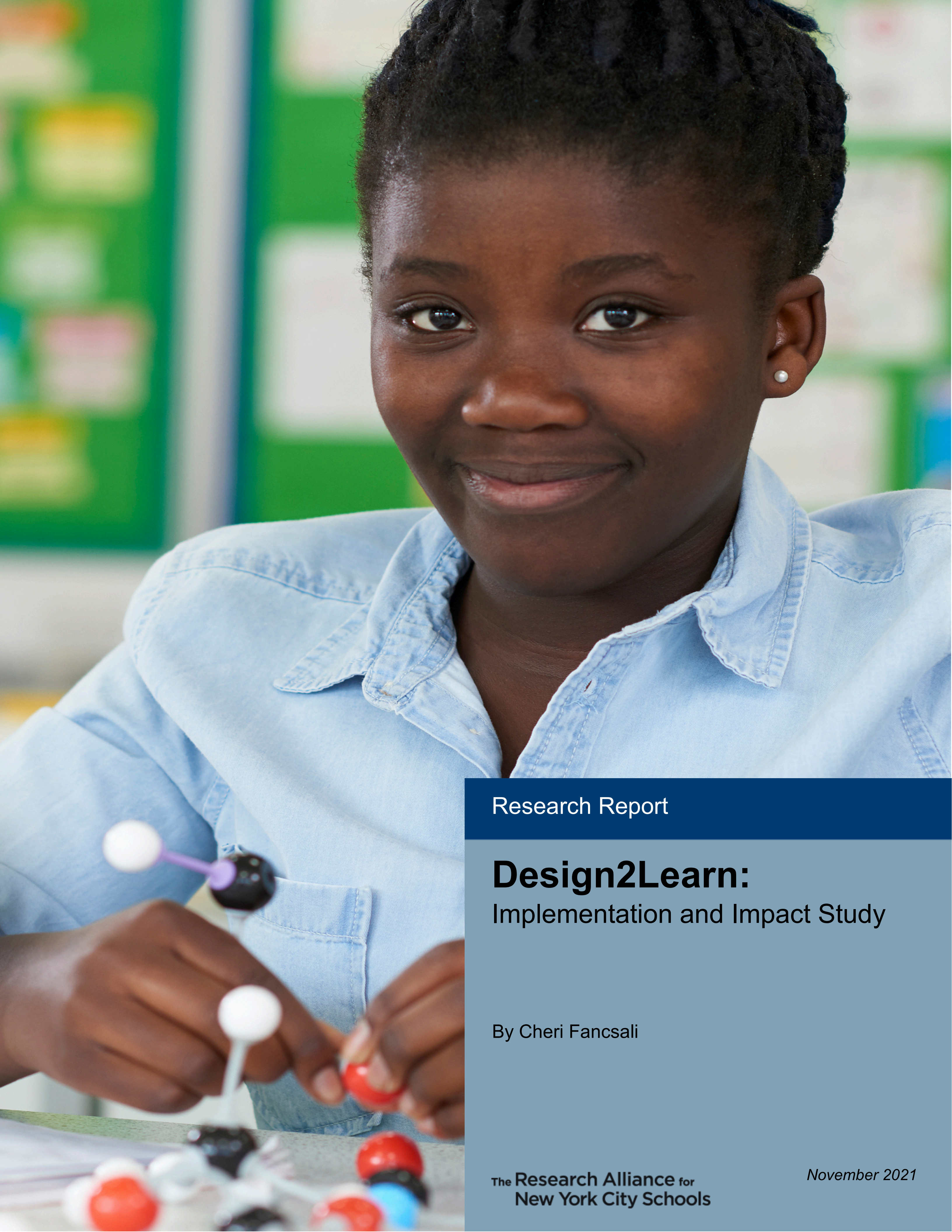


Research Report

Design2Learn:

Implementation and Impact Study

By Cheri Fancsali

The Research Alliance for New York City Schools

November 2021

About the Research Alliance

Housed at NYU Steinhardt, the Research Alliance for New York City Schools is an independent, nonpartisan research center that conducts rigorous studies on topics that matter to the City's public schools. We strive to advance excellence and equity in education by providing evidence about the policies and practices that promote students' development and academic success.

Acknowledgments

This study was supported by the U.S. Department of Education office of Elementary and Secondary Education, Investing in Innovation investing in Innovation Development Grant #U411C150068. This report represents the work of many individuals. At the Research Alliance, Ethan Crasto, Paulina Toro Isaza, Ben Schwab, Marlee Tavlin, Kristin Black, Wendy Castillo, and Rachel Cole provided analytic support and participated in data collection. James Kemple provided oversight for the research design and analyses; Chelsea Farley edited the report; and Dariana Almeyda-Vega provided copyediting and production support. Marc Moss of Abt Associates provided technical assistance through the i3 grant program. At ExpandED Schools, Emma Banay, Lizzie Murchison, Fran Agnone, Isabella Fonte, Saskia Traill, and Katie Brohawn were extremely helpful thought partners in the research. We are also deeply indebted to the staff, educators and students from the sites involved in this work. They were generous with their time and in sharing their experiences with us.

Research Alliance publications are supported by a small group of funders who underwrite our core operations, including research capacity, communications, and public engagement efforts. These funders include Carnegie Corporation of New York, the Catherine and Joseph Aresty Foundation, the New York Community Trust, the Wallace Foundation, and the William T. Grant Foundation.

Our publications reflect the findings, interpretations, and conclusions of the Research Alliance and not necessarily those of our funders or individual Steering Committee members.

Contents

Design2Learn: Implementation and Impact Study

Executive Summary

Design2Learn (D2L) is an afterschool program aimed at increasing science interest, engagement, and achievement among middle school students who historically have been underrepresented in STEM. The D2L model, designed and implemented by ExpandED Schools, uses three core strategies to foster students' interest and engagement in science: 1) collaborative teaching between an in-school science teacher and two afterschool educators, 2) curricular bridging, which connects in-school science instruction with afterschool activities, and 3) design-based learning, which emphasizes hands-on activities and inquiry-based instruction. To support implementation of the program, ExpandED Schools and its partner, the New York Hall of Science (NYSCI), provided participating educators with a week-long professional development (PD) institute in the summer, occasional Saturday workshops, strategic planning sessions, and regular on-site coaching throughout the school year. D2L was designed as a three-year intervention, serving students in 6th through 8th grade.

In 2016, the U.S. Department of Education Investing in Innovation (I3) program awarded ExpandED Schools a development grant to support the design and study of D2L. The Research Alliance for New York City Schools served as the external evaluator of the program and conducted a randomized control trial to assess the impact of D2L on key student outcomes. As part of this study, we randomly assigned 32 afterschool programs to offer D2L or be part of a control group that offered their usual science programming. Using mixed methods, we examined D2L's implementation, as well as its impact on students' interest and engagement in science, their attendance, and their science grades.

Key Findings

- Overall, study findings showed strong fidelity of implementation for the key program inputs. Of the 15 sites that implemented D2L, 14 fully participated in the PD offered, and all 15 sites participated in strategic planning and instructional and programming coaching sessions as expected.
- D2L sites implemented the three key pillars of the model—collaborative teaching, curricular bridging, and design-based instruction—at fairly high levels, and overall program quality was rated highly.
- Differences between D2L and the control site activities were small. Control sites also offered students engaging science activities, and the program quality was rated similarly. Students in control sites experienced the same amount of curricular bridging as their D2L peers, but less exposure to design-based instruction and educator collaboration.
- Both D2L and control sites offered fewer than the 72 hours of afterschool activities expected per year. On average, D2L sites offered 46 hours of activities per year, and control sites offered 63 hours per year.

- Student attendance overall was low, and D2L students attended slightly less frequently than their control peers. On average, D2L students attended about 32 hours of programming, and control students attended about 34 hours of afterschool programming. Few students participated in the afterschool program for more than one year, despite the expectation that they would attend for three consecutive years.
- D2L had no effect on students' interest and engagement in science, perseverance, or critical thinking behaviors, as measured by student surveys. It also had no effect on students' academic outcomes (science grades, school attendance).

Summary and Conclusions

The D2L model was motivated by the need to build students' science skills and interest in the middle grades—a vulnerable time when science achievement and interest tend to sharply decline, particularly for students who are underrepresented in STEM. Though the program inputs were executed as intended (ExpandED and NYSCI provided, and sites participated in, professional development and supports), and D2L sites seemed to implement the three pillars of the model as designed (design-based instruction, collaboration and co-teaching with in-school educators, and connections to the in-school curriculum), there were some substantial challenges and issues that likely influenced the results. First, our data suggest the control sites were also implementing high-quality science activities characterized by the three pillars of D2L, at least to some extent. Second, both D2L and control sites offered, and students attended, far fewer hours of activities than expected, and D2L student attendance was slightly lower than their control peers. Further, very few students participated in the afterschool programming for three years (6^{th} through 8^{th} grade) as intended. Together, these findings suggest that there may not have been a substantial difference between the D2L and control experiences, and students may not have received adequate exposure to the intervention to influence students' interest and engagement in science or science achievement.

The findings from this study raise important questions about and directions for future research on afterschool interventions designed to increase student interest and achievement in science. First, the difficulty in recruiting and retaining middle grade students for both the D2L and control sites suggests the issue was not unique to D2L or a result of the D2L offerings. Prior research indicates that attracting students in this age range to afterschool programs is a widespread problem that can only be overcome with aggressive recruitment efforts (Grossman, Raley, & Walker, 2005). Future studies should explore whether better recruitment and retention strategies could attract and retain middle grade students, resulting in more consistent and sustained participation. Future research might also explore the efficacy of shorter-term and more flexible and informal experiences in the middle grades. Additionally, these findings raise the question of whether this multi-year intervention might be better suited for elementary-aged students, whose afterschool attendance is often more reliable because of parents' needs for childcare.

Second, our findings raise questions about how much exposure to D2L activities students would have to receive to make a difference in their science interest and skills. As designed, D2L was expected to expose students to about 200 hours of designed-based activities connected to the in-school science curriculum over three years (approximately 72 hours per year). However, we found that D2L students attended for only 32 hours in a year on average, and generally for only one year. That amount totals only 15 percent of the expected exposure. Would more exposure to D2L have led to impacts on the hypothesized student outcomes? Or, is there a flaw in the theory underlying the model's design? Given the importance of science literacy to succeed in 21st century society and careers, and the imperative to address equity issues raised by the underrepresentation of students in STEM, it is essential to conduct further research to answer these questions and understand the mechanisms that lead to increased interest and achievement in science.

Chapter 1: Introduction

Background

Science literacy—the knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity (National Research Council, 1996)—is widely considered an indispensable skill in the 21st-century, essential to success in higher education and careers (Commission on Mathematics and Science Education, 2009). Yet, according to the 2015 National Assessment of Educational Progress, just over one third of 8th graders nationwide (34 percent) scored proficient in science, and fewer than one fourth of 12th graders (22 percent) did so (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics). Further, according to the most recent Program for International Student Assessment findings, the U.S. ranks 18th in science achievement, lagging behind many other nations (OECD, 2018).

To help tackle this issue, ExpandED Schools, in partnership with the New York Hall of Science (NYSCI), created Design2Learn (D2L), an afterschool program aimed at increasing science interest, engagement, and achievement among middle school students who have been historically underrepresented in STEM (especially Black, Latinx and female students). Design2Learn uses three strategies to foster students' interest and engagement in science: 1) collaborative teaching between an in-school science teacher and two afterschool educators from a local community-based organization (CBO), 2) curricular bridging, which connects in-school science instruction with afterschool activities, and 3) design-based learning, which emphasizes hands-on activities and inquiry-based instruction. D2L provides participating educators with a week-long training in the summer, occasional Saturday workshops, and regular on-site coaching throughout the school year to support their implementation of this three-year program.

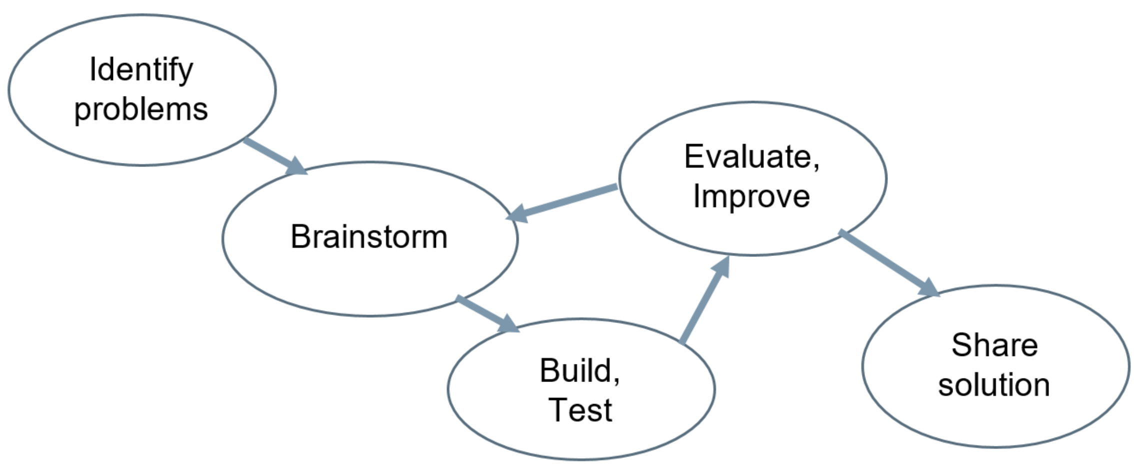
The D2L project was funded through a U.S. Department of Education Investing in Innovation (i3) development grant. ExpandED Schools engaged the Research Alliance for New York City Schools to conduct an independent evaluation of D2L, which was a requirement of the i3 grant. The evaluation was designed to both provide formative feedback about the program and to assess D2L's impact on student outcomes. This report focuses on the implementation and impact of D2L in its first three years of operation: 2016-2017, 2017-2018, and 2018-2019. We examine D2L's influence on students' science interest and engagement, as well as exploratory outcomes, such as science grades and school attendance. We also examine the extent to which D2L's program components were carried out as planned, how faithfully educators implemented key D2L program features, and how ExpandED Schools adjusted the program design to improve implementation and address programmatic challenges. Last, we draw on these findings to highlight relevant insights about future iterations of D2L, and for the afterschool field more broadly.

About Design2Learn

At the heart of the D2L model is a novel strategy for delivering high-quality afterschool science programming that combines three practices: 1) design-based learning; 2) collaborative teaching; and 3) curricular bridging. We describe each of these practices in the next section, summarizing related research that supports their effectiveness. While there is preliminary evidence that demonstrates the potential of these individual strategies, they have not previously been tested as a seamless afterschool intervention designed to improve student interest, engagement, and achievement in science. ExpandED hypothesized that these strategies are mutually reinforcing and more powerful when offered together. D2L was designed to provide a framework for delivering engaging, relevant science experiences to middle school youth that reinforce in-school learning, take advantage of the flexibility afforded by the afterschool setting, and leverage the strength of CBO program staff who often have strong youth development backgrounds.

Design-based learning frames instruction around a process, often referred to as a "design challenge," where students identify a problem; iteratively brainstorm, build, and evaluate solutions to the problem; and then share their final solution (see Figure 1). As students engage in these hands-on, real-world tasks, higher-order thinking skills become more tangible and visible.

Figure 1: Design Process



Source: New York Hall of Science

Design-based activities are intrinsically motivating to students because they leverage the desire to create, solve problems, and learn how things work. Children engaged through design-based activities, particularly those who do not believe they are good at science, become more motivated because they witness how core science content can help solve problems that are personally relevant (Bennett, 2013). Promising evidence exists for both academic and non-academic outcomes of design-based learning (Capobianco et al., 2014; Han et al., 2015; Harris et al., 2015; Korur et al., 2017; NASEM, 2021; Saritepeci, 2020; Vaino et al., 2018; Vongkulluksn et al., 2018).

In the D2L model, **collaborative teaching** is defined as certified teachers and informal educators (i.e., staff employed by nonprofit organizations providing school-based after-school programs) jointly delivering instruction during out-of-school time. This approach is informed by research demonstrating the effectiveness of co-teaching by two or more certified teachers (e.g., a general education teacher and a reading specialist) during the school day (Friend & Cook, 2010). Co-teaching provides more opportunity for teacher-student interaction, increases differentiated instruction, and serves as a predictor of student performance, suggesting that optimal learning occurs when educators collaborate (Moorehead and Grillo, 2013; Zito, 2011). The D2L model builds on this powerful approach by incorporating the expertise of CBOs delivering after-school programming as informal educators who are trained in mentoring and peer-assisted learning strategies. Informal educators contribute specialized competencies in culturally sensitive social work and positive youth development practice (Krishnamurti, Ballard, and Noam, 2014).

Curricular bridging is defined as the alignment between school and after-school curricula to create seamless student learning experiences (Noam, 2003). According to Noam, Biancarosa, and Dechausay (2002), congruity of environments, learning goals, and teaching styles is associated with increased student performance. Curricular bridging also shows promise as a strategy for engaging students. Out-of-school time programs that include links to academic curricular goals connect children’s “divergent worlds,” making learning “more meaningful and relevant to [students’] life experience” (Noam, 2003).

Table 1, on the next page, shows D2L’s key program inputs, the short- and longer-term student outcomes the program is designed to affect, and the mediators hypothesized to connect the inputs to the short- and longer-term outcomes.

Table 1: Design2Learn Logic Model

Key Components (Inputs/Activities)	Educator Mediators	Student Outcomes	
		Short-term	***Long-term***
Professional Development • *Summer institute*: 5 days (30 hours per year). • *School-year PD workshops*: 2 Saturdays per year (one in the fall and one the spring). ***Strategic Planning Retreats*** • ExpandED-facilitated retreats: 2 hours per site per year (fall). ***Instructional Coaching*** • *ExpandED*: monthly site visits, 3 hours each visit, up to 8 per year. • *NYSCI*: 2 site visits per year, 3 hours each visit (fall and spring). ***Programmatic Coaching*** • ExpandED conducts an assessment of support needs and creates site-specific plans once per semester.	***Educator Practices in Afterschool*** • **Design-Based Instruction:** Educators deliver design-based afterschool science instruction: 72 hours per year (1 hour per day, 2 days per week for 36 weeks). • **Collaboration:** Site teams (in-school science teacher and two afterschool educators) participate in collaborative site-based planning: 36 hours per year (1 hour per week for 36 weeks). Afterschool instruction is collaboratively taught by in-school and afterschool educators (1 hour per week for 36 weeks). • **Curricular Bridging:** Educators bridge in-school and afterschool content. ***Educator Attitudes and Knowledge*** • Educators increase confidence in delivering inquiry-based afterschool science instruction. • Educators increase knowledge about how to deliver inquiry-based science instruction. • Educators increase knowledge about how to deliver standards-based science instruction. • Educators believe in efficacy of D2L model.	Increased: • Science interest and engagement (primary outcome of interest). Secondary outcomes: • Science identity. • Science career interest. • Science career knowledge. • Science enjoyment. • Participation in science activities (e.g., reading books about science, watching science-related TV shows). • Attendance in afterschool program. • Attendance in school.	Increased: • Science achievement test scores. • Science course grades.

Inputs: Professional Development, Retreats, and Coaching

As seen in Table 1, the Design2Learn model included four key program components: 1) professional development, 2) strategic planning retreats, 3) instructional coaching, and 4) programmatic coaching. ExpandED Schools and NYSCI delivered a five-day (30 hours) PD summer institute and two Saturday workshops during the school year to teams of two afterschool educators and an in-school science teacher from each site. The summer institute provided foundational knowledge of the model, including information about how to develop and implement design-based instruction that aligns with the New York City Department of Education's Science Scope & Sequence and the Next Generation Science Standards Science and Engineering Practices. In the second year of implementation, ExpandED Schools provided additional curricular resources and materials in response to educators' requests, including a "toolkit" of 18 lessons. Educator teams experienced and facilitated a number of design-based activities and developed their own lesson plans, which incorporated engineering design tasks that align to the standards covered during the institute. In-school teacher and CBO educator teams were expected to attend together to allow for site-level planning and to facilitate collaboration and team building. The summer institute was repeated in the summer of 2017 and 2018 for both new sites and returning sites.

The Saturday workshops (6 hours each) were held at NYSCI and facilitated by NYSCI and ExpandED Schools staff, in order for educator teams to continue the learning begun during the summer institute. Educators focused on strategies for planning and delivering afterschool activities that bridged the Enhanced NYC Science Scope & Sequence and the elements of design-based instruction (e.g., identifying problems, asking thoughtful questions, designing solutions).

At the beginning of each school year, ExpandED Schools facilitated a two-hour strategic planning retreat at each site, which served as an opportunity for leaders from the school and the partnering CBO to work with the site's educator team to plan for the year. This included developing program recruitment and curricular bridging strategies, discussing co-teaching responsibilities, and setting up weekly planning meetings for the educators. In addition, in Years 2 and 3, ExpandED conducted an assessment of support needs and created a support plan for each site, once per semester. This allowed ExpandED to customize and differentiate support based on sites' specific needs.

Finally, ExpandED and NYSCI staff conducted visits to each site to provide instructional and programmatic coaching and technical assistance related to implementing the three core practices of the D2L model. During the visits, ExpandED and NYSCI staff selected observation focus areas with participants, observed the afterschool activities, and facilitated a post-observation conference with the afterschool educators. In each site's first year of implementation, ExpandED conducted eight visits over the course of the school year, and NYSCI conducted two visits. In subsequent years of implementation, ExpandED and NYSCI each conducted two visits per site per year.

Mediators: Educator Practices, Knowledge, and Attitudes

The mediating factors that were expected to link the program inputs to outcomes included changes in educator practices, knowledge, and attitudes. As noted in Table 1, D2L educators were expected to offer approximately 72 hours of design-based learning opportunities in their afterschool sessions per year. Educators and sites were allowed flexibility in how they scheduled the activities. For example, they could offer one hour of D2L activities per day, two days per week for 36 weeks, or they could offer D2L activities once per week, for a longer block of time (e.g., two hours). In terms of collaboration, the in-school teacher and two afterschool educators were expected to jointly deliver afterschool instruction for at least one hour per week. Further, each educator team was asked to meet for at least one hour per week to plan for the collaborative teaching, with D2L compensating them for this time. To enact curricular bridging, the educator teams were expected to gather information about the scope and sequence of students' in-school science classes and align the afterschool curriculum with those in-school classes, to create a seamless learning experience for students. D2L aimed to recruit a science teacher from the afterschool program's host school to participate in the D2L educator team, ideally someone teaching the target grades (6^{th} grade in Year 1, 7^{th} grade in Year 2, 8^{th} grade in Year 3), to help facilitate this alignment.[1]

D2L PD and other supports focused on helping educators acquire the knowledge and attitudes needed to implement the three pillars of the D2L model. Specifically, the PD and support were designed to increase educators' confidence in delivering inquiry-based afterschool science instruction and their knowledge about how to deliver inquiry- and standards-based science instruction. Educators were also expected to learn strategies for collaborating and connecting the afterschool activities to the in-school science curriculum. Taken together, these changes were expected to promote educators' belief in efficacy of D2L model.

The ultimate aim of D2L was to combine the strategies above into a seamless intervention with measurable impacts on student interest, engagement, and achievement in science.

Short- and Long-Term Student Outcomes

D2L hypothesized that the program would lead to improved knowledge and attitudes about science among students, as well as improved academic outcomes. In the short-term, students who participated in D2L were expected to show increases in science interest and engagement, science identity, science career interest, science career knowledge, science enjoyment, participation in science activities (e.g., reading books about science, watching science-related TV shows), and attendance in the afterschool program.

Over the long-term, the program hypothesized that students would show improved academic performance in science. This would include increased science achievement as measured by students' scores on the New York State 8th-grade science assessment, as well as improved grades in science courses.

Organization of this Report

The remainder of this report includes a description of the study design, an overview of our implementation and impact findings, and a discussion of implications for practice and future research. Chapter 2 presents details on the evaluation's study design, methodology, and sample, including participant eligibility, recruitment, and assignment, and baseline characteristics. Chapter 3 describes D2L's implementation, including measures of implementation fidelity and student attendance in the afterschool programs. Chapter 4 describes findings on the impact of D2L on student outcomes. The final chapter reflects on these findings to offer recommendations for future implementation of D2L and other science-focused afterschool programs.

Chapter 2: Study Design, Research Questions, and Data Sources

Evaluation Design

To assess the implementation and impact of the Design2Learn model, the Research Alliance conducted a three-year randomized controlled trial. ExpandED Schools recruited sites for the study by drawing on its existing network of programs targeting high-needs schools with high proportions of Black, Latinx, and low-income students. ExpandED Schools invited CBOs delivering afterschool programming in New York City public schools serving 6th-8th graders to apply to be part of the study. To be eligible, sites had to meet the following conditions: be publicly funded, serve at least 25 6th grade students in the afterschool program in 2016-2017, and have the commitment of a team of two afterschool educators and an in-school science teacher from the host school to participate. Finally, sites had to agree to be randomly assigned to either the D2L group or the control group, participate in the study for the full three school years, and take part in data collection and PD activities.

Prior to random assignment, we matched sites that were similar on key demographic factors, including eligibility for free or reduced-price lunch, prior ELA and mathematics state test achievement levels, and racial/ethnic makeup of 6th grade students served in the sites' host school. Matched pairs were then randomly assigned to treatment or control conditions prior to the first PD session. Matching sites on key characteristics before random assignment helps reduce the chance of accidental bias due to unequal treatment and control groups.

ExpandED Schools provided PD and coaching supports described above, plus $12,400 per year to cover educator time and materials to sites randomized to receive the D2L "treatment." They provided control sites $2,000 to cover less extensive outlays of educator time and materials, and offered all control sites the opportunity to participate in other, non-STEM focused PD offered by ExpandED Schools.

At the beginning of the study, we asked sites to recruit at least 25 students to participate in afterschool activities. In reality, sites struggled to recruit students, and many recruited far less than 25. Conversely, some control sites served far more than 25 students in their STEM programming. For recruitment, we asked both D2L and control group sites to use a consistent "script" that described the afterschool programming and activities being offered. This script mitigated the chance that students in the treatment group were attracted to the unique features of the D2L program and differed from the control group at the outset, for example in terms of their motivation or level of interest and engagement in science.

Research Questions

The evaluation addressed research questions related to the implementation and impact of Design2Learn. In terms of implementation, we explored the following questions:

- To what extent were the key components of the model implemented with fidelity?
- What was the overall level of programming quality as assessed by the Dimensions of Success standardized assessment for after-school programs?
- How were the experiences of treatment and control students different (i.e., what was the contrast between D2L and control conditions)?

Impact questions included one "confirmatory" research question—that is, the primary research question that we were trying to answer about the effectiveness of D2L for students. Our confirmatory research question was:

- What is the impact of one year of exposure to D2L on students' interest and engagement in science compared to the alternative STEM programs offered in the control sites?

In addition, we conducted several exploratory analyses to investigate secondary questions, including the following.

- What is the impact of one year of D2L on students' science grades and school attendance compared to the alternative STEM programs offered in the control sites?
- What is the impact of D2L on key subgroups of students, including girls, English language learners, students with disabilities, and Black and Latinx students?
- What is the relationship between site fidelity of implementation (or strength of implementation) and program impacts on students? Do sites with stronger implementation see larger impacts?

Data Sources

The D2L evaluation used a variety of data sources to measure program implementation and educator and student outcomes. The Research Alliance collected student and educator survey data, observed D2L PD sessions, and in a sample of sites, conducted activity observations and educator interviews. We also used school records data to assess student characteristics, prior performance, science grades, and school attendance. Additionally, we conducted secondary analyses on afterschool attendance data, as well as data from observations of D2L and control sites conducted by ExpandED Schools staff using the Dimensions of Success (DOS) observation tool.[2] Following is an overview of the data sources and measures used in this study. Additional details about data sources and response rates are described in the appendices.

Student Surveys

We administered surveys to students in both the treatment and control groups, in the fall and spring of the 2016-2017, 2017-2018, and 2018-2019 school years. This allowed us to assess growth in their interest and engagement in science during the course of the year. The spring survey also measured other attitudes toward science (e.g., the value and importance of science), participation in science activities, and students' reaction to their afterschool program.

The pre and post surveys were largely designed around the Partnerships in Education and Resilience's (PEAR Institute's) Common Instrument, a self-report survey that measures a variety of STEM-related attitudes, including 21st Century skills that are correlated with STEM interest and engagement (PEAR Institute, 2019). It was developed for informal settings like afterschool programs and aims to assess the impacts of those settings on student attitudes. In addition to the Common Instrument items, we created several survey questions specifically for this study.

We tested the reliability of the survey sub-scales and found them to have strong internal reliability (Cronbach's alphas ranged from .67 to .96, with all but one scale well above the industry standard of .70). For all students who answered at least 50 percent of the items within a sub-scale, we computed a scale score by converting responses to a numeral (e.g., 1= strongly disagree, 4= strongly agree) and averaging all items together. See the Appendix for response rates by number of years of exposure.

Educator Surveys

We administered surveys to in-school teachers and afterschool educators in both D2L and control sites in May of each school year. These surveys were designed to measure educators' confidence in their ability to implement afterschool STEM programming, the frequency with which they engaged instructional activities aligned with D2L's logic model, as well as challenges and successes in implementation. For control sites, we modified the surveys slightly by removing items on collaboration between in-school teachers and afterschool educators. Items on both educator surveys were drawn from surveys previously developed by ExpandED and the Research Alliance for other studies, with some additional items created by the Research Alliance specifically for this study.

Dimension of Success (DOS) Observation

ExpandED Schools staff conducted observations at D2L and control sites using the DOS observation tool twice per year as part of their regular programming. The DOS was developed by the PEAR Institute as a framework that defines key aspects of a quality STEM learning experience (Shah et al., 2018). ExpandED Schools shared site-level observation data with the Research Alliance for use in our assessment of D2L implementation.

Program Attendance Data

The Research Alliance obtained afterschool attendance data collected by ExpandED Schools as part of their program records. Because sites had rolling admissions, we used the attendance data to identify students in the sample. Any student who attended at least one hour of D2L or control programming was considered eligible to participate in the study.

Student Records Data

Drawing on the Research Alliance's archive of NYC Department of Education student records data, we assessed student characteristics (race/ethnicity, gender, socio-economic status) and prior academic performance (math and ELA test scores) for the D2L and control groups. We also used records data to measure students' in-school attendance and science class grades.

Site Visits

The Research Alliance conducted yearly site visits at eight participating sites (four D2L sites and four control sites). These visits included observations of afterschool activities, and teacher and educator interviews. ExpandED Schools provided a list of D2L sites divided into those deemed as having strong and weaker implementation, based on previous observations. We invited two stronger and two weaker D2L sites, and their matched control site pairs, to participate in the observations.

Observations: We conducted observations of the afterschool STEM programming in both D2L and control sites to gather descriptive evidence about what D2L looks like in practice and to describe the difference between D2L and business-as-usual STEM activities. Each observation was approximately 90 minutes long and was conducted using protocols based on previously created afterschool and pedagogical observation tools, including the Youth Program Quality Assessment (PQA) tool (Akiva, 2005), the Out-of-School Time (OST) observation instrument (Pechman et al, 2008), and the Classroom Assessment Scoring System (CLASS, Pianta et al, 2008). The observation protocol we used included items on student engagement, design-based instruction, classroom management, relationships (staff-to-student; student-to-student), relevance to students' lives, connections to STEM concepts, student collaboration, curricular bridging, and the extent to which activities included Next Generation Science Standards practices.

Teacher and Educator Interviews: During the site visits, we also conducted interviews with educators to assess their perspectives on implementing STEM afterschool programs and to illuminate differences between D2L and control programs. In each of the sites, we conducted between one and four interviews, depending on the number of educators present on the day of our site visit. Each interview lasted approximately 45 to 60 minutes. For both D2L and control sites, our protocol included questions concerning educator background, collaboration, design-based instruction, curriculum bridging, student reactions, and successes and challenges. For D2L sites, we included further questions about the educators' impressions of D2L and requests for further supports in future years.

Site and Student Characteristics

Site Characteristics

In total, 32 afterschool programs applied to and were eligible to be in the study: 18 in 2016 (Wave 1) and an additional 14 in 2017 (Wave 2). For both Wave 1 sites (which began in the summer of 2016) and Wave 2 sites (which began in the summer of 2017), we created matched pairs based on key demographic factors, such as the socio-economic status, prior achievement levels, and racial/ethnic makeup of students served by the afterschool programs' host school. Within each matched pair, one site was randomly assigned to implement D2L, while the other site would implement "business–as-usual" STEM programming (i.e., the alternative STEM activities offered in the control sites). One control site dropped out of the study after random assignment but before implementation began in the first year. (See the Appendix for site attrition and response rates by study year.)

CBOs ran the afterschool programs at each site. ExpandED asked each CBO (both treatment and control sites) to identify a team of two afterschool educators and an in-school science teacher to be part of the study. Participating sites were located in schools across four of the five NYC boroughs: Brooklyn (13), Manhattan (5), Queens (6), and the Bronx (8). Schools represented 18 community school districts and two charter schools. Compared with other NYC schools serving students in grades six through eight, students in our study sample were similar in terms of the gender breakdown, the percent of Asian and White students, and the percent of students eligible for free/reduced priced lunch. The schools in our study were less likely to serve Black students, and more likely to serve Latinx students than middle schools districtwide. (See the Appendix for more information.)

Student Characteristics

D2L was designed to target middle grade students who have historically been underrepresented in science. As such, in recruiting sites to participate in the study, ExpandED sought out programs from the highest-needs districts in NYC and whose student population was predominately Black, Latinx, and low income. As shown in Table 2 on the next page, students in our analytic sample for the confirmatory impact analysis largely reflected the population of interest. Specifically, the majority of students in both treatment and control sites were Black or Latinx, and two thirds were eligible for free or reduced-price lunch.

In total, our analytic sample includes 1,125 students with at least one hour of attendance in any of the three years of implementation, a post-program survey, and school record data (510 treatment, 615 control) from 31 sites. As shown in Table 2, none of the differences in the background characteristics of D2L and control students were statistically significant, indicating the randomization worked as expected, and the treatment and control samples were equivalent.

In order to identify background characteristics to control for in our impact analyses, we also looked at the effect size (in standard deviations) of the treatment and control group differences. All of the D2L and control group differences at baseline were less than .25 standard deviations (SD). Differences of .25 SD or greater at the start of the intervention could bias the estimated impact of the intervention and are considered “nonequivalent” by What Works Clearinghouse standards (IES, 2020).

Differences by race/ethnicity, gender, grade level, eligibility for free/reduced priced lunch and English language services, and prior math and ELA achievement were less than .1 SD. Differences in eligibility for special education services were slightly larger (.15 SD), with treatment students less likely to have an Individualized Education Plan than their control peers. Students in the D2L sites also had slightly higher interest and engagement in science at baseline, as measured by the pre survey (.15 SD) administered at the beginning of each school year. Using WWC standards, D2L and control group differences at baseline that were greater than .05 standard deviations were adjusted for statistically in the impact model.

Table 2: Demographic Characteristics of D2L Treatment and Control Students

Characteristics		D2L Students	Control Students	Estimate Difference	P-Value	SD
Race/Ethnicity (%)						
	Asian	13.1	12.8	0.3	1.0	0.01
	Black	30.0	33.2	-3.2	0.8	-0.07
	Latinx	37.8	35.3	2.5	0.8	0.05
	Other	0.1	0.1	0.0	0.5	-0.09
	White	7.8	4.7	3.2	0.5	0.11
Age (average)		12.3	12.4	-0.1	0.5	-0.12
Female		47.9	46.6	1.3	0.8	0.03
Grade Level (average)		6.5	6.6	-0.1	0.6	-0.10
Services (%)						
	English Language Services	9.8	8.0	1.8	0.6	0.06
	Free/Reduced Price Lunch	68.8	70.5	-1.7	0.8	-0.04
	Special Education Services	16.1	22.0	-6.0	0.1	-0.15
Prior Achievement (Z-score)						
	Prior ELA Achievement	0.0	-0.1	0.1	0.5	0.13
	Prior Math Achievement	0.0	-0.1	0.1	0.5	0.15
Baseline Science Interest and Engagement (scale of 1-4)		2.9	2.9	0.1	0.3	0.15
Number of Sites		16	15			
Number of Students		510	615			

Source: Research Alliance calculations based on the D2L pre and post student surveys and school records data obtained from the NYC Department of Education.

Notes: Differences between D2L and control students are regression adjusted. Missing baseline interest and engagement, prior ELA and prior Math were imputed using the site average for students at the same age and grade level. Due to rounding, the baseline science interest and engagement scores appear to be the same for D2L and control students (D2L = 2.94, control = 2.86). P-value is showing the probability that the difference between D2L and control is by chance. P-values of less than .05 (5% probability) are generally considered statistically significant. None of the baseline measures we explored showed statistically significant differences between D2L and control students.

Chapter 3: Design2Learn Implementation

Assessing implementation of D2L is key to understanding the results of the impact analyses. Drawing on multiple data sources, we assessed implementation in two ways: the extent to which the model was implemented as designed (fidelity of implementation), and the extent to which the D2L and control conditions and experiences were different (treatment/control contrast).

Fidelity of Implementation

To assess the extent to which D2L was implemented as designed, we looked across multiple sources of data to determine if the key program inputs (e.g., resources, materials, PD) were delivered and the intended recipients of those inputs participated. We also examined the extent to which the D2L programming—the activities that students' experienced—reflected the model as specified in the theory of change. Findings for each are described below.

Program Inputs and Participation

Our evaluation measured fidelity of implementation (FOI) for four key program inputs: professional development, strategic planning, instructional coaching, and programmatic coaching. We assessed participation in PD (including the summer institute and Saturday workshops) and strategic planning sessions using the attendance rosters collected by the ExpandED Schools and NYSCI professional development facilitators. We measured instructional and programmatic coaching through reports completed by facilitators after each day of coaching. Each site was scored on their adherence to the FOI criteria for the four components. Site scores ranged from 0 to a total possible score of 16.

For each of the four components, we worked with ExpandED Schools to identify thresholds for the minimum level of implementation necessary to reasonably expect the hypothesized outcomes. ExpandED Schools drew on their experience supporting afterschool programming to make assumptions, such as how many team members would need to be present and for how many PD days, in order for each educator team to receive enough training to implement D2L with at least a basic level of fidelity to the model. Some assumptions related closely to the particulars of the D2L model; for example, given the emphasis on educator collaboration, FOI benchmarks required multiple team members to be present at the same PD training in order to bond and co-plan. In consultation with ExpandED Schools, we decided that the *program* achieved implementation fidelity if two thirds of D2L sites met the thresholds on all components.

Overall, these data showed strong fidelity of implementation for these components. Of the 15 sites that implemented D2L, 14 met the fidelity threshold for participation in PD, and all 15 sites met the thresholds for participating in strategic planning and instructional and programming coaching sessions (see Table 3 below).

Table 3: Fidelity of Implementation Results

Component	Site Level Threshold	Scoring	Number of Sites Meeting Threshold
1. Professional Development			
a. Summer institute PD for site and afterschool educator teams	2/3 of teachers from the site attended at least 3/5 days.	0=less than 2/3 of teachers from site attended at least 3/5 days. 1=2/3 of teachers from site attended at least 3/5 days.	15/15
b. Saturday PD workshops (two Saturdays/year)	2/3 of teachers from the site attended at least 1/2 days.	0=less than 2/3 of teachers from site attended at least 1/2 days. 1=2/3 of teachers from site attended at least 1/2 days.	14/15
c. Summer institute: full team attendance	2/3 of teachers from the site attended summer institute.	0=less than 2/3 of teachers from site attended summer institute. 1=2/3 of teachers from site attended summer institute. 2=All teachers from site attended summer institute.	15/15
d. Saturday PD workshops: full team attendance	All team members attended at least of the 2 workshops together.	0=team members did not attend fall or spring workshop together. 1=at least 2 team members attended 1 of the 2 workshops together. 2=team members attended both workshops together.	12/15
2. Planning			
a. Strategic Planning Retreats	Either team members or site leadership attended the planning retreat.	0=only educators attended the planning retreat. 1=educators and site leadership attended the planning retreat 2=educators, site leadership and school leadership attended the planning retreat together.	15/15
3. Instructional Coaching			
a. Instructional Coaching (ExpandED): TA assistance/ site-based coaching monthly	Sites receive at least 4 days of coaching in the first year and at least 1 day of coaching in subsequent years.	0=0-3 days of coaching provided (first year, 0 days in years 2 & 3). 1=4-7 days of coaching provided (first year, 1 day in years 2 & 3). 2=8 days of coaching provided (first year, 2 days in years 2 & 3).	15/15
b. Instructional Coaching (NYSCI): TA assistance/ site-based coaching twice per year	Sites received at least 1 day of coaching.	0=0 days of coaching. 1=1 day of coaching. 2=2 days of coaching.	14/15
4. Programmatic Coaching ExpandED conducts an assessment of support needs and creates a site-specific support plan once per semester	Sites receive at least one assessment and support plan.	0=no assessment/support plan. 1=1 assessment/support plan. 2=assessment and support plan updated each semester.	15/15

D2L Programming

To explore the extent to which D2L programming reflected the key aspects of the model, we drew on data from student and educator follow-up surveys administered at the end of each program year. The surveys asked about respondents' experiences related to curricular bridging, design-based instruction, and educator collaboration. We also investigated the implementation of these three components in our interviews with educators. Specifically, we asked educators to comment on how they enacted the three pillars, and the challenges and facilitators to doing so. We also looked at program quality ratings collected by ExpandED observers. The observation instrument is in alignment with D2L's vision of quality pedagogy for STEM afterschool programs, including aspects of design-based learning, such as inquiry and youth voice. Finally, we looked at the number of hours of D2L activities sites offered, and the rate at which students attended those activities.

Curricular Bridging

Students were asked three items on the year-end survey about the extent to which activities were connected to the science topics and instruction they received in school—an indication of curricular bridging. Looking across all curricular bridging survey items, D2L students reported fairly frequent curricular bridging. Responding on a four-point scale (1=not at all, 2=sometimes, 3=often, 4=almost always), D2L students had an average scale score of 2.9 (just below "often"), as shown in the "D2L" column in Table 4 below.[3] This measure included statements like, "The Afterschool activity leaders in this program talked about things we were learning in school"; 60 percent of D2L students said this happened "often" or "almost always".

Our educator interview data revealed different ways in which teachers and educators approached curriculum bridging. Some educators started with an activity they wanted to implement and aligned it to the Next Generation Science Standards or the school's science curriculum; others started with topics covered in the in-school science curriculum and then developed afterschool activities to complement them.

One D2L educator explained how her site approached curricular bridging as follows:

> **As soon as we have an idea, we scan through the standards, and we see which concepts it covers. Or, we look at a project the science teacher might be doing and say, "Okay, what's the important idea we want them to walk away from this with for next year," and then we would go into the concept and try to build it from there, from the curriculum. Sometimes, it was teacher-driven, and sometimes, it was right off the paper of the curriculum, but it always meshed up together nicely.**

Another D2L educator reported that she began by consulting the in-school science scope and sequence:

> **That's how we plan our lessons…Based off what they're doing during day school, we most likely give them the hands-on version to give them a better understanding of what they're doing during day school.**

The interview data also showed that afterschool educators relied on the participation of the classroom teacher to guide bridging. Classroom teachers were familiar with science standards and curriculum, but also with the day-to-day changes in curricular plans. In sites where the in-school teacher taught a different grade or group of students than those in the afterschool program, bridging was more challenging. For example, one D2L site where this was the case reported that bridging was difficult because science standards had changed recently (and the afterschool educators were not familiar with the changes). Further, some sites said curricular bridging was challenging because it felt like too much of a continuation of the school day to students, leading to a lack of interest. Another interviewee reported that she had trouble finding engaging hands-on activities that were relevant to the in-school topic of weather. Along similar lines, an educator at a different site said it was difficult to find resources that matched both the topics in the curriculum and the design-based pedagogy that D2L promotes.

Design-Based Instruction

Design-based instruction uses hands-on and real-world activities, inquiry, and projects to engage students in substantive science learning. In a year-end survey, we asked students to report the frequency with which they experienced activities indicative of design-based instruction in their afterschool program, such as hands-on instruction and working together to solve problems. Responding on the same four-point scale (1=not at all, 2=sometimes, 3=often, 4=almost always), D2L students reported that they experienced design-based instruction “often”, with an average scale score of 3.0 across the five items (again, see the “D2L” column Table 4). For example, in response to the statement, “We did hands-on science activities”, 77 percent of D2L students said this happened “often” or “almost always.”

In an example of design-based instruction that we observed, students were asked to design their own spider webs, first by drawing one out and then by using craft materials to create a three-dimensional model of the web. The activity, developed by ExpandED, focused on animal habitats and defense mechanisms. As part of the activity, the in-school educator prompted students to think about their spider web designs by posing questions such as: “Imagine you’re a spider—what adaptations could you come up with to better catch prey than competitors?” “What might its web-making behavior help it do over the other spiders?” As described by one D2L educator, reflection and engaging in scientific inquiry were key to design-based instruction:

> **At the end [of the project], they present [their design], and they reflect. We don’t ever end without reflecting on what have you done, because that is what scientists do; they go back to the board. Sometimes, you start with an idea, and at the end it didn’t come out the way you want, and that’s okay. We’ll try again the next time, and we don’t want to get things right the first time around, and so on. That’s all a part of the process.**

In our interviews, educators reported that the primary challenges to implementing design-based activities were acquiring specific materials, finding the time to prepare for the activities, and having enough time in the afterschool sessions to allow for prolonged investigations (e.g., projects that took multiple sessions to complete).

Table 4: Differences Between D2L and Control Students on Design-Based Instruction and Curricular Bridging Implementation

	D2L Students	Control Students	Estimated Difference		P-Value	Effect Size
Curricular Bridging	2.9	2.9	-0.1		0.4	-0.1
Design-Based Instruction	3.0	2.8	0.3	*	0.0	0.3
Number of students	492	600				
Number of sites	16	15				

Source(s): Research Alliance calculations based on data obtained from the D2L student survey.

Note: Differences for each measure between D2L and control sites are regression adjusted to account for differences associated with student demographic characteristics and baseline interest and engagement. The analysis also accounts for the clustering of students within sites. * Differences were statistically significant at p <.01, using a two-tailed t-test.

Educator Collaboration

The last key feature of D2L was educator collaboration and co-facilitation of the activities. According to the model, educators were to have a dedicated planning time of at least one hour per week (for which they were paid). Further, it was expected that the in-school science teacher would co-facilitate at least one session per week. In the end-of-year survey, we asked D2L educators about this collaboration. As shown in Table 5 on the next page, most D2L educators met program expectations; they reported meeting with their colleagues at least once per week to plan (83%) and co-teach the activities (97%).

Interviewed educators reported very few challenges related to the collaboration process. One exception was time. Several participants stated that—even though D2L compensated teachers for their time—it was sometimes difficult to find the extra hour to meet. One D2L educator commented in an interview, “You’ve been there all day, and now it’s 6:00, and it’s like, I can’t think about STEM anymore. Not having enough time to plan or just trying to do too much in one day—we were just over-exhausted from everything.”

Program Quality

A final source of information on students’ experiences is from the DOS observation data. The DOS includes elements of program quality that fall in four broad domains: Features of the Learning Environment, Activity Engagement, STEM Knowledge and Practices, and Youth Development in STEM. This tool was selected because ExpandED Schools uses it with all of the sites in its network to measure program quality and provide formative feedback.

Twice per year, ExpandED staff visited each site and observed the afterschool programming using the DOS structured observation protocol. The activity was rated on each of 12 quality dimensions, using a four-point rubric for each dimension. A score of 1 indicated there was no evidence of the dimension. A score of 2 indicated there was “inconsistent evidence” of the quality dimension, a score of 3 indicated “reasonable evidence,” and a score of 4 indicated “compelling evidence” of the quality dimension.

According to the developers of the DOS protocol, scores of 3 or 4 indicate high program quality. We averaged scores across all 12 dimensions to get an overall program quality score for each site. As shown in the first column of numbers in Table 6 below, the average score across all D2L sites was 3.0 in fall of 2018 and 3.2 in spring of 2019, the final year of implementation, suggesting an overall high level of quality on the program elements measured by the DOS.

Table 5: Frequency of Collaboration and Co-Teaching for D2L and Control Educators

	D2L Educators	Control Educators	Difference	
Frequency of collaboration (% at least once a week)	83.3	28.6	54.7	*
Frequency of co-teaching (% at least once a week)	96.7	44.4	52.3	*
Number of Sites	16	15		
Number of Educators	30	30		

Source: Research Alliance calculations based on 2019 D2L educator survey data.

Note: * Differences were statistically significant at p<.01, using a two-tailed t-test.

Table 6: Average 2018-19 Dimensions of Success Score for D2L and Control Sites

	D2L Sites	Control Sites	Difference	
Fall 2018	3.0	2.8	0.2	
Spring 2019	3.2	2.7	0.5	*
Number of Sites	16	15		

Source: Research Alliance calculations based on 2018-19 DOS data provided by ExpandED Schools.

Note: * Differences were statistically significant at p <.01, using a two-tailed t-test.

Activity Hours and Attendance

D2L sites were initially expected to deliver 72 hours of afterschool instruction—approximately 2 hours per week—over the course of the year. This expectation was later lowered to take into account time programs need to ramp up and down at the beginning and end of the school year, as well as holidays, field trips, and programmatic changes during state testing periods. The actual number of hours offered varied greatly by site and year of implementation, ranging from a low of 12 hours to a high of 98 hours. On average, D2L sites offered 46 hours of activities per year, about two thirds (64%) of what was expected (see the first column of numbers in Table 7). Only six D2L sites (38%) offered 72 or more hours of activity in any one year.

Table 7: Hours of Programming Offered in D2L and Control Sites, by Year

	Hours of Activity Offered	
Year	**D2L**	**Control**
2016-2017	57.3	72.9
2017-2018	30.3	59.0
2018-2019	56.0	60.0
Average Across Years	45.6	62.6
Number of Sites	16	15

Source: Research Alliance calculations based on student attendance data collected by ExpandED Schools.

Notes: Due to inconsistent and incomplete records from sites, we estimated the hours of activity offered from the maximum number of hours any student at the site attended.

Furthermore, we found that student attendance in these activities was low. On average, D2L students attended about 32 hours of programming in their first year of participation (see the first column of numbers in Table 8 below). This was less than half (44%) of the 72 hours of activity sites were expected to provide. Average student attendance also varied substantially by site (ranging from 9 to 51 hours per year). Just four of the 16 D2L sites reached 32 or more hours of attendance per student per year.

In interviews with educators, low student attendance was frequently noted as an implementation challenge. The reasons cited for low attendance included competing afterschool activities that were held at the same time (for example, drama, step, and basketball were mentioned as popular activities held at the same time that drew students away from D2L); student responsibilities at home, such as taking care of younger siblings; and the weather and time of year (attendance was lower during warm weather and at the end of the school year). As one educator noted: “They don’t want to come when the weather’s warm. They want to be outside.”

Table 8: Average Hours of Afterschool Attendance for D2L and Control Students

	D2L Students	Control Students
Average Attendance (hours)	32	34
Number of Students	1,264	1,703
Number of Sites	16	15

Source: Research Alliance calculations based on student attendance data collected by ExpandED Schools.

Note: Averages are based on students' attendance during their first year of participation. Students' first year of participation could have occurred in 2016-2017, 2017-2018, or 2018-2019.

In summary, our implementation data present mixed findings on the extent to which D2L was delivered as intended. In terms of the program inputs—delivery of and participation in the professional development, strategic planning, and coaching supports that ExpandED and NYSCI provided—we found strong fidelity to the model. In terms of sites' implementation of the D2L core pillars, we also found fidelity. D2L students reported frequently experiencing activities that were grounded in design-based learning and connected to their in-school science learning. In terms of program quality, observers found "reasonable evidence" that D2L sites were providing students with quality STEM learning experiences. However, D2L sites offered on average far fewer hours of programming than expected, and student attendance in those activities was generally low. We also found a great deal of variation across sites in terms of the number of activity hours offered and student attendance. Finally, students generally did not receive three years of programming as intended. These findings raise questions about whether students experienced enough of the intervention to reasonably expect an impact on their outcomes. In the next section, we evaluate the extent to which the D2L and control conditions were different.

Treatment/Control Contrast

Understanding the experience of students in the control sites, and how it compares to those in D2L sites, is key to understanding the results of our impact analyses. Sites randomly assigned to the control condition implemented the school-day and afterschool STEM programming they would normally implement ("business as usual"). The STEM programming included a wide variety of science-related activities, such as robotics, environmental and mechanical engineering, bridge building, digital music making, computer science coding, and architecture. To assess the difference between the D2L and control conditions, we compared D2L and control student and educator responses to the same survey items related to the core pillars of the D2L model: use of design-based instructional pedagogy, bridging the in-school and afterschool curriculum, and educator collaboration. We also compared the number of activity hours offered by D2L and control sites, and student attendance in those activities. If the contrast between what students experienced in the D2L and control sites was not strong, and control students participated in afterschool activities that looked similar to their D2L peers in terms of the model's key characteristics, then we would not expect their outcomes to be different.

Our findings suggest that the afterschool programming in control sites *did* include the key characteristics of D2L, at least to some extent. Control site activities included curricular bridging, design-based instruction, and educator collaboration. As shown in Table 4 above, there was no difference in average scale scores for D2L and control students on the curricular bridging items. On the design-based instruction survey items, there was a small statistically significant difference, with control students reporting less design-based instruction than their peers. Nonetheless, control students still reported fairly frequent design-based instruction. For example, 56 percent of control students said they experienced hands-on instruction often or almost always.

Control educators also reported collaborating together to deliver the STEM activities, although not necessarily with in-school educators nor as frequently as D2L educators (see Table 5). We found that 29 percent of control educators collaborated with their peers, and 44 percent co-taught activities (compared with 83 and 97 percent respectively for D2L). In terms of program quality, ratings were somewhat lower for control sites, though D2L/control site differences were statistically significant only in the spring (see Table 6). Further, D2L sites showed a small improvement from fall to spring, whereas control sites saw a small decline. This may suggest that the D2L supports helped enhance program quality over time.

Finally, in terms of the number of hours of activities offered, control sites also fell below project expectations of 72 hours per year, but they offered more hours of activities than D2L sites (46 versus 63 hours, as shown in Table 7). Student attendance was similar at D2L and control sites (32 versus 34 hours, as shown in Table 8).

Overall, the data indicate that D2L was implemented well and hewed closely to the model design—with the notable exception of the number of hours offered and attended. However, control students also participated in activities characterized by key elements of the D2L model: curricular bridging, and to a lesser extent, design-based instruction and educator collaboration. And though control sites offered more hours of activity than D2L sites, student attendance was about the same. Together, these findings suggest there may not have been a strong contrast between the D2L and control conditions.

Chapter 4: Design2Learn Impact

As discussed in Chapter 1, the D2L theory of change hypothesized that participation in afterschool activities characterized by 1) collaborative teaching between an in-school science teacher and two afterschool educators, 2) curricular bridging, which connects in-school science instruction with afterschool activities, and 3) design-based learning, which emphasizes hands-on activities and inquiry-based instruction, would increase students' interest and engagement in science. Increased interest and engagement were theorized to lead to improved achievement, including science test scores, grades, and attendance. In this chapter, we summarize findings from our analysis of D2L's impact on these outcomes.

To estimate the impact of D2L, we compared the outcomes of students in the D2L sites to those in their matched control sites, controlling for baseline levels of science interest and engagement and background characteristics. As described in Chapter 2, science interest and engagement were measured through surveys administered at the beginning of the school year (baseline) and end of the school year (follow-up). The end-of-year survey also collected information about students' self-reported perseverance when solving problems and critical thinking behaviors—21st-century skills that are correlated with STEM interest and engagement (PEAR Institute, 2019). Using extant school record data, we tested the impact of D2L on students' science grades and in-school attendance. In cases where students were missing a baseline survey, we imputed the baseline interest and engagement measure using the site mean baseline score for students in the same grade and gender. Additional details on the analytic model specifications are located in the Appendix.

Results of our impact analyses show that one year of exposure to D2L had no effect on students' interest and engagement in science at follow-up (see Table 9 on the next page). Both D2L and control students had an average follow-up interest and engagement score of 2.7 out of 4. Further, students' scale scores on the follow-up survey were slightly lower than their scores on the baseline measure (2.9 vs. 2.7), suggesting that science interest did not increase for either the D2L or control group.

We also saw no impact on students' self-reported perseverance (e.g., "If I fail to solve a problem, I will try again until I find the solution"), critical thinking behaviors (e.g., "I like to think of different ways to solve a problem"), or academic outcomes (science grades, school attendance).

Table 9: Impact of D2L on Student Outcomes					
Student Outcomes	**D2L Students**	**Control Students**	**Estimated Difference**	**P-Value**	**Effect Size**
Follow-Up Science Interest and Engagement	2.7	2.7	0.0	0.7	0.0
21st Century Skills					
Critical Thinking (1-4)	2.9	2.9	-0.1	0.4	-0.1
Perseverance (1-4)	2.9	2.9	0.0	0.5	-0.1
Academic Outcomes					
Science Grades (1-100)	80.7	80.6	0.1	0.9	0.0
Average Daily Attendance (%)	94.0	93.3	0.7	0.5	0.1
Number of Students	510	615			
Number of Sites	16	15			

Sources: Research Alliance calculations based on the D2L pre and post student surveys and school record data obtained from the NYC Department of Education.

Notes: Data on science grades were only available for 408 D2L and 473 control students. Data on school attendance were only available for 444 D2L and 508 control students. No D2L/control group differences were statistically significant. Differences for each measure between D2L and control sites are regression adjusted to account for site differences associated with student demographic characteristics and baseline interest and engagement. The analysis also accounts for the clustering of students within sites.

In addition to these analyses, we conducted several exploratory and non-experimental analyses. This included analyses to explore the impact of D2L on specific groups of students (e.g., groups defined by gender, race/ethnicity, English language learner and special education status); to explore the impact of two years of exposure; and to explore whether fidelity of implementation and students' degree of exposure to the intervention were related to the outcomes of interest. We found no significant differences between D2L and control students in these analyses, nor did we find evidence that more participation or fidelity of implementation were linked to program outcomes (see the Appendix for details).

Chapter 5: Summary and Conclusions

D2L aspired to improve middle school students' interest and engagement in science, and ultimately their science achievement, through a three-year intervention that combined three key strategies: design-based instruction, curricular bridging, and collaborative teaching. In this study, we did not find evidence that D2L influenced these intended outcomes. To understand why this might be, it is important to consider the extent to which D2L was implemented as designed, the extent to which the experience of students in D2L was different from their peers in the control condition, and the amount (or "dosage") of D2L that students experienced.

In terms of the first consideration, we found strong fidelity of implementation for key program inputs. ExpandED delivered D2L professional development and associated supports, and sites participated in these activities as expected. Further, student and educator reports suggested that D2L activities were characterized by the three key pillars of the model. For example, students reported that they "often" experienced science activities that were hands-on and inquiry based—key features of design-based instruction—and that the afterschool educators connected those activities to what students were learning in school. The afterschool educators in D2L sites also reported frequently planning together and co-teaching.

Speaking to the second consideration, however, students and educators in the control sites also reported experiencing design-based science instruction, curricular bridging, and educator collaboration. In fact, students in the control sites reported the same frequency of curricular bridging as their D2L peers. And though they reported less exposure to design-based instruction and collaboration, these practices were in place in the control sites to a substantial degree. Further, data from observations of the activities indicated that control site activities were of a similar quality to the D2L sites. These findings suggest there may not have been a sufficient amount of contrast in the D2L and control conditions.

In terms of the third consideration, dosage, we found that sites offered—and students participated in—far fewer hours of activity than expected. As designed, D2L was intended to provide students with up to 72 hours of additional science activities per year. On average, D2L students attended 32 hours of program activities, slightly less than control students and less than half of the amount expected per year. Further, students by and large did not attend the afterschool programming for more than one year, let alone the three years that program planners initially hoped for. Thus, it is unclear if students experienced enough of the intervention—both in terms of number of hours and consecutive years—to reasonably expect an impact on their outcomes.

Together, the findings from this study raise important questions about afterschool interventions designed to increase middle school student interest and achievement in science. First, the low participation rates of students in this study point to the difficulty in recruiting and retaining middle grade students for an afterschool initiative. Perhaps sixth through eighth grades were not the best grades to target, or perhaps more could have been done to improve recruitment and retention of students in these grades. Prior research suggests attracting students in this age range to afterschool programs is a widespread problem that calls for aggressive recruitment efforts (Grossman, Raley, & Walker, 2005). We know that interest and engagement in science drops precipitously by middle school, suggesting the need to implement interventions during this time period. However, if most middle school students do not attend such programs regularly, future research should investigate the efficacy of shorter-term and more flexible and informal experiences at this grade level. It may be that in order to prevent declining interest and achievement in science during middle school, interventions should preemptively target elementary-aged students. Thus, future research might also explore the efficacy of a similar, multi-year intervention geared toward students in upper elementary grades, where afterschool attendance is more reliable because of parents' childcare needs.

Second, though our findings suggest D2L students frequently experienced the three pillars of the model, the control sites also employed these strategies (though to a somewhat lessor degree) and were of similar program quality. Thus, there may not have been enough of a contrast between students' experiences in D2L and the control condition to reveal impact. This raises questions about the extent to which these practices are already in place in typical afterschool programs. It may be that most afterschool programs already employ such practices, or, because the control sites were recruited from the larger network of afterschool programs supported by ExpandED, it is possible that the control sites are not typical of most afterschool science programs and do not reflect "business as usual." They may be of higher quality or more in line with the D2L model than typical programs.

Third, our findings raise questions about how much exposure to D2L activities students need to make a difference in their science interest and skills. Middle school students in NYC receive approximately 108 hours of in-school science instruction (the equivalent of one unit) per year. [4] Full participation in D2L (72 hours) would increase students' exposure to science by about two thirds. If students in this study had more exposure to D2L, would it have led to impacts on the hypothesized student outcomes? Or, is there a flaw in the theory that curricular bridging, design-based instruction, and collaboration in the context of afterschool science programming can improve students' interest, engagement, and achievement in science? Given the importance of science literacy to succeed in 21st century society and careers, and the imperative to address equity issues raised by the underrepresentation of students in STEM, it is essential to conduct further research to answer these questions and understand the mechanisms that lead to increased interest and achievement in science.

Endnotes

1 This section describes the D2L model as ExpandED expected it to be implemented. In some sites, however, it was impossible to recruit an in-school science teacher from the targeted grade to participate in the D2L afterschool activity. In cases when this was not possible, the educator team was asked to communicate with the relevant in-school science teacher or teachers, for example, by attending department meetings.

2 ExpandED took on this role because the DOS protocol must be conducted by trained and certified observers. ExpandED staff had this experience.

3 In Tables 4-8, we show results for both D2L and control sites. In the text, we first describe findings from the “D2L” columns of these tables—to understand the extent to which the program was implemented as designed. Then we describe the information from the “Control” columns—to explore the treatment/control contrast.

4 Students in 7th and 8th grade must receive one unit of science each year according to the NYC middle school academic policy guide.

References

Afterschool Alliance. (2015). *Full STEM Ahead: Afterschool Programs Step Up as Key Partners in STEM Education*. Washington, DC: Author. Retrieved from http://afterschoolalliance.org/AA3PM/STEM.pdf

Akiva, T. (2005). "Turning training into results: The new youth program quality assessment." *High/Scope Resource*, 24(2), 21-24.

Bennett, D. & Monahan, P. (2013). “NYSCI Design Lab: No bored kids!” In M. Honey & D. Kanter (Eds.). *Design, make, play: Growing the next generation of stem innovators.* New York: Routledge.

Capobianco, B.M., Yu, J.H. & French, B.F. (2014). “Effects of Engineering Design-Based Science on Elementary School Science Students’ Engineering Identity Development across Gender and Grade.” *Res Sci Educ,* 45, 275–292 (2015). https://doi.org/10.1007/s11165-014-9422-1

Commission on Mathematics and Science Education (US) (2009). *Opportunity Equation: Transforming Mathematics and Science Education for Citizenship and the Global Economy*. New York, NY: Carnegie Corporation of New York. Retrieved from: https://www.carnegie.org/media/filer_public/80/c8/80c8a7bc-c7ab-4f49-847d-1e2966f4dd97/ccny_report_2009_opportunityequation.pdf

Friend, M. & Cook, L. (2010). “Co-Teaching: An Illustration of the Complexity of Collaboration in Special Education.” *Journal of Educational and Psychological Consultation*, 20:9–27.

Grossman, Jean, Raley, Rebecca, and Walker, Karen (2005). *Getting It Right: Strategies for After-School Success.* Philadelphia, PA: Public/Private Ventures. Retrieved from: https://ppv.issuelab.org/resource/getting-it-right-strategies-for-after-school-success.html

Han, S., Capraro, R., & Capraro, M. M. (2015). “How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of student factors on achievement.” *International Journal of Science and Mathematics Education*, 13(5), 1089-1113.

Harris, C. J., Penuel, W. R., D'Angelo, C. M., DeBarger, A. H., Gallagher, L. P., Kennedy, C. A., ... & Krajcik, J. S. (2015). “Impact of project-based curriculum materials on student learning in science: Results of a randomized controlled trial.” *Journal of Research in Science Teaching*, 52(10), 1362-1385.

Krishnamurthi, A., Ballard, M., & Noam, G. G. (2014). *Examining the Impact of Afterschool STEM Programs*. Washington, DC: Afterschool Alliance.

Korur, F., Efe, G., Erdogan, F., & Tunç, B. (2017). “Effects of toy crane design-based learning on simple machines.” *International Journal of Science and Mathematics Education*, 15(2), 251-271.

Moorehead, T. & Grillo, K. (2013). “Celebrating the Reality of Inclusive STEM Education: Co-Teaching in Science and Mathematics.” *Teaching Exceptional Children*, 45(4), 50-57.

National Academies of Sciences, Engineering and Medicine (2021). *Cultivating Interest and Competencies in Computing: Authentic Experiences and Design Factors. A Consensus Study Report*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25912.

National Research Council & the Institute of Medicine (2004). *Engaging Schools: Fostering High School Students' Motivation to Learn.* Committee on Increasing High School Students' Engagement and Motivation to Learn. Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. 1996. *National Science Education Standards*. Washington, DC: The National Academies Press. https://doi.org/10.17226/4962.

Noam, G. G. (2003). “Learning with excitement: Bridging school and after-school worlds and project-based learning.” *New Directions for Youth Development*, 2003 (97), 121-138.

Noam, G., Biancarosa, G. & Dechausay, N. (2002). *Learning to Bridge–Bridging to Learn: A Model and Action Plan to Increase Engagement Between Schools and Afterschool Programs in Boston.* Program in Afterschool Education and Research (PAER), Harvard University. A Report Commissioned by Boston's After-School for All Partnership Learning Goal Research. Cambridge, MA: Harvard University.

OECD (2016). *Country Note: Key Findings from PISA 2015 for the United States.* Paris, France: Author. Retrieved from: http://www.oecd.org/pisa/pisa-2015-United-States.pdf.

PEAR Institute (2019). *A Guide to PEAR's STEM Tools: Common Instrument Suite & Dimensions of Success.* Cambridge, MA: Author. Retrieved from: http://stelar.edc.org/sites/stelar.edc.org/files/PEAR%20CIS-DoS%20Overview_2019.pdf

Pechman, E., C. Russell, & J. Birmingham (2008). *Out of School Time (OST) Observation Instrument.* Washington DC: Policy Studies Associates.

Pianta, R.C., K.M. La Paro, & B.K. Hamre (2008). *Classroom Assessment Scoring System™: Manual K-3.* Baltimore: MD: Paul H Brookes Publishing.

Saritepeci, M. (2020). “Developing computational thinking skills of high school students: Design-based learning activities and programming tasks.” *The Asia-Pacific Education Researcher*, 29(1), 35-54.

Shah, A. M., C. Wylie, D. Gitomer, & G. Noam (2018). "Improving STEM program quality in out-of-school-time: Tool development and validation." *Science Education,* 1-22.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2015). *2015 science results.* Washington DC: Institute of Education Sciences. Retrieved from: https://www.nationsreportcard.gov/science_2015/files/overview.pdf

U.S. Department of Education, Institute of Education Sciences (2020). "What works clearinghouse standards handbook (Version 4.1)." Washington DC: Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Vaino, K., Vaino, T., & Ottander, C. (2018). “Designing an ice cream making device: A design-based science learning approach.” *Science Education International,* 29(3).

Vongkulluksn, V. W., Matewos, A. M., Sinatra, G. M., & Marsh, J. A. (2018). “Motivational factors in makerspaces: a mixed methods study of elementary school students' situational interest, self-efficacy, and achievement emotions.” *International Journal of STEM Education*, 5(1), 1-19.

Zito, M. (2011). *Is Working Together Worth It? Examining the Relationship Between the Quality of Teacher Collaboration, Instruction, and Student Achievement.* Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of Doctor of Education.

The Research Alliance conducts rigorous studies on topics that matter to the City's public schools. We strive to advance equity and excellence in education by providing nonpartisan evidence about policies and practices that promote students' development and academic success.

The Research Alliance for New York City Schools

NYU | STEINHARDT

627 Broadway, 7th Floor
New York, NY, 10012

(212) 992-7697

research.alliance@nyu.edu
www.rancys.org