOWERED GAMES

At the end of the spring semester, after the 20 hours of classroom instruction have concluded, PS holds the ENpowered Games event which is attended by all students who received ENpowered classroom instruction during the semester. Program students from all school partners (both schools that participated in the evaluation and ones that did not) are expected to attend the event.

In addition, each corporate partner sends at least one representative to serve as a volunteer at the event. Corporate partner volunteers can act in a variety of capacities during the ENpowered Games event, but most commonly act as judges where they rate student small groups on their presentations and performance on challenges, using a rubric provided by PS.

At the ENpowered Games event, students are judged on two core products showcasing their work and knowledge gained: completion of a timed engineering design challenge and the presentation of their completed engineering notebook. For the first activity, students are presented with a design challenge that requires them to apply the knowledge gained from all the small group work completed throughout the semester. The panel of judges observes the students as they address the final challenge in their small groups and assigns each group a score based on their final design. In the second activity, the students present the content of their engineering notebook, which is completed throughout the duration of the classroom-based programming, as a small group to the panel of judges. At the end of the ENpowered Games event, there is an award ceremony acknowledging the top 10 performing student groups and the top performing school overall.

DATA COLLECTION PLAN AND KEY MEASURES

DATA SOURCES

In this section, we outline the data collection instruments PRG used to conduct the implementation evaluation of ENpowered.

IMPLEMENTATION SUMMARY FORM

PS staff tracked several pieces of implementation data on PRG's *Implementation Summary Form* (ISF), an Excel worksheet. The form was completed by PS staff throughout ENpowered program implementation each year. The form documented information used to report whether or not two trained instructors were assigned to each school during each year (Professional Development key component) and the date of the ENpowered Games event (ENpowered Games key component).

PROGRAM MANAGER CHECK-IN REPORT

Program managers meet with instructors on a biweekly basis during classroom instruction to provide instructors with continuous feedback on lesson pacing, troubleshooting if a school is behind schedule, and to discuss strategies for classroom management and diverse learning needs. Program managers document these check-ins using the *Program Manager Check-In Report*, an Excel worksheet that documents the date of each check-in and the lessons completed each week. PS submits the completed worksheet to PRG at the end of programming each spring. PRG uses the information to report on the number of meetings held each school year (Professional Development key component).

ENGINEERING PROGRAM FACILITATOR OBSERVATION FORM

PS program managers observe at least one, but ideally two, ENpowered classroom-based lessons at each study school during the implementation period. Observations are conducted in person in the classroom. During each observation, program managers fill out an *Engineering Program Facilitator (EPF) Observation Form*, which documents the following:

- The site (school) name, grade of students, and the date of the observation
- The names of the instructors

- The overall class environment
- A score from 1 to 4, where 1 indicates poor instruction and 4 indicates optimal instruction for:
    - How well the lesson was planned and prepared
    - How effectively and to what degree the instructor used materials available to enhance student learning
    - How well the two instructors worked together to present the lesson
    - The quality of the classroom culture
    - How well the instructors followed established routines during the lesson
    - How efficiently the instructors used the classroom time to present the lesson
- Whether the observation requires follow-up intervention

The form collects data that are used to report on the number of observations completed (Professional Development key component).

ATTENDANCE TRACKER

At the beginning of each lesson, PS instructors record the date of the lesson, lesson number, and student-level attendance (present/absent) on the electronic *Attendance Tracker* developed by PS. At the end of each spring semester once classroom-based instruction has concluded, PS sent PRG an electronic attendance report that reports, for each student, the number of lessons attended and the number of lessons the student was absent. We use these data to report the number of lessons offered (Classroom Programming key component).

ENPOWERED GAMES ATTENDANCE LOGS

Program coordinators at each school track their students' attendance at the ENpowered Games event held at the end of each spring. Program coordinators record, for each student, whether the student was present at the ENpowered Games. PS then scanned the attendance sheets and submitted copies to PRG. We use these data to report whether or not students from each study school attended the games (ENpowered Games key component).

SALESFORCE VOLUNTEER TRACKING SOFTWARE

At each school, during at least one ENpowered classroom lesson, a local STEM professional volunteers to join the class and discuss their career with the students. In addition, PS recruited volunteers from each corporate sponsor to attend the culminating ENpowered Games event each spring. PS collected information about volunteer hours in their SalesForce platform and submitted an aggregate report to PRG at the end of each spring semester. We use the aggregate volunteer reports to report on volunteer attendance during the classroom curriculum (Classroom Programming key component) and the number of volunteers at the ENpowered Games (ENpowered Games key component).

ENPOWERED GAMES JUDGE RUBRICS

Volunteers serve as judges of student presentations on their engineering notebooks, completed during classroom programming, and completion of the design challenge, completed at the ENpowered Games. Judges complete a rubric provided by PS that assesses the quality of the engineering notebook presentation and a second rubric assessing the quality of the design challenge. PS staff collect these rubrics from the volunteers at the end of the ENpowered Games event, compile scores, and submit final scores for each group to PRG after the conclusion of the event. PRG uses the rubric scores to verify that each student group presented on their engineering notebook and completed a design challenge (ENpowered Games key component).

ANALYSIS APPROACH

To assess the degree to which each key component of the intervention was implemented with fidelity, we reviewed data for each of the three intervention components, during each year of implementation. For each component, indicator scores were summed to create a total component score for each school year. To determine whether a key component was implemented with fidelity for the intervention sample, we calculate the percentage of intervention schools that implemented the component with fidelity during each school year. The specific thresholds for fidelity for each school year are defined at both the school- and sample level in Tables B.1 through B.3 for each key component.

PROFESSIONAL DEVELOPMENT

Professional Development key component fidelity is measured using three indicators. Schools with a score of 5 are considered to have implemented the Professional Development component with fidelity for the school year. The component was considered to have been implemented with fidelity in the sample for the school year if at least 75% of intervention schools implemented the component with fidelity.

Table B.1. Key Component 1: Professional Development

Indicators	Definition	Unit of Implementation	Data Source	Score for Level of Implementation at Unit Level
1.1 Trained instructors assigned to teach class	Two trained instructors from PS assigned to teach at each school	School	ISF	0 = fewer than 2 trained instructors assigned to school 1 = 2 trained instructors assigned to school
1.2 Program managers conduct classroom observations	Program managers conduct monthly observations of classroom lessons at each school to monitor fidelity	School	EPF Observation Form	0 = 0 observations conducted at the school 1 = 1–2 observations conducted 2 = 3 or more observations conducted
1.3 Program managers provide continuous feedback to instructors	Program managers hold biweekly meetings with instructors to provide continuous feedback on pacing and classroom management	School	Check-in log	0 = 0–1 meetings held 1 = 2–3 meetings held 2 = 4 or more meetings held
All Indicators	Score range: 0–5 Unit-level adequate implementation score: 5			Adequate implementation at sample level: 75% of schools with score of 5

CLASSROOM PROGRAMMING

The Classroom Programming key component fidelity is measured using two indicators. Schools with a score of 3 are considered to have implemented the Classroom Programming component with fidelity for the school year. The component was considered to have been implemented with fidelity in the sample for the school year if at least 75% of intervention schools implemented the component with fidelity.

Table B.2. Key Component 2: Classroom Programming

Indicators	Definition	Unit of Implementation	Data Source	Score for Level of Implementation at Unit Level
2.1 Ten weeks of programming offered	Each school implements 10 weeks of programming, defined as 10 lessons if implementing once per week or 20 lessons if implementing twice per week	School	Attendance Tracker	1 time/week: 0 = <4 lessons held 1 = 5–7 lessons held 2 = 8+ lessons held 2 times/week: 0 = <8 lessons held 1 = 8–16 lessons held 2 = 17+ lessons held
2.2 STEM professional volunteer visits classroom	Engineering or STEM professional from the local community visits each classroom to discuss career	School	SalesForce Volunteer Tracking Software	0 = STEM professional does not visit the school during the classroom unit 1 = STEM professional visits the school 1 or more times during the classroom unit
All Indicators	Score range: 0–3 Adequate implementation score: 3			Adequate implementation at sample level: 75% of schools with score of 3

ENPOWERED GAMES

The ENpowered Games key component fidelity is measured using five indicators. Two indicators are scored at the program level and three are scored at the school level. School-level indicator scores are aggregated up to the program level before a fidelity score is given. The component was considered to have been implemented with fidelity for the school year if it achieved a total component score of 5.

Table B.3. Key Component 3: ENpowered Games

Indicators	Definition	Unit of Implementation	Data Source	Score for Level of Implementation at Unit Level	Adequate Implementation at Program Level
3.1 ENpowered Games held	PS holds the ENpowered Games competition at the end of the spring semester	Program	ISF	0 = event not held 1 = event held	Score of 1
3.2 STEM professionals volunteer at event	At least 50 volunteers from STEM professional community attend the ENpowered Games	Program	SalesForce Volunteer Tracking Software	0 = fewer than 50 volunteers attend the event 1 = 50+ volunteers attend the event	Score of 1
3.3 Students attend the ENpowered Games	Students from each school attend the ENpowered Games	School	ENpowered Games Attendance Log	0 = no students from the school attend the event 1 = at least one student from the school attends the event	0 = <80% of schools have a score of 1 1 = 80% of schools have a score of 1
3.4 Students present their engineering notebook	Students complete engineering notebook presentation	School	Judge rubric	0 = no student groups from the school complete a presentation 1 = 0–50% of student groups complete a presentation 2 = 51%+ student groups complete a presentation	0 = <80% of schools have a score of 2 1 = 80% of schools have a score of 2
3.5 Students present challenge design plans	Students present their design challenge to judges	School	Judge rubric	0 = no student groups from the school complete a presentation 1 = 0–50% of student groups complete a presentation 2 = 51%+ student groups complete a presentation	0 = <80% of schools have a score of 2 1 = 80% of schools have a score of 2
All Indicators	Score range: 0–5			Adequate implementation at sample level: score of 5	

RESULTS

Table B.4 presents the cohort-level fidelity scores for each of the three key components.

Table B.4. ENpowered Implementation Fidelity Findings

	Year 1 (SY 2021–22)	Year 2 (SY 2022–23)	Year 3 (SY 2023–24)
Key Component 1			
Percentage of schools that met adequate implementation threshold (score of 5)	17% (1 of 6)	86% (6 of 7)	67% (4 of 6)
Sample met fidelity (75% of schools achieve score of 5)	No	Yes	No
Key Component 2			
Percentage of schools that met adequate implementation threshold (score of 3)	100% (6 of 6)	100% (7 of 7)	83% (5 of 6)
Sample met fidelity (75% of schools achieve score of 3)	Yes	Yes	Yes
Key Component 3			
Program achieved adequate fidelity (score of 5)	Yes	No	Yes

Schools implemented Key Component 1 (Professional Development) with fidelity during Year Two (2022–23) of the study but fell short of implementing with fidelity during Years One and Three. During Year One (2021–22), the grantee was experiencing a staffing shortage and was unable to assign two instructors to each program class. In Year Three, the program managers completed fewer than three observations at two schools during the implementation period. Program managers met with instructors for each school at least biweekly during implementation in all implementation years.

For Key Component 2 (Classroom Programming), the program was implemented with fidelity during all three implementation years. All schools received at least 10 weeks of programming in all three years, and all but one school (during Year Three) had a STEM professional visit the ENpowered class to discuss their careers.

Regarding Key Component 3 (ENpowered Games), the grantee implemented the component to fidelity during Years One and Three, but fell short during Year Two. During 2022–23, only 31 STEM professionals volunteered to support the ENpowered Games event, falling short of the target of 50 for that indicator. Otherwise, the grantee implemented the component with fidelity in that the ENpowered Games event was held each spring and students from each study school attended the event, presented on their engineering notebooks, and completed the design challenge presentation.

APPENDIX C. IMPACT STUDY METHODS

The purpose of this appendix is to provide additional details of the impact study methods and data used to answer the confirmatory research questions. The impact study aimed to isolate the causal impact the ENpowered program had on middle school students' achievement in math and science. ENpowered was designed to increase middle school students' engagement with, and interest and achievement in science, technology, engineering, and math (STEM) subjects with the long-term aim of increasing the representativeness of traditionally underrepresented communities in STEM fields. The target population for ENpowered was middle school students from traditionally underrepresented racial and ethnic groups (i.e., Black and Hispanic/Latino/a students). The impact study was a quasi-experimental design study that compared end-of-course grades in math and science among students who participated in the ENpowered program (treatment) with those of students who are similar, but did not participate (comparison).

In this appendix, we provide additional details on outcome and covariate operationalization, analytic model specification, and the methods used to establish baseline equivalence between the treatment and comparison groups.

VARIABLE OPERATIONALIZATION

In this section, we present a description of the individual-level covariates and outcome variables used in the confirmatory impact analyses.

COVARIATES

Table C.1 provides a description of the individual-level covariates that were included in the benchmark analytic models. Covariate data were largely complete given that data were provided by Chicago Public Schools(CPS). Where data for an individual student were missing, we used dummy variable adjustment to impute missing values to the sample mean.[28]

Table C.1. Covariate Operationalization

Variable Name	Variable Type, Construction, and Data Source
Age at baseline	A continuous variable calculated by subtracting the student's date of birth from the date the program began implementing classroom programming for the school year.
Race/ethnicity	The study sample is largely homogenous regarding race and ethnicity. As a result, we specify two dummy variables representing students who were identified as Black (1) or some other racial group (0), and students who were identified as Hispanic/Latino/a (1) or Non-Hispanic (0). Although race and ethnicity are not mutually exclusive characteristics, CPS reported race and ethnicity in the same field in the data request files, and therefore we do not have race identification for students identified as Hispanic or Latino/a.
Gender	Dummy variable indicating a student's gender as female (1) or not (0). One student in the analytic sample was marked as nonbinary, and is coded as not female; all others in this category were reported as male.
Learning disadvantage	A dummy variable indicating whether the student is designated as an English Language Learner (ELL) status or an Individualized Education Plan (IEP) status (1) during the implementation year (1) or neither (0).

[28] Puma. M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Table C.1. Covariate Operationalization (Continued)

Variable Name	Variable Type, Construction, and Data Source
Free and/or reduced price lunch status	Dummy variable indicating whether the student was designated as eligible for a free or reduced price lunch (1) or not during the implementation year.
Home language other than English	Dummy variable indicating whether the student was identified as having a home language other than English (1) or not (0).
Grade level	A set of 3–1 dummy variables indicating whether the student was enrolled in a given grade level (6th, 7th, or 8th) or not.
Study cohort	A set of 3–1 dummy variables indicating whether the student was enrolled in the study during a given implementation year (1) or not (0).
School enrollment	A series of 8–1 dummy variables indicating whether the student was enrolled at a given middle school (1) or not (0). The set of dummy variables is included in the analytic model as fixed effects for the *classroom-assigned sample only*. For the school-assigned sample, we specify school membership as the third level in the multilevel regression model.

OUTCOME VARIABLES

Table C.2 outlines how outcome measures for the confirmatory Research Questions were constructed.

Table C.2. Outcome Variable Operationalization

Variable Name	Variable Type, Construction, and Data Source
Achievement in math (Research Question 1)	Count variable ranging from 0 to 4 representing the first quarter report card (baseline) or final (outcome) grade earned in the standard mathematics course completed during the implementation year. CPS provided letter grades to PRG and we re-coded these to numeric values such that: 0 = F 1 = D 2 = C 3 = B 4 = A First quarter report card grades are reported to CPS in November of the implementation year, prior to the program beginning in January. End-of-course final grades are reported to CPS in June of the implementation year, following the end of programming in May. We use the grades reported for the CPS standard mathematics course that the student was enrolled in during the implementation year (i.e., we excluded grades from elective or advanced placement math courses, such as Algebra).

Table C.2. Outcome Variable Operationalization (Continued)

Variable Name	Variable Type, Construction, and Data Source
Achievement in science (Research Question 2)	Count variable ranging from 0 to 4 representing the first quarter report card (baseline) or final (outcome) grade earned in the standard science course completed during the implementation year. CPS provided letter grades to PRG and we re-coded these to numeric values such that: 0 = F 1 = D 2 = C 3 = B 4 = A First quarter report card grades are reported to CPS in November of the implementation year, prior to the program beginning in January. End-of-course final grades are reported to CPS in June of the implementation year, following the end of programming in May. We use the grades reported for the CPS standard science course that the student was enrolled in during the implementation year (i.e., we excluded grades from elective or advanced placement science courses, such as STEM lab).

ANALYTIC APPROACH

MODEL SPECIFICATIONS

For the sake of clarity, we present the multilevel difference-in-difference (DID) model in mixed-effects form. The DID impact analytic model is a three-level model in which time-variant individual observations are nested within individual students, who are nested within either schools (school-level assignment) or classrooms (classroom-level assignment). This accounts for the clustered nature of the data. The general DID regression model specification is:

$$O_{ijt} = \alpha + \beta_1(TX_{ij}) + \theta_2(TIME_{ijt}) + \gamma_3(TX_{ij} * TIME_{ijt}) + \tau_n(X_n) + \mu_j + \varphi_{ij} + \varepsilon_{ijt}$$

where:
O_{ijk} represents the outcome of interest for participant *i* at time point *t* in cluster *j*;

TX_{ij} is the treatment group indicator (treatment = 1, comparison = 0) for individual *i* in cluster *j*;

$TIME_{ijt}$ is the pre to post indicator (where pre-exposure = 0, post-exposure = 1) for individual *i* in cluster *j*;

X_n is an n vector of individual-level covariates at baseline. We include the propensity score as an IPTW at the student level (level 2). Covariates included in the analytic model were centered at the grand mean and included:[29]

- Age
- Race/ethnicity
- Gender
- Learning disadvantage
- Free and/or reduced-price lunch status
- Cohort

[29] Operationalizations of the covariates included in the analytic model are provided in Table C.1.

- Grade level
- School fixed effects (*classroom cluster model only*)

The coefficients provide the pre- and post-program estimates where:

- α and $(\alpha + \beta_1)$ represent the baseline regression-adjusted mean of the outcome for the comparison group and the treatment group, respectively
- θ_2 represents the regression-adjusted average difference in the outcome from pre- to post-exposure for the comparison group
- γ_3 is the DID estimator and it represents the regression-adjusted average differential in pre- to post-program change for the treatment and comparison groups
- $\varepsilon, \varphi, \mu$ are the unexplained variance components. This includes the individual-level residuals, random intercept at the individual level, and random intercept at the cluster level.

Creating a Single Impact Estimate and Standard Error

After creating two separate impact estimates and standard errors for each discrete sample, we follow the guidance outlined by Price and Wolf (2024) to create a single weighted estimate for the impact coefficient and standard error.

For each model, we will obtain the treatment effect estimate, the standard error of the treatment effect estimate, and the variance of the treatment effect (as denoted below):

	Model 1 Time nested within students nested within classrooms	**Model 2** Time nested within students nested within schools
Treatment effect estimate	α_1	β_1
Standard error of treatment effect	$SE(\alpha_1)$	$SE(\beta_1)$
Variance of the treatment effect	$Var(\alpha_1) = (SE(\alpha_1))^2$	$Var(\beta_1) = (SE(\beta_1))^2$

Using the parameters outlined in the table above, we then create weights that are proportional to the inverse of the variance of the estimates:

Model 1 $$Wgt_1 = \frac{[Var(\alpha_1)]^{-1}}{[Var(\alpha_1)]^{-1} + [Var(\beta_1)]^{-1}}$$

Model 2 $$Wgt_2 = \frac{[Var(\beta_1)]^{-1}}{[Var(\alpha_1)]^{-1} + [Var(\beta_1)]^{-1}}$$

Then, calculate the combined treatment effect estimate (θ_{Trt}) as a weighted mean of the three separate estimates:

$$\theta_{Trt} = (\alpha_1 * Wgt_1) + (\beta_1 * Wgt_2)$$

To calculate the standard error of the combined treatment effect ($SE(\theta_{Trt})$), we use a standard formula for the variance of the sum of random variables:

$$Var\theta_{Trt} = [(Wgt_1)^2 Var(\alpha_1)] + [(Wgt_2)^2 Var(\beta_1)]$$

$$SE(\theta_{Trt}) = \sqrt{Var\theta_{Trt}}$$

The expected result from the combined estimate is a more precise (smaller) standard error than if we were to run a single model for the pooled analytic sample.

To get a two-tailed *p*-value from the combined treatment estimate and standard error, we calculate the *t*-statistic as:

$$tstat = \theta_{Trt}/SE(\theta_{Trt})$$

and obtain the *p*-value using the T.DIST.2T function in Excel:
T.DIST.2T(ABS(tstat), df) , where df = degrees of freedom, which is calculated as the number of units minus 2.

TREATMENT OF MISSING DATA

We did not impute missing outcome data, including baseline outcome data. Impact analyses samples will include only those observations that have non-missing Quarter 1 report card grade (baseline) and non-missing final end-of-course grade (post-program) data. We establish baseline equivalence on the analytic samples using non-missing pretest data only.

Missing covariate data were handled according to the techniques outlined by the National Center for Education Evaluation. With the assumption that data are missing at random (MAR) missing, covariate data were imputed according to guidance provided by Puma et al. (2009) using dummy variable adjustment (for details, see pp. 34–35).

CALCULATION OF EFFECT SIZE

We calculate effect size in accordance with the guidelines published in the *What Works Clearinghouse (WWC) Procedures Handbook, Version 5.0* (2022). For each confirmatory outcome, the standard deviation for each condition is estimated from the sample data. We calculate the pooled standard deviation using the following formula:

$$S_p = \sqrt{\frac{(n_t - 1)S_t^2 + (n_c - 1)S_c^2}{(n_t + n_c - 2)}}$$

where: n_t and n_c are the sample sizes, and S_t and S_c are the student-level standard deviations for the analytic treatment and control groups, respectively.

The standardized effect size, known as Hedges' *g*, is calculated using the following formula:

$$g = \frac{\theta_{Trt}}{S_p}$$

where: θ_{Trt} is the precision-weighted treatment effect estimate calculated as described in the Analytic Approach section, and S_p is the pooled standard deviation (detailed above).

BASELINE EQUIVALENCE

In calculating baseline equivalency, we calculate either standardized mean differences (continuous variables) or differences in the probability of occurrence (dichotomous variables) of the baseline outcome and covariate measures for treatment and control groups. We examine baseline equivalence for each sample separately.

For continuous variables, we follow the steps outlined above to calculate Hedges' *g*. For dichotomous variables, we follow the formula for calculating the Cox Index. Baseline equivalence of the treatment and comparison samples is established in accordance with the requirements that have been identified in the WWC *Standards and Procedures Handbook, Version 5.0* (2022). According to these guidelines, assessments of the magnitude of difference between groups are based on the following rules:

- If the standardized effect size (Hedge's *g* or Cox Index) is less than or equal to 0.05, equivalence has been established;
- If the standardized effect size is greater than 0.05 but less than or equal to 0.25 and the variable is included in the model, equivalence has been established with statistical adjustment;
- If the standardized effect size is greater than 0.25, baseline equivalence is not established.

For each of our two primary outcomes, Table C.3 presents the treatment and control group means for each characteristic and the balance statistic in the form of standardized differences (Hedges' *g* or Cox Index) for the benchmark analytic sample.

Table C.3. Baseline Equivalence of Treatment and Comparison Groups

		Treatment Group			Comparison Group			Treatment – Control	Standardized
Analytic Sample	**Baseline Measure**	***N***	**Mean**	**SD**	***N***	**Mean**	**SD**	**Difference**	**Difference**
Classroom-level sample									
Math achievement	Q1 math course grade	176	2.70	1.02	183	2.68	1.00	0.02	0.04
Science achievement	Q1 science course grade	177	3.02	0.87	184	3.02	0.96	0.00	0.02
School-level sample									
Math achievement	Q1 math course grade	248	2.71	0.94	375	2.71	1.03	0.00	–0.03
Science achievement	Q1 science course grade	217	3.00	0.81	315	3.01	0.89	–0.01	–0.06

APPENDIX D. DETAILED ANALYTIC RESULTS

Tables D.1 and D.2 provide the full output from the benchmark impact analytic models for Research Questions 1 and 2, respectively.

Table D.1. Benchmark Model Results – Research Question 1

	Classroom Sample (*n* = 718)				School Sample (*n* = 1,246)			
Variable	**Coef.**	**SE**	**z**	***p***	**Coef.**	**SE**	**z**	***p***
Treatment time interaction	0.01	0.17	0.03	0.976	0.07	0.11	0.65	0.513
Treatment indicator	0.02	0.14	0.15	0.879	–0.03	0.15	–0.23	0.819
Time indicator	–0.01	0.12	–0.07	0.943	0.04	0.05	0.87	0.386
Female	–0.02	0.11	–0.16	0.871	–0.09	0.08	–1.12	0.261
Hispanic/Latino/a	–0.33	0.22	–1.51	0.130	–	–	–	–
Black	–0.13	0.21	–0.62	0.536	0.34	0.16	2.11	0.035
Learning disadvantage	0.21	0.18	1.17	0.243	0.05	0.09	0.53	0.594
Home language	–0.27	0.08	–3.41	0.001	–0.67	0.14	–4.68	0.000
Free/reduced price lunch	0.54	0.15	3.62	0.000	0.34	0.08	4.39	0.000
Age at baseline	–0.15	0.16	–0.94	0.348	–0.02	0.13	–0.17	0.866
Cohort 1	–0.05	0.16	–0.29	0.771	–0.24	0.18	–1.38	0.169
Cohort 2	–0.01	0.13	–0.05	0.958	–0.02	0.11	–0.18	0.861
Cohort 3	(omitted)				(omitted)			
School dummy 1	0.34	0.41	0.82	0.411	–	–	–	–
School dummy 2	0.38	0.38	1.02	0.310	–	–	–	–
School dummy 3	0.65	0.15	4.39	0.000	–	–	–	–
Grade 6	(omitted)				–0.60	0.28	–2.16	0.031
Grade 7	–	–	–	–	–0.24	0.18	–1.30	0.193
Grade 8	(omitted)				(omitted)			
Intercept	2.69	0.07	38.95	0.000	2.71	0.10	26.79	0.000

Table D.2. Benchmark Model Results – Research Question 2

	Classroom Sample (n = 722)				School Sample (n = 1,064)			
Variable	**Coef.**	**SE**	**z**	***p***	**Coef.**	**SE**	**z**	***p***
Treatment time interaction	–0.03	0.10	–0.34	0.735	–0.06	0.21	–0.30	0.766
Treatment indicator	–0.02	0.07	–0.22	0.825	–0.04	0.19	–0.20	0.842
Time indicator	–0.15	0.09	–1.70	0.088	0.03	0.14	0.21	0.833
Female	–0.11	0.12	–0.88	0.378	–0.26	0.07	–3.91	0.000
Hispanic/Latino/a	–0.15	0.22	–0.67	0.502	–	–	–	–
Black	–0.03	0.25	–0.13	0.893	0.22	0.16	1.41	0.157
Learning disadvantage	0.48	0.10	4.55	0.000	0.18	0.07	2.61	0.009
Home language	–0.28	0.11	–2.63	0.008	–0.59	0.13	–4.52	0.000
Free/reduced price lunch	0.25	0.15	1.64	0.101	0.23	0.11	2.01	0.045
Age at baseline	–0.06	0.08	–0.75	0.453	0.03	0.06	0.55	0.585
Cohort 1	–0.10	0.16	–0.61	0.544	–0.33	0.12	–2.79	0.005
Cohort 2	0.10	0.08	1.25	0.211	–0.10	0.14	–0.69	0.490
Cohort 3	(omitted)				(omitted)			
School dummy 1	–0.73	0.18	–4.06	0.000	–	–	–	–
School dummy 2	0.04	0.23	0.16	0.871	–	–	–	–
School dummy 3	0.60	0.15	4.00	0.000	–	–	–	–
Grade 6	(omitted)				–0.23	0.19	–1.22	0.223
Grade 7	–	–	–	–	0.03	0.15	0.19	0.847
Grade 8	(omitted)				(omitted)			
Intercept	3.02	0.03	91.72	0.000	3.05	0.12	25.22	0.000

APPENDIX E. SENSITIVITY STUDIES

Our benchmark analytic approach was to fit two separate weighted difference-in-differences (DID) regression models for students who were enrolled in the study by way of classroom or school assignment procedures and then construct a weighted average of the two impact estimates. We used the mixed command in Stata and structured the models such that time (observations) is nested within students who are then nested within either the classroom or school clusters. We conducted a few sensitivity analyses to test the robustness of our analytic approach. In each case sensitivity results are substantively equivalent to benchmark results (no program impact is observed). Estimates for each of the sensitivity studies are presented in Tables E.1 and E.2 below.

We examined the following sensitivity models:

1. An ordinary least squares (OLS) model with cluster robust standard errors, which has fewer statistical assumptions than a random effects model and produces similar results (McNeish et al., 2017).
2. A Poisson count model, which may be a more appropriate fit for discrete count data such as grades (Hilbe, 2014). Our data do not indicate the outcome measures of math and science grades are over dispersed.
3. A fully interacted (with treatment) model that will reduce any bias from covariate adjustment in the presence of heterogeneous treatment effects and should also maximize precision (Lin, 2013).
4. A pooled model that combines the classroom- and school-assigned samples into the same multilevel model, omitting the school dummy variables used in the benchmark classroom-assigned model and where observations are nested within students who are nested within their appropriate cluster.

Table E.1. *Results of Sensitivity Models – Research Question 1*

Analytic Sample	Number Reporting	Impact Estimate	Standard Error	*p*-value	Pooled Standard Deviation	Effect Size
Benchmark model						
Classroom sample	718	0.01	0.17	0.976	0.98	0.01
School sample	1,246	0.07	0.11	0.513	0.98	0.08
OLS model						
Classroom sample	718	–0.08	0.18	0.663	0.98	–0.08
School sample	1,246	0.11	0.11	0.300	0.94	0.12
Poisson model						
Classroom sample	718	0.00	0.06	0.976	0.98	0.00
School sample	1,246	0.03	0.04	0.529	0.94	0.04
Fully interacted model						
Classroom sample	718	0.01	0.17	0.976	0.98	0.01
School sample	1,246	0.07	0.11	0.513	0.94	0.08
Pooled model						
Pooled sample	1,964	0.05	0.09	0.611	0.95	0.05

Note: ~ $p < 0.10$; *$p < 0.05$; **$p < 0.01$.

Table E.2. Results of Sensitivity Models – Research Question 2

Analytic Sample	Number Reporting	Impact Estimate	Standard Error	*p*-value	Pooled Standard Deviation	Effect Size
Benchmark model						
Classroom sample	722	–0.03	0.10	0.735	0.91	–0.04
School sample	1,064	–0.06	0.21	0.766	0.82	–0.08
OLS model						
Classroom sample	722	–0.02	0.11	0.834	0.91	–0.03
School sample	1,064	–0.02	0.21	0.913	0.82	–0.03
Poisson model						
Classroom sample	722	–0.01	0.03	0.723	0.91	0.02
School sample	1,064	–0.02	0.07	0.767	0.82	0.04
Fully interacted model						
Classroom sample	722	–0.03	0.10	0.735	0.91	–0.04
School sample	1,064	–0.06	0.21	0.766	0.82	–0.08
Pooled model						
Pooled sample	1,786	–0.05	0.13	0.699	0.86	–0.06

Note: ~ $p < 0.10$; *$p < 0.05$; **$p < 0.01$.