Original Manuscript

Evaluation Review
1-32

DOI: 10.1177/0193841X14553299
erx.sagepub.com

Museums, Zoos, and Gardens: How Formal-Informal Partnerships Can Impact Urban Students' Performance in Science

Meryle Weinstein[1], **Emilyn Ruble Whitesell**[2], **and Amy Ellen Schwartz**[1,2,3]

Abstract

Background: Informal science education institutions (ISEIs) are critical partners in public science education, as they support the science efforts of school systems by providing authentic opportunities for scientific inquiry. This study reports findings from an evaluation of urban advantage (UA), a collaboration between the New York City Department of Education and eight ISEIs designed to improve science education in New York City (NYC) middle schools. Now in its 10th year, the program harnesses the resources and expertise of NYC's ISEIs to (a) enhance the science content knowledge

---

[1] Steinhardt School of Culture, Education, and Human Development, New York University, New York, NY, USA
[2] Robert F. Wagner Graduate School of Public Service, New York University, New York, NY, USA
[3] Maxwell School of Citizenship and Public Affairs, Syracuse University, Syracuse, NY, USA

Corresponding Author:
Meryle Weinstein, Steinhardt School of Culture, Education, and Human Development, New York University, 665 Broadway, Suite 805, New York, NY 10012, USA.
Email: meryle.weinstein@nyu.edu

of middle school science teachers, (b) develop teachers' skills at using inquiry-based approaches in their classrooms, and (c) improve the science achievement of middle school students. **Objectives:** We examine whether the UA program has led to increased student achievement on the eighth-grade New York State standardized science exam for students in participating schools; in supplemental analyses, we examine the effects on longer term (ninth-grade) outcomes. **Research Design:** We use a difference-in-differences framework with school fixed effects to estimate the impact of attending a UA school in eighth grade on science achievement. **Measures:** Our key outcome is performance on New York State's eighth-grade intermediate-level science assessment; longer term outcomes include enrollment at specialized science, technology, engineering, and math high schools as well as taking and passing the high school (Regents) science exams. **Results:** We find that attending a UA school increases student performance on the eighth-grade science exam by approximately 0.05 SD, and there is some evidence of small effects on Regents taking and passing rates.

Keywords
science education, informal science education institutions, urban education, professional development

Introduction

Concern about the United States' competitiveness in an increasingly global economy is fueled in part by U.S. students' lackluster performance in math and science. Science, technology, engineering, and math (STEM) fields play a critical role in the nation's economy and individuals' career opportunities. The National Governor's Association reports that "STEM occupations are among the highest paying, fastest growing, and most influential in driving economic growth and innovation" (National Governors Association, 2011). Indeed, according to the U.S. Department of Commerce's Economics and Statistics Association, STEM employment grew at a rate of 7.9% compared to just 2.6% in non-STEM fields from 2000 to 2009, and STEM employment is projected to grow at a rate of 17.0% from 2009 to 2018 (Langdon, McKittrick, Beede, Khan, & Doms, 2011). Although STEM skills are growing in importance, U.S. students show mediocre achievement and little improvement in these subjects. For example, U.S. students ranked 20th out of the 34 Organisation for Economic Co-operation and Development (OECD) countries on the Programme for International Student Assessment science assessment in 2012 (OECD, 2014). On the nationally representative National Assessment of Educational Progress (NAEP), only 32% of eighth-grade students scored at or above proficient in science in 2011, which reflects a slight increase from 30% in 2009 (NAEP, 2011).

Given disappointing performance in math and science, policy makers and educators are increasingly worried about a mismatch in U.S. students' skills and what will be required to be successful in the 21st-century job market. One approach to improving U.S. performance is to focus on teaching and learning through scientific inquiry and application. This is reflected in the Next Generation Science Standards (NGSS), which are explicitly designed to be taught in a real-world, applied context and are based on the National Research Council's (NRC) *Framework for K–12 Science Education* (NRC, 2012; NGSS, 2013). This framework is designed for both formal and informal science educators, in accordance with the NRC's stance that informal science education institutions (ISEIs) are an important contributor to students' scientific engagement and achievement. Although formal-informal collaborations are still the exception rather than the rule, the NRC is hardly alone in its support of school-ISEI partnerships; in fact, the National Science Teachers Association, National Science Board, Institute of Museum and Library Sciences, and the Center for Informal Learning and Schools have all called for greater partnership with ISEIs to support science education both in and out of school.

One such large-scale, formal-informal collaboration is New York City's (NYC's) Urban Advantage (UA) program, which explicitly draws upon the expertise and resources of the city's ISEIs, bringing these institutions together with NYC public middle schools to improve science teaching and learning. UA differs fundamentally from traditional museum-school partnerships, as it is a hybrid model of formal-informal collaboration where the resources of ISEIs are selected, designed, and shaped specifically to align with the science curriculum of NYC's middle schools. This partnership is more intensive than most programs that involve outside resources in schools, as there is a deep relationship between the ISEIs, the NYC Department of Education (NYCDOE), and the participating schools. Additionally, since the program is designed specifically for NYC schools, it is more closely tied to the science curriculum than typical programs that focus on general science enrichment.

The UA program operates on a large scale in NYC, and as of its 10th year (2013–2014), it reached approximately 33,000 students and more than 500 teachers in one third of all NYC middle schools[1] (177 schools). In this article, we use a difference-in-differences framework to estimate the impact of

UA on middle school science test scores as well as on early high school outcomes, such as taking and passing a science Regents exam in eighth or ninth grade. Findings will be useful for program staff who seek to improve the UA program in NYC or to develop programs in other cities, but the results will be more broadly relevant for researchers, policy makers, and practitioners interested in the role of formal-informal partnerships in improving science education.

Background

ISEIs and Formal-Informal Collaborations

Several large research, education, and science organizations (e.g., NRC, National Science Teachers Association, National Science Board) emphasize the importance of informal science learning, as ISEIs provide a unique resource for students to engage in authentic scientific inquiry and connect science content knowledge and the scientific process to their daily lives. For example, the NRC has explicitly called for greater integration of school-based instruction, learning in ISEIs, and interactions with scientific concepts in daily life (NRC, 2009). Several studies confirm positive outcomes associated with informal science education; for example, exposure to informal science education can influence long-term career choices by making STEM careers an appealing and viable career choice (Darke, Clewell, & Sevo, 2002; Dorsen, Carlson, & Goodyear, 2006; Fadigan & Hammrich, 2004).

In accordance with their mission, the vast majority of ISEIs provide targeted educational services, such as teacher and student programs. In the Centre for Informal Learning and Schools' survey of 345 ISEIs, 73% reported providing "support in the way of programs, workshops, materials, curricula, etc. for districts, schools, teachers, or students in the broad area of science education besides a one-day field trip" (Phillips, Finkelstein, & Wever-Frerichs, 2007, p. 1492). For teachers, ISEIs frequently provide residency programs, research opportunities, and professional development (PD). These programs vary in their intensity, with some leading to official degrees or certification (Saxton, Gupta, & Steinberg, 2010). For students, ISEIs typically feature family outreach programs, camp-ins, activity kits, science materials, and out-of-school time programs (Astor-Jack, Balcerzak, & McCallie, 2010; Hein, 1998; Hofstein & Rosenfeld, 1996; Inverness Research Associates, 1996; Kisiel, 2010; Ramey-Gassert, Walberg, & Walberg 1994).

Unfortunately, studies suggest that science museum resources are generally underused by the teachers and students they hope to serve, and most interactions between schools and ISEIs are infrequent and primarily dependent on the actions of individual teachers (Kisiel, 2010). For example, 53% of the institutions responding to the Centre for Informal Learning and Schools' survey reported their programs could handle more participants than they currently serve, while only 24% indicated they turn away potential participants due to capacity constraints (Phillips et al., 2007). Research suggests external factors such as rising costs and accountability concerns likely influence teacher and school interactions with ISEIs (Anderson, Kisiel, & Storksdieck, 2006; DeWitt & Storksdieck, 2008). Structured collaborations between ISEIs and schools are intended to integrate science education resources across contexts, provide more cohesive educational experiences, and overcome some of the barriers that make it difficult for schools to fully capitalize on the wealth of resources available at ISEIs.

Professional Development

Contemporary efforts to improve science education in the United States emphasize the importance of effective PD to help science teachers increase content knowledge and improve pedagogy. PD is especially important for secondary STEM teachers, as it has historically been difficult for schools to attract and retain educators with strong STEM content knowledge. For example, using data from the 2007–2008 Schools and Staffing Survey, the National Center for Education Statistics (NCES) reports that approximately 10–20% of secondary math and science teachers are not certified to teach their subjects, with teachers less qualified in schools that serve large populations of minority students (NCES, 2010). This issue gained national prominence in the wake of President Obama's 2011 State of the Union address, in which he famously outlined plans to recruit and train 100,000 new high-quality STEM teachers in the next 10 years (The White House, 2011).

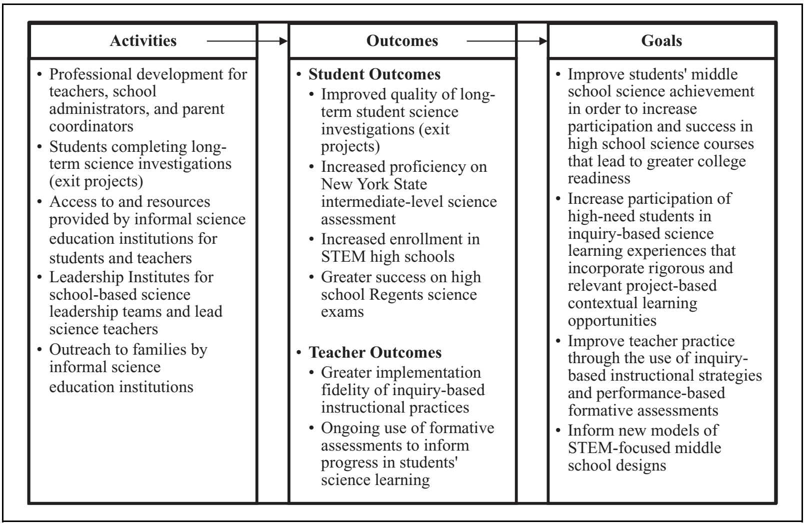
Best practices in effective PD emphasize that high-quality experiences use an inquiry-based approach, deepen teachers' content knowledge, are grounded in standards and research, are ongoing and intensive, and are relevant for teachers' curriculum and classroom contexts (e.g., Desimone, Porter, Garet, Yoon, & Birman, 2002; Fishman, Marx, Best, & Tal, 2003; Loucks-Horsley & Matsumoto, 1999; Loucks-Horsley, Stiles, Mundry, Love, & Hewson, 2009; NRC, 2000; Supovitz & Turner, 2000).

The UA Program

UA was launched in September 2004 as a collaboration between eight NYC ISEIs (American Museum of Natural History, Brooklyn Botanic Garden, New York Botanical Garden, New York Hall of Science, Queens Botanical Garden, Staten Island Zoological Society, and the Wildlife Conservation Society's Bronx Zoo and New York Aquarium) and the NYCDOE to provide teachers and students with opportunities to engage in authentic science practice. The ultimate goals of UA are broad, including improving middle school science achievement, increasing success in high school science, and improving college readiness, but UA has articulated proximal outcomes. These outcomes, which we use in our empirical models, include performance on the New York State eighth-grade intermediate-level science (ILS) assessment, enrollment in STEM high schools, and success in high school (Regents) science exams. Working with the first author, UA staff have developed a logic model to articulate and link inputs, activities, outcomes, and goals; a simplified version is shown in Figure 1

As shown in the logic model, the UA program includes several elements (or "activities") aimed toward improving outcomes and ultimately meeting program goals. Notably, UA provides 48 hr of PD (across three cycles in the same academic year) for teachers new to UA and 10 hr each year for teachers continuing in the program. The PD model is designed by UA program who represent the broader scientific community (e.g., geologists, astronomers, biologists) and science educators, using an immersion-into-inquiry strategy. This approach provides authentic hands-on learning experiences for teachers, who focus on the nature of scientific work, specific science content areas, and the essential features of inquiry in the form of long-term investigations. As part of their training, teachers conduct their own scientific investigations, experiencing firsthand what it means to "do science," which is consistent with the teacher-as-learner model of PD (Loucks-Horsley et al., 2009; NRC, 2000; Thompson & Zeuli, 1999).

After choosing a UA partner institution to attend for targeted PD, teachers learn how to plan effective field trips, embed resources in instruction, and teach students the components of experimental design as well as how to develop scientific explanations based on claims, evidence, and reasoning. As a part of this PD, teachers are trained to use a variety of UA classroom tools, such as the Investigation Design Diagram and the Developing a Scientific Explanation Tool, which are intended to support teachers as they implement UA principles in their classrooms. Teachers who are new to UA attend three PD cycles throughout their first year in the program and



Activities → **Outcomes** → **Goals**

Activities

- Professional development for teachers, school administrators, and parent coordinators
- Students completing long-term science investigations (exit projects)
- Access to and resources provided by informal science education institutions for students and teachers
- Leadership Institutes for school-based science leadership teams and lead science teachers
- Outreach to families by informal science education institutions

Outcomes

- **Student Outcomes**
  - Improved quality of long-term student science investigations (exit projects)
  - Increased proficiency on New York State intermediate-level science assessment
  - Increased enrollment in STEM high schools
  - Greater success on high school Regents science exams
- **Teacher Outcomes**
  - Greater implementation fidelity of inquiry-based instructional practices
  - Ongoing use of formative assessments to inform progress in students' science learning

Goals

- Improve students' middle school science achievement in order to increase participation and success in high school science courses that lead to greater college readiness
- Increase participation of high-need students in inquiry-based science learning experiences that incorporate rigorous and relevant project-based contextual learning opportunities
- Improve teacher practice through the use of inquiry-based instructional strategies and performance-based formative assessments
- Inform new models of STEM-focused middle school designs

Figure 1. Urban Advantage program logic model.

may be unable to apply all UA principles until their second year in the program (after they have completed all three cycles). Furthermore, with the help of targeted PD from UA, teachers likely continue to develop over time, potentially becoming more effective in implementing UA principles and practices as they gain experience in the program.

In addition to high-quality PD, participating teachers and schools receive material and monetary support from the UA. For example, UA supplies schools with science kits containing materials to be shared among UA teachers, and it provides funds for individual teachers to purchase materials of their choice for their classrooms. Additionally, UA provides vouchers that give teachers, administrators, students, and families free admission to any of the eight partner institutions (ISEIs), and schools also receive vouchers for transportation. These vouchers are intended both to strengthen the quality of class field trips and to support science learning outside the classroom. Schools use these resources in a variety of ways. For example, some teachers use their individual vouchers to visit ISEIs and plan future field trips, while others use them for personal development. Similarly, some teachers integrate class field trips into their curriculum and exit projects, while other teachers plan enrichment-based trips that are less directly related to specific units or projects. Parent coordinators are encouraged to organize family field trips on the weekends, although many schools simply distribute family vouchers to students and encourage them to take family trips independent of the school.

The nature of the relationship between UA teachers and ISEI staff varies across teachers based on individual preferences and constraints. Some teachers report choosing partner institutions based on personal factors (e.g., schedule, transportation, proximity), while others identify a partner institution based on course curricula and the ability to integrate ISEI resources into specific units. Some teachers engage with partner institutions primarily at PD, while others develop this relationship more extensively. For example, many teachers draw on the expertise of the staff at ISEIs to help them plan field trips, and some invite ISEI-based educators to come to their schools for specific class activities or to attend Family Science Nights.

The UA program has evolved over time, both in terms of strategic expansion and in terms of development of program elements. In its early years, UA accepted teachers into the program on a volunteer basis; teachers learned about the program from their principals or from other teachers, and those who were interested self-selected into UA. Over time, the program has developed a more rigorous protocol for accepting both teachers and schools.[2] This is partly due to increased demand and partly due to budget reductions, which reflect the fiscal constraints of the program's funders (the NYC Council and the NYCDOE). Rather than expand to provide the program to more schools, the UA staff opted to grow within already-participating schools by opening the program to sixth-grade teachers and adding additional teachers per grade.[3] This reflects the belief that UA is best implemented in schools with a greater share of science teachers participating in the program, as having a high concentration of UA teachers in a school may provide opportunities for teacher collaboration within schools and give students repeated exposure to UA concepts across different grades.

UA staff and partner institutions have also worked to develop program resources[4] and PD over time. For example, as the balance of UA teachers has shifted to teachers continuing in the program, UA staff have developed more PD offerings tailored to the needs of continuing teachers.[5] These specially designed workshops are open only to teachers who have already participated in the PD for first-year UA teachers, focus in greater depth on specific content related to the science exit projects, and provide greater opportunities for experienced teachers to examine student work and assess students' thinking. To help ensure ongoing participation in UA, attendance at these workshops is required for teachers to continue to receive resources and classroom materials from the program (Short, Elgendy, Roditi, &

Holmes, 2012). UA also capitalizes on the expertise of more experienced UA teachers by involving them in the PD of newer UA teachers; UA has designated "lead teachers" who facilitate designated UA PD events and support other UA teachers in their schools more informally.[6]

Method

Data and Measures

Our analysis draws on a rich student-level longitudinal database for NYC public schools and students from 2003-2004 (hereafter, 2004) to 2009-2010 (hereafter, 2010).[7] Every student record contains detailed demographic, program, and academic information, including nativity, race, gender, language ability, poverty (free/reduced-price lunch status), attendance rates, participation in special education and language programs, and standardized test scores. These data are combined with publicly available data from the *Annual School Reports* and *State Report Cards* prepared annually by the NYCDOE and the New York State Education Department. In addition to the rich detail and breadth of our data, unique student identification numbers allow us to track students from their entrance in the NYC public school system until their departure. Importantly for our analyses, we are able to follow students from middle school to high school. Our sample includes eighth graders from 2004 to 2010, for a total of more than 460,000 student-year observations.

We use UA's logic model to guide our definition of relevant student outcomes. In accordance with the logic model, we measure short-term outcomes using the eighth-grade ILS exam. New York State requires that all eighth-grade students take the ILS test, which consists of approximately 80 questions, including multiple choice, open response, and performance tasks. The test covers three broad standards: scientific inquiry, living environment, and physical setting. We measure student performance on the ILS with a *standardized score* (z-score), a measure of relative performance standardized across students within a grade to have a mean of 0 and standard deviation of 1. Students performing above (below) average relative to other students in their grade, in that year, have positive (negative) z-scores.

Again in accordance with UA's logic model, we use several different measures for long-term outcomes. First, we examine students' likelihood of enrolling in an STEM high school. In NYC, students have substantial choice in what high school they attend, and many high schools offer multiple specialized academies students can choose from, such as health professions, technology, law, journalism, computer science, humanities, and performing arts.[8] Schools vary in terms of both how many specialized tracks (if any) they offer and what types of curricula these programs provide. If UA fosters a greater appreciation for and understanding of science, then it is possible that UA students will be more interested in STEM schools, more qualified to enroll in them, or both.

For the purposes of this analysis, we define "all-STEM" schools as those that offer *only* science-rich academies—that is, all students in the school are in a science-specific program. We define "partial-STEM" schools as those that offer both science-based academies and nonscience academies to students. We operationalize enrollment at STEM schools in ninth grade with dichotomous variables. All-STEM takes a value of 1 if a student is enrolled in an all-STEM school, and partial-STEM takes a value of 1 if a student is enrolled in a partial-STEM school.[9]

We also explore students' likelihood of taking and passing science Regents exams in the eighth or ninth grade. In New York State, students have some choice about what Regents exams they take and when to take them, but they are required to pass at least one science Regents exam in order to graduate. Students often take either the Earth Science Regents exam or the Living Environment (biology) Regents exam in eighth or ninth grade, which then allows them to take higher-level science courses, such as chemistry and physics, later in high school. In some ways, taking and passing a science Regents exam in eighth or ninth grade is a proxy for interest in science as well as a reflection of future opportunities to take higher-level science courses.

In our sample, 44% of students take either the Earth Science or the Living Environment Regents exam in eighth or ninth grade, although test-taking behavior varies considerably by grade and increases over time. In our analysis, we explore students' taking and passing behavior on these two science Regents exams, as they are the two most commonly taken by all NYC high school students. Here again, we use dichotomous variables reflecting whether or not a student (a) took and (b) passed the exam. There are three different passing rates for which we show results (passing at the 55, 65, and 85 cut-points), which reflect different degrees of proficiency and qualify students for different types of diplomas.[10]

We measure UA program participation at the school level, and our results can be interpreted as intent-to-treat (not treatment-on-the-treated) estimates. This is because we are unable to match students to their specific teachers, and thus we must estimate the impact of attending a UA school—not the impact of having a UA teacher.[11] Furthermore, it is possible that elements of the UA program treatment "spill over" from UA teachers to non-UA teachers in their schools. Although spillover across teachers within schools is a potential source of unobserved heterogeneity, defining the treatment at the school level means we have "captured" this spillover with our treatment variable. To be clear, we classify a school as a UA school if at least one teacher in the school is participating in the UA program; thus, schools vary in their concentration of UA teachers.[12]

In any given year, the comparison group includes all NYC public schools that both (a) have an eighth grade and (b) are not participating in UA. Because schools join UA in different years (treatment is rolled out across schools over time), UA schools will be in the comparison group in the years prior to joining UA and then be in the treatment group once they join UA. Thus, our analysis exploits variation in program participation across schools over time.

Models

We estimate a series of models to assess the impact of attending a UA school on student outcomes. When we use the short-term outcome (standardized science test scores), these models are standard fixed effects models. For the long-term (dichotomous) outcomes, though, we use linear probability models. Our baseline specification models student outcomes as a function of attending a UA school in the current year, and we include student covariates (ST_{it}), a year trend (t), and school effects (α_j). This model is as follows:

$$Y_{ijt} = \beta_0 + \beta_1 UA_{jt} + \beta_2 ST_{it} + t + \alpha_{\mathrm{j}} + \varepsilon_{ijt}$$

In this model, Y is the outcome of interest for student i in school j in year t, and UA is an indicator variable that takes a value of 1 if school j was participating in UA in year t. In this specification, β_1 represents the impact of attending a UA school (in the current year) on student achievement, controlling for student characteristics. By including school fixed effects, we compare students in the same schools in the years before and after the school joins UA. By looking within schools over time, we guard against the possibility that our estimates reflect unobserved differences in the schools that join UA, compared to the schools that do not join UA.

School effects are also useful in controlling for differential selection into UA. If UA schools are fundamentally different than other schools in unobservable and time-invariant ways (e.g., higher performing, better managed), this will be captured by school effects. Of course, it is also possible that schools select into UA based on time-varying school factors. For example, schools may choose to join UA after a year of particularly high or low science test scores. To assess and control for potential differential selection based on performance in the year before joining UA, the second specification includes an indicator variable (Pre*UA*) that takes a value of 1 in the year before a school joins UA. This model is as follows:

$$Y_{ijt} = \beta_0 + \beta_1 \mathrm{Pre}UA_{jt} + \beta_2 \mathrm{UA}_{jt} + \beta_3 ST_{it} + t + \alpha_j + \varepsilon_{ijt}$$

Here, β_1 represents the difference between UA schools and non-UA schools in the year prior to joining the program, and β_2 is the impact of UA on student outcomes.

Finally, the third specification distinguishes between the first year a school joins UA and all subsequent years (the second year in the program and beyond). This takes into account the fact that the UA program is not likely to be fully implemented in the first year, as described previously. The specification we use to model this relationship is as follows:

$$Y_{ijt} = \beta_0 + \beta_1 \mathrm{Pre}UA_{jt} + \beta_2 \mathrm{Base}Yr_{jt} + \beta_3 \mathrm{Post}Yrs_{jt} + \beta_4 ST_{it} + t + \alpha_j + \varepsilon_{ijt}$$

In this model, Base*Yr* is an indicator variable that takes a value of 1 if, in year t, school j is in its first year of participating in UA. Post*Yrs* is an indicator variable that takes a value of 1 if the school is in at least its second year of UA implementation. In this model, β_1 represents the difference between UA schools and non-UA schools in the year prior to joining the program, β_2 is the impact of UA during the first year a school is in UA, and β_3 is the impact of UA in all other years after a school joins UA. Note that we do not ever "turn off" the Post*Yrs* indicator variable. That is, Post*Yrs* will equal 1 for all years after the base year, even if the school withdraws from the program. We do not "turn off" this treatment because after teachers go through UA PD and schools receive UA resources, it is impossible to fully retract the treatment from a school. Thus, these impact estimates will likely be lower bounds for the true program effect.

Note that in some specifications we include a control for a student's lagged math or English language arts (ELA) score. This is not a true value-added model, as we do not have a lagged science test score (no science test is given in seventh grade). Because math and science test scores are highly correlated, however, the lagged math score proxies for prior STEM performance.

Table 1. Mean Characteristics of UA and Non-UA Schools, 2010.[13]

Student Body	UA	Not UA	School Location	UA	Not UA
Total enrollment	**720**	**593**	% Manhattan	20.7	21.0
	(425)	**(350)**		(40.6)	(40.8)
% Black	**33.4**	**39.3**	% Brooklyn	30.0	34.8
	(29.1)	**(29.4)**		(46.0)	(47.7)
% Hispanic	41.8	41.1	% Bronx	24.7	24.6
	(26.3)	(26.5)		(43.3)	(43.1)
% Asian	11.7	9.2	% Queens	20.0	18.4
	(15.9)	(14.1)		(40.1)	(38.8)
% White	12.9	10.3	% Staten Island	4.7	1.3
	(19.3)	(17.7)		(21.2)	(11.4)
% LEP	12.1	11.8	*N*	150	305
	(10.4)	(12.8)			
% Free lunch	70.3	70.9			
	(20.0)	(20.0)			
% Reduced lunch	8.7	8.2			
	(5.0)	(4.9)			
% Passing reading exam	37.8	36.6			
	(20.0)	(21.6)			
% Passing math exam	49.7	49.5			
	(21.9)	(22.8)			
% Passing science exam	53.0	51.4			
	(22.4)	(23.5)			
N	150	305			

Note. UA = Urban Advantage. Standard deviation in parentheses. Boldface indicates differences are statistically significant at .05 level or less. Percentage passing is the percentage scoring in level 3 or 4 out of 4.

Results

Table 1 provides descriptive statistics for UA and non-UA schools in 2010. UA schools are in many respects quite similar to other NYC public schools serving eighth graders. For example, the distribution of UA schools across NYC's boroughs is consistent with the distribution of non-UA schools, and both UA and non-UA schools have a similar mix of students in terms of demographic characteristics, educational needs, and test scores.[14] Across UA schools, as with city schools as a whole, there is substantial variation in school characteristics. As the large standard deviations (*SDs*) show, UA serves both quite large and very small schools, schools where all students are eligible for free or reduced-price lunch and those where only a small proportion are eligible, and schools where the majority of students are Black or Hispanic as well as those with a more balanced mix of student ethnicities.

The only consistent difference between UA and non-UA schools across all years is size. In the 2010 academic year, the average enrollment at UA schools was more than 700, compared to less than 600 at non-UA schools. Across all years in the sample, the average size of a UA school ranges from about 650 students to more than 1,000 students, compared to between 400 and 800 for non-UA schools. Other small differences in terms of student demographics and academic characteristics emerge in some years, but there are no other consistent differences between UA and non-UA schools across the sample period.

Before turning to our empirical models, we first show descriptive evidence that, on average, UA schools outperform non-UA schools on the eighth-grade ILS exam. Figure 2 shows average passing rates[15] on the ILS test from years 2004 (1 year prior to the inception of UA) through 2010. In this figure, we allow the sample of UA schools to change over time, as schools enter and exit the program. Note that in 2004, approximately 43% of NYC eighth graders were proficient in science, which is considerably less than the New York State public school average of 70% (University of the State of New York, 2005).

In the first 2 years of UA, there are no significant differences in student performance between UA and non-UA schools. However, in the third year, differences begin to emerge, with students at UA schools outperforming students at non-UA schools. In 2007, 43.4% of eighth graders at UA schools passed the science exam compared to 41.2% at non-UA schools. By 2010, the passing rates in UA and non-UA schools had improved significantly, although UA schools were still outperforming non-UA schools (passing rates of 61.2% vs. 56.2%).

This finding is consistent with the school improvement literature that argues 3 years is the minimum amount of time needed to see results from interventions (Fullan & Stiegelbauer, 1991). It is reasonable to expect UA to take several years to develop into an effective program. For example, by its third year of implementation, UA had a more developed and stronger program that included a more comprehensive set of materials and resources for teachers to use in the classroom. As previously described, it may also take individual schools and teachers several years to implement the program effectively.

Short-Term Outcome: Eighth-Grade ILS Performance

Although Figure 2 provides descriptive evidence that UA schools, on average, have higher passing rates on the eighth-grade ILS exam, it does not speak to

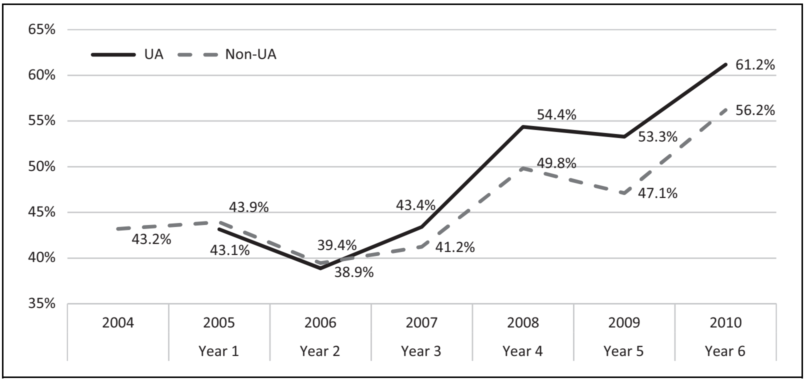


Figure 2. Student-weighted passing rates on the New York State eighth-grade intermediate-level science (ILS) exam, Urban Advantage (UA) and non-UA schools. *Note*. This figure uses school-level passing rates, weighted by the number of eighth graders in the school. In each year, there will be a different mix of schools in UA, as schools enter and exit the program.

causality. We now turn to our empirical models to estimate the impact of UA on student outcomes. We first estimate the short-term impact of UA on eighth-grade ILS test scores. For these models, our analytic sample includes all eighth graders who have science test scores in our sample period (2004–2010).

As seen in Table 2, the coefficient on the UA variable is statistically significant in the most basic model, which includes only a year trend (Model 1); thus, schools participating in UA outperform non-UA schools on average. After accounting for differences in student characteristics, however, the estimated coefficient decreases substantially and becomes insignificant (Model 2). Next, we turn to school fixed effects models to examine changes within schools over time. When we add school fixed effects (Model 3), we find a positive and statistically significant impact of UA; specifically, when we compare students in the same school (in years before and after a school joins UA), we see a positive impact of program participation of 0.039 *SD*s. This result is robust to controlling for performance in the year before joining UA (Model 4), and as the coefficient on the Pre*Yr* variable is essentially zero and statistically insignificant, we do not see evidence of either positive or negative selection into UA based on performance on the eighth-grade ILS in the prior year.

As described previously, UA is not likely to be fully implemented in a school until at least the second year in which a school has been participating. To model this appropriately, we distinguish between the first year a

Table 2. Impact of Attending a UA School on Eighth-Grade ILS Exam z-Scores, 2004–2010.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
UA in current year	0.058**	0.011	0.039***	0.040***		
	(0.029)	(0.017)	(0.011)	(0.013)		
Pre-year				0.004	0.012	0.007
				(0.016)	(0.017)	(0.015)
Yr entered UA					0.043**	0.025*
					(0.018)	(0.015)
Post-UA years					0.051***	0.036**
					(0.019)	(0.017)
Lagged *Z*-math						0.560***
						(0.004)
Black		−0.736***	−0.401***	−0.401***	−0.401***	−0.184***
		(0.016)	(0.009)	(0.009)	(0.009)	(0.006)
Hispanic		−0.538***	−0.233***	−0.233***	−0.233***	−0.096***
		(0.015)	(0.008)	(0.008)	(0.008)	(0.005)
Asian		0.098***	0.159***	0.159***	0.159***	0.012**
		(0.015)	(0.009)	(0.009)	(0.009)	(0.006)
Female		−0.062***	−0.073***	−0.073***	−0.073***	−0.041***
		(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
LEP		−0.916***	−0.815***	−0.815***	−0.815***	−0.405***
		(0.013)	(0.012)	(0.012)	(0.012)	(0.010)
SPED		−0.625***	−0.594***	−0.594***	−0.594***	−0.281***
		(0.008)	(0.007)	(0.007)	(0.007)	(0.005)

(continued)

Table 2. (continued)

Poor		−0.157***	−0.083***	−0.083***	−0.083***	−0.044***
		(0.009)	(0.006)	(0.006)	(0.006)	(0.004)
Constant	23.153*	17.265**	42.015***	42.001***	45.836***	31.156***
	(13.221)	(8.259)	(4.657)	(4.658)	(5.717)	(4.959)
Year trend	Y	Y	Y	Y	Y	Y
School effects	N	N	Y	Y	Y	Y
Observations	415,988	415,988	415,988	415,988	415,988	415,988
R^2	.001	.281	.383	.383	.383	.562

Note. UA = Urban Advantage; ILS = intermediate-level science; N = no; Y = yes; LEP = limited English proficient; SPED = special education. Models include controls for missing demographic variables, educational needs, and lagged test scores (only Specification 6). Robust clustered (school-year) standard errors in parentheses.

***$p < .01$, **$p < .05$, * $p < .1$.

school is in UA and all subsequent years (Model 5). We find that UA does have a statistically significant impact on ILS scores during the first year of implementation (0.043) and that the impact of UA in subsequent years is slightly larger (0.051) although not statistically different from the estimated coefficient for the first year. Thus, we see a positive effect of joining UA on students' science test scores, and there is evidence that the effect persists and potentially grows over time.

Model 6 augments the model by including lagged math achievement; as described previously, this is not a pure value-added model, as we do not have lagged science test scores. Including lagged math scores, however, allows us to control for some measure of students' academic ability/prior performance. In this specification, the impact of UA is still positive and statistically significant, although point estimates are smaller. Specifically, we find a positive effect of 0.025 *SD*s in a school's first year in UA and 0.036 *SD*s in subsequent years. We also estimate a specification that controls for lagged ELA performance, and results (not shown) are qualitatively similar.[16] In separate models that estimate the impact of attending a UA school on a student's probability of passing the science exam, we do not find any significant effect.[17]

Next, in Table 3, we provide subgroup results in which we estimate our preferred specification (Model 5) on subsamples of the population; note, however, that these results are also robust to including controls for prior test scores (Model 6).[18] Specifically, we estimate the impact of attending a UA school on subgroups by race, gender, educational needs, and poverty status. When we split the sample in this way, we see particularly large effects for Black students, male students, and those in special education; results for other subgroups suggest a positive although not always statistically significant relationship between attending a UA school and science achievement (e.g., White, female, and poor students).

For Black students, attending a UA school in its first year of implementation improves science test scores by 0.080 *SD*s, and the effect in subsequent years is slightly larger (0.086). Male students score 0.052 *SD*s higher in a school's first year in UA and 0.072 *SD*s higher in subsequent years, compared to male students at non-UA schools. For students in special education, the effect of attending a UA school in its first year of implementation is 0.071 *SD*s, with a larger effect in subsequent years (0.094). These results are encouraging, as on average, Black students and those in special education have relatively low performance on the science exam (see Table 2). The strong effect of UA on students in special education may be partially explained by special education teachers' participation in UA. Both science

Table 3. Subgroup Results: Impact of Attending a UA School on Eighth-Grade ILS Exam z-Scores, 2004–2010.

Variables	(1) Black	(2) Hispanic	(3) Asian	(4) White	(5) Female	(6) Male	(7) LEP	(8) SPED	(9) Poor
Pre-year	0.001	0.001	0.012	0.021	−0.003	0.027	0.017	0.013	0.002
	(0.027)	(0.021)	(0.026)	(0.029)	(0.018)	(0.019)	(0.035)	(0.031)	(0.019)
Yr entered UA	0.080***	0.011	−0.013	0.064**	0.037*	0.052***	−0.026	0.071**	0.030
	(0.029)	(0.024)	(0.027)	(0.026)	(0.019)	(0.020)	(0.034)	(0.031)	(0.020)
Post-UA years	0.086***	0.038	0.011	0.015	0.030	0.072***	0.016	0.094***	0.046**
	(0.031)	(0.024)	(0.029)	(0.028)	(0.021)	(0.021)	(0.041)	(0.033)	(0.021)
Black					−0.356***	−0.443***	−0.204***	−0.265***	−0.342***
					(0.010)	(0.011)	(0.033)	(0.017)	(0.009)
Hispanic					−0.234***	−0.220***	0.029	−0.112***	−0.172***
					(0.009)	(0.009)	(0.028)	(0.016)	(0.009)
Asian					0.189***	0.136***	0.262***	0.006	0.230***
					(0.011)	(0.010)	(0.033)	(0.022)	(0.011)
Female	−0.007	−0.106***	−0.075***	−0.132***			−0.014	−0.194***	−0.082***
	(0.005)	(0.005)	(0.007)	(0.007)			(0.009)	(0.009)	(0.004)
LEP	−0.812***	−0.691***	−1.118***	−1.201***	−0.764***	−0.858***		−0.411***	−0.770***
	(0.024)	(0.012)	(0.028)	(0.026)	(0.013)	(0.013)		(0.014)	(0.013)
SPED	−0.544***	−0.529***	−0.809***	−0.779***	−0.664***	−0.549***	−0.252***		−0.562***
	(0.010)	(0.008)	(0.021)	(0.014)	(0.008)	(0.008)	(0.017)		(0.007)
Poor	−0.060***	−0.056***	−0.042***	−0.182***	−0.106***	−0.059***	0.016	−0.028**	
	(0.008)	(0.007)	(0.010)	(0.010)	(0.006)	(0.006)	(0.013)	(0.012)	
Year trend	Y	Y	Y	Y	Y	Y	Y	Y	Y
School effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	126,739	157,568	59,602	57,858	200,346	197,552	41,931	34,973	268,048
R^2	.226	.276	.370	.330	.356	.358	.174	.227	.317

Note. UA = Urban Advantage; ILS = intermediate-level science; Y = yes; LEP = limited English proficient; SPED = special education. Robust clustered (school-year) standard errors in parentheses. Models include controls for missing demographic variables and educational needs (LEP and SPED). Constant not shown. All models include a year trend and school fixed effects (consistent with Table 2, Specification 5). Models controlling for lagged math test scores (consistent with Table 2, Specification 6) are similar.

***$p < .01$, **$p < .05$, *$p < .1$.

Table 4. Impact of Attending a UA School on Eighth-Grade Math and English Language Arts (ELA) *z*-Scores, 2004–2010.

	Math		ELA	
Variables	(1)	(2)	(3)	(4)
Pre-year	0.018	0.009	0.008	−0.002
	(0.018)	(0.014)	(0.015)	(0.014)
Yr entered UA	0.044**	0.024	0.033**	0.033**
	(0.018)	(0.015)	(0.016)	(0.014)
Post-UA years	0.019	0.004	0.021	0.014
	(0.020)	(0.017)	(0.016)	(0.015)
Lagged *Z*-score		0.679***		0.585***
		(0.004)		(0.004)
Black	−0.411***	−0.138***	−0.377***	−0.165***
	(0.011)	(0.006)	(0.011)	(0.007)
Hispanic	−0.274***	−0.100***	−0.277***	−0.119***
	(0.010)	(0.005)	(0.0 10)	(0.006)
Asian	0.404***	0.225***	0.064***	0.062***
	(0.012)	(0.007)	(0.011)	(0.007)
Female	0.025***	0.062***	0.192***	0.131***
	(0.003)	(0.002)	(0.003)	(0.002)
LEP	−0.598***	−0.072***	−1.073***	−0.529***
	(0.011)	(0.008)	(0.015)	(0.012)
SPED	−0.635***	−0.249***	−0.583***	−0.262***
	(0.007)	(0.005)	(0.007)	(0.005)
Poor	−0.062***	−0.016***	−0.115***	−0.055***
	(0.006)	(0.004)	(0.006)	(0.004)
Constant	23.450***	12.099***	6.862	4.161
	(5.519)	(4.309)	(4.745)	(4.195)
Year trend	Y	Y	Y	Y
School effects	Y	Y	Y	Y
Observations	421,956	421,956	405,568	405,568
R^2	.327	.605	.324	.544

Note. UA = Urban Advantage; ILS = intermediate-level science; Y = yes; LEP = limited English proficient; SPED = special education. Robust clustered (school-year) standard errors in parentheses. Columns 1 and 3 are consistent with Table 2, Specification 5; columns 2 and 4 are consistent with Table 2, Specification 6. Models include controls for missing demographic variables, educational needs, and lagged test scores (columns 2 and 4).
***$p < .01$, **$p < .05$, *$p < .1$.

teachers and special education teachers supporting science can participate in UA, and for many special education teachers, this is the only science-specific support they receive.

Table 5. Impact of Attending a UA School on the Probability of Enrolling in an STEM High School, 2004–2010.

	All STEM				Partial STEM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UA in current year	−0.009	−0.016***	−0.000	−0.001	0.053***	0.049***	0.000	−0.006
	(0.005)	(0.005)	(0.003)	(0.003)	(0.012)	(0.012)	(0.009)	(0.011)
Pre-year				−0.003				−0.015
				(0.004)				(0.012)
Student characteristics	N	Y	Y	Y	N	Y	Y	Y
Year trend	Y	Y	Y	Y	Y	Y	Y	Y
School effects	N	N	Y	Y	N	N	Y	Y
Observations	369,974	369,974	369,974	369,974	369,974	369,974	369,974	369,974
R^2	.002	.047	.115	.115	.011	.020	.131	.131

Note. STEM = Science, technology, engineering, and math; N = no; UA = Urban Advantage; Y = yes; LEP = limited English proficient; SPED = special education. Robust clustered (school-year) standard errors in parentheses. Results exclude students who are not observed for both eighth and ninth grades. In Table 5, columns 4 and 8 are consistent with Table 2, Specification 4. All models in Table 6 are consistent with Table 2, Specification 4. Models include controls for student demographics and educational needs; these coefficients, along with the constant, are not shown.

***$p < .01$, **$p < .05$, *$p < .1$.

Table 6. Impact of Attending a UA School on the Probability of Taking and Passing the Earth Science Regents Exam, 2004–2010.

	Sample: All Students				Sample: Students Who Took the Test		
	Took Test (1)	Passed 55 (2)	Passed 65 (3)	Passed 85 (4)	Passed 55 (5)	Passed 65 (6)	Passed 85 (7)
UA in current year	0.063***	0.048***	0.043***	0.010***	−0.007	−0.001	−0.012
	(0.010)	(0.008)	(0.007)	(0.003)	(0.005)	(0.007)	(0.009)
Pre-year	0.050***	0.035***	0.029***	0.005*	−0.011**	−0.007	−0.008
	(0.012)	(0.0 10)	(0.009)	(0.003)	(0.006)	(0.007)	(0.007)
Student characteristics	Y	Y	Y	Y	Y	Y	Y
Year trend	Y	Y	Y	Y	Y	Y	Y
School effects	Y	Y	Y	Y	Y	Y	Y
Observations	369,974	369,974	369,974	369,974	121,252	121,252	121,252
R^2	.101	.114	.123	.107	.128	.177	.190

Note. UA = Urban Advantage; Y = yes; LEP = limited English proficient; SPED = special education. Robust clustered (school-year) standard errors in parentheses. Results exclude students who are not observed for both eighth and ninth grades. In Table 5, columns 4 and 8 are consistent with Table 2, Specification 4. All models in Table 6 are consistent with Table 2, Specification 4. Models include controls for student demographics and educational needs; these coefficients, along with the constant, are not shown.

$^{***}p < .01$, $^{**}p < .05$, $^{*}p < .1$.

Although we see no evidence of positive selection into UA based on ILS test scores in the year prior to joining UA, it is possible that schools on a trajectory of improvement are more likely to choose to participate in UA. If this is the case, there may be reverse causality, with test score improvements predicting UA participation. If UA schools are improving overall (potentially as a result of whole-school reforms, other programs, or increased teacher effectiveness), we would expect to see similar positive "impacts" of UA on math and ELA test scores. As a falsification test, we perform the analyses from Models 5 and 6 using math and ELA test scores (*z*-scores) as dependent variables (Table 4), and results suggest that UA has no systematic effect on eighth-grade math or ELA performance. In some specifications, we estimate a positive and statistically significant effect of UA in the schools' first year of UA, but this does not persist in subsequent years. This suggests that UA participation is not simply a proxy for overall school improvement and provides further support for our estimates of the impact of UA on science achievement.

Long-Term Outcomes: High School Outcomes

The results from the linear probability models are presented in Tables 5 and 6. As with our previous models, the sample includes students in eighth grade between 2004 and 2010, and we follow these students to their ninth-grade schools to measure ninth-grade outcomes (enrolling in an STEM school, Regents taking, and Regents performance). Thus, the long-term outcomes for students who were eighth graders in 2010 are measured in 2011. Because these models require ninth-grade outcomes, we exclude students who exit the NYC public schools after eighth grade.

Table 5 provides results for models estimating the impact of attending a UA school on the probability of enrolling in an STEM high school. The table reports results from four specifications, which are the same as described previously (Models 1 through 4). Overall, there does not seem to be a systematic impact of attending a UA school on the probability of enrolling in either an all-STEM or a partial-STEM high school.[19]

It is not entirely surprising that there are no significant effects of UA on the type of school a student attends for high school, as many school-level factors that are not likely to be correlated with UA participation may be important to students' high school placement. For example, school guidance counselors, parent coordinators, and other supports to eighth-grade students and parents can have great influence over the high school choice and placement process. Also, many middle schools tend to send students to specific types of high schools, consequently acting as de facto feeder schools. Thus, students may choose middle schools based in part on where they want to enroll in high school, which makes the middle school to high school transition endogenous. Additionally, we only observe the high school students ultimately attend—not their preferences. In NYC's complicated high school choice system, many factors, including but not limited to student preferences, are used to match students to high schools.

Table 6 shows the effect of attending a UA school on taking and passing the Earth Science Regents exam in eighth or ninth grade, using Model 4 (as previously described).[20] In column 1, we see a positive effect (0.063) of attending a UA school on the probability of taking the Earth Science Regents exam. Columns 2 through 4 show that attending a UA school also has a positive and statistically significant effect on the probability of passing the Earth Science Regents exam at all three cut points, although the effect is larger for passing at the 55 and 65 cut points (0.048 and 0.043, respectively) than for the 85 cut point (0.010). When we restrict the sample to just those students who took the test, there is no statistically significant impact of UA on passing the Earth Science Regents.

Taken together, these results suggest that UA induces more students to take the exam, but by definition these are marginal students. When considering their absolute probability of passing, these students are more likely to pass the exam, but this effect is driven by the increased likelihood of taking the exam. Of students who take the exam, students who attended UA schools in eighth grade do no better or worse than students who attended non-UA schools.

Results are qualitatively similar but weaker for the effect of UA on taking and passing the Living Environment Regents exam. We find a small positive effect of UA on taking the Learning Environments Regents exam (0.024) but no consistent effect of UA on passing the exam. Thus, attending a UA school slightly increases students' probability of taking the exam without affecting their probability of passing it.[21] For all long-term outcome models (enrolling at an STEM high school and science Regents outcomes), subgroup analyses revealed similar results for students grouped by race, gender, education needs, and poverty.[22]

Although our results find some effect of attending a UA school on taking science Regents exams in eighth or ninth grade, there are several important factors that may influence Regents-taking behavior that are not included in our models. First, because of the regulations around licensing and who is eligible to teach a Regents-level course, not all middle schools can offer science Regents exams to their students. This is a structural factor that is not likely to be influenced by UA participation.[23] Additionally, the high school attended has a role in Regents-taking and passing behavior, and this is unobserved in these models.

Conclusion

Despite the growing consensus that collaboration between formal and informal education organizations is an important component of improving science education in the United States, there are still relatively few examples of ongoing, intensive collaborations between schools and ISEIs, and there has been little research on the impact of such partnerships. This study provides the first estimates of the impact of a formal/informal science program on academic achievement and finds that exploiting the "urban advantage" with collaborations between formal and informal education institutions can be an effective way to improve science education in urban schools.

This study was made possible through an ongoing, productive collaboration with staff at UA and the American Museum of Natural History (AMNH). Together, we have developed a long-term research partnership and established a multiyear research agenda that relies on rich program data and the expertise of program staff. By talking extensively with program staff and UA teachers, undertaking qualitative explorations of program implementation, and analyzing both program and administrative data, we have been able to develop a strong understanding of UA, which is critical to our empirical analysis.

Through this collaboration, we have created a longitudinal data set that allows for a stronger identification strategy to estimate effects than is typical in program evaluation. By combining UA program data with administrative data from the NYCDOE, we have created a data set with several features that are rare for science program evaluations. We have a large sample size, the program itself operates on a large scale, we follow schools and students over several years, and we have several measures of academic outcomes, including science test scores in both middle and high schools.

In short, we find evidence that UA improves performance in science, as students who attend UA schools in eighth grade have higher performance on the New York State eighth-grade science exam. Our estimated impact of approximately 0.05 *SD*s reflects small improvements in science achievement; for comparison, results from the Tennessee STAR (student–teacher achievement ratio) experiment indicate that reducing class size from 22–26 to 13–17 students increased third graders' science test scores by 0.05 to 0.1 *SD*s (Konstantopoulos & Chung, 2009). The UA program, however, is substantially less costly to implement than these sizeable reductions in class size or other intensive interventions. Furthermore, our estimated effect likely underestimates the true program effect, as we have taken a conservative approach in defining the UA treatment. As described previously, data constraints require that we define the treatment variable at the school level, and so some students who do not have UA teachers will be in the treatment group (they attend a UA school but do not necessarily have a teacher who participates in UA).

We do not observe a consistent "effect" on math or ELA test scores, suggesting our impact estimates reflect a true program effect and are not merely reflecting coincident overall school improvement. Exploratory subgroup analyses find that the impact is largest for Black students and those in special education, which is particularly heartening, as these students often struggle in science. Additionally, we find that UA has no significant impact on whether a student enrolls in a STEM high school, but there is evidence that students who attend UA schools are slightly more likely to take a science Regents exam in eighth or ninth grade.

Although UA was initially designed for the NYC context, the program is more broadly generalizable to other cities with at least one informal science education institution, such as a zoo, garden, or science museum. The UA model can be successfully replicated in other locations, and in fact several other cities have implemented or are in the process of adapting a UA program for their community. For example, with support from AMNH and the National Science Foundation, Denver launched a UA program in the 2011 school year; this partnership brings together three Denver-area public school districts with the Denver Museum of Nature and Science, the Denver Zoo, and the Denver Botanic Gardens. Currently, there are discussions about developing a UA program for Boston, and UA has also been contacted by museums in Israel to learn more about the program.

Although this study uses rigorous empirical methods to estimate the impact of the UA program overall, it is important to note that we cannot identify the causal mechanism through which UA affects student science achievement. UA is a multifaceted program, and in this study, we cannot identify the particular aspects of the UA program that are most important for student outcomes. For example, UA may increase student science achievement by building the human capital and content knowledge of science teachers, by increasing the resources available for science education in schools, or by exposing students to ISEIs. Likely, multiple components of the program are important, and different schools utilize them in individualized ways to meet the needs of their teachers and students. A recent mixedmethods study identified several school-level factors associated with success in implementing UA, such as teacher collaboration, the concentration of UA teachers in a school, and administrative support (Weinstein, Whitesell, & Leardo, 2013).

A key limitation of this study is the scope of the outcomes we are able to measure. Although this study makes a significant and meaningful contribution to the literature on science interventions and partnerships by estimating effects on test scores, our academic outcomes present a somewhat limited picture of UA's impact. We are unable to provide evidence of UA's effect on other important student outcomes (e.g., interest in science, inquiry skills) or teacher practices and beliefs (e.g., mindsets, satisfaction, pedagogical skills).

An additional empirical limitation of our analysis is the inability to match students to teachers or to measure the percentage of science teachers in a school who are participating in UA. Many students who do not have UA teachers are included in the treatment group, as they attend a school where at least one teacher participates in UA. If UA program effects are stronger for students whose teachers participate in the program (as compared to students in the same schools whose teachers are not in UA), our estimates will understate the effect of having a UA teacher. Ongoing research will build on this work by exploring the relationship between program implementation (such as the concentration of UA teachers within schools and the extent to which schools utilize vouchers to ISEIs) and program effects.

Acknowledgments

The authors would like to thank the staff of Urban Advantage at the Gottesman Center for Science Teaching and Learning at the American Museum of Natural History and the Research and Policy Support Group for their help with this study. Major public support for Urban Advantage is provided by the Speaker and the City Council of New York and the New York City Department of Education.

Authors' Note

The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is supported by Urban Advantage; major public support for Urban Advantage is provided by the Speaker and the City Council of New York and the New York City Department of Education. The research reported here was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B080019 to New York University.

Notes

1. Middle schools are defined as all schools with an eighth grade.
2. Urban advantage (UA) does not accept all schools or teachers who apply to be in the program. Current UA schools as well as schools that are hoping to join the program submit applications to UA, and UA program staff consider factors such as how many teachers in the program have been or are planning to participate in UA and how frequently schools have been or anticipate being able to use vouchers to attend informal science education institutions.
3. In its first year of operation (2005), UA was open to eighth-grade teachers. UA expanded to include seventh-grade teachers in 2006 and sixth-grade teachers in 2010.
4. For example, UA introduced the Investigation Design Diagram tool to teachers in 2007–2008 and the Developing Science Exploration Tool in 2010.
5. The percentage of UA teachers who are continuing in the program has increased over time: 2005 (0%), 2006 (32%), 2007 (45%), 2008 (50%), 2009 (76%), 2010 (53%), 2011 (77%), 2012 (82%), and 2013 (70%).
6. Lead teachers were introduced as a component of the UA program in 2005–2006; a separate Leadership Institute was initiated in 2008–2009. We are unable to measure the extent to which lead teachers (or experienced UA teachers more generally) provide support to other teachers in their schools; to the extent that some lead teachers help support non-UA teachers in their schools, it is possible that there is UA program spillover from UA teachers to their non-UA colleagues.
7. The database is housed at New York University's Institute for Education and Social Policy and is updated annually with data from the New York City Department of Education (NYCDOE).
8. NYC does not assign students to high schools based on location, but rather students are matched to high schools using a complicated ranking system that takes into account student preferences and qualifications.
9. Although we are able to identify whether or not a student enrolls in a partial-STEM school, we are not able to determine whether a student attends a science-focused academy within a partial-STEM school.

10. Prior to 2005, students were required to earn a score of 55 or higher to count a Regents exam toward earning a local diploma, while 65 was the score needed for a Regents diploma or an Advanced Regents diploma. Now, earning a Regents diploma is the minimum requirement; this means that 65 is the lowest passing score. Earning an 85 on a Regents exam is used as a cutoff for admission to certain selective colleges and universities.
11. We are not able to match students to their science teachers because of limitations in the data collected by the NYCDOE. Over the time period in our analysis, there is only one year (2010) in which the DOE provided a reliable match of students to science teachers.
12. Unfortunately, we do not have accurate data on the percentage of teachers in each school who participate in UA. In the early years of UA, the program did not systematically collect consistent program information on teachers and schools, and neither teachers' movements between schools nor teachers' movement in and out of program participation were accurately reflected in UA program data. For more recent years, UA does have accurate information on UA teachers and schools, and we are planning to use these data in future studies.
13. UA and non-UA schools also have similar exam-taking rates; for both groups of schools, taking rates are high (about 95% for math, 93% for English-language arts, and 90% for eighth-grade science).
14. Results for all years are similar and available from the authors.
15. A school's passing rate is the percentage of eighth-grade students who scored a 3 or 4 out of 4 on the intermediate-level science exam.
16. Results are not shown but are available from the authors.
17. Results are not shown but are available from the authors.
18. Results are not shown but are available from the authors.
19. We do not find a statistically significant effect for either outcome when we used our preferred specification (Model 4).
20. We do not estimate specifications distinguishing between the base year and postyears, as we find small or null average effects of UA in less inclusive specifications.
21. Results are not shown but are available from the authors.
22. Results are not shown but are available from the authors.
23. For a student to be eligible to take a specific science Regents exam, the teacher must be certified in that content area. That is, only teachers certified in earth science can teach an Earth Science Regents course; only teachers certified in biology can teach a Living Environment Regents course. If a middle school science teacher is only certified in general science or in general middle school instruction, her students are not eligible to take the Earth Science or Living Environment Regents exam.

References

Anderson, D., Kisiel, J., & Storksdieck, M. (2006). Understanding teachers' perspectives on field trips: Discovering common ground in three countries. *Curator: The Museum Journal*, *49*, 365–386.

Astor-Jack, T., Balcerzak, P., & McCallie, E. (2010). Professional development and the historical tradition of informal science institutions: Views of four providers. *Canadian Journal of Science, Mathematics, and Technology Education*, *6*, 67–81.

Darke, K., Clewell, B. C., & Sevo, R. (2002). Meeting the challenge: The impact of the national science foundation's program for women and girls. *Journal of Women and Minorities in Science and Engineering*, *8*, 285–303.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*, 81–112.

DeWitt, J., & Storksdieck, M. (2008). A short review of school field trips: Key findings from the past and implications for the future. *Visitor Studies*, *11*, 181–197.

Dorsen, J., Carlson, B., & Goodyear, L. (2006). *Connecting informal STEM experiences to career choices: Identifying the pathway*. ITEST Learning Resource Center. Retrieved from http://stelar.edc.org/sites/stelar.edc.org/files/itestliteraturereview06.pdf

Fadigan, K. A., & Hammrich, P. L. (2004). A longitudinal study of the educational and career trajectories of female participants of an urban informal science education program. *Journal of Research in Science Teaching*, *41*, 835–860.

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, *19*, 643–658.

Fullan, M., & Stiegelbauer, S. (1991). *The new meaning of educational change*. New York, NY: Teachers College Press.

Hein, G. E. (1998). *Learning in the museum*. London, England: Routledge.

Hofstein, A., & Rosenfeld, S. (1996). Bridging the gap between formal and informal science learning. *Studies in Science Education*, *28*, 87–112.

Inverness Research Associates. (1996). *An invisible infrastructure: Institutions of informal science education*. Washington, DC: Association of Science-Technology Centres.

Kisiel, J. F. (2010). Exploring a school-aquarium collaboration: An intersection of communities of practice. *Science Education*, *94*, 95–121.

Konstantopoulos, S., & Chun, V. (2009). What are the long-term effects of small classes on the achievement gap? Evidence from the lasting benefits study. *American Journal of Education*, *116*, 125–154.

Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011, July). *STEM: Good jobs now and for the future* (Issue Brief #03-11). Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics*, *9*, 258–271.

Loucks-Horsley, S., Stiles, K. E., Mundry, M. S. E., Love, N. B., & Hewson, P. W. (2009). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.

National Assessment of Educational Progress. (2011). *2009 and 2011 science assessments*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

National Center for Education Statistics. (2010). *Status and trends in the education of racial and ethnic groups*. Washington, DC: Department of Education. NCES 2010-015, Table 9.1.

National Governors Association. (2011). *Building a science, technology, engineering, and math education agenda: An update of state actions*. Washington, DC: NGA Center for Best Practices.

National Research Council. (2000). *Educating teachers of science, mathematics, and technology: New practices for the new millennium*. Washington, DC: The National Academies Press.

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

Next Generation Science Standards. (2013, June). The next generation science standards: Executive summary. Retrieved from http://www.nextgenscience.org/sites/ngss/files/Final%20Release%20NGSS%20Front%20Matter%20-%206.17.13%20Update_0.pdf

Organisation for Economic Co-operation and Development. (2014). PISA 2012 results in focus. Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Phillips, M., Finkelstein, D., & Wever-Frerichs, S. (2007). School site to museum floor: How informal science institutions work with schools. *Journal of Science Education*, *29*, 1489–1507.

Ramey-Gassert, L., Walberg, H. J., & Walberg, H. J. (1994). Reexamining connections: Museums as science learning environments. *Science Education*, *78*, 345–363.

Saxton, L. J., Gupta, P., & Steinberg, R. N. (2010). CLUSTER: University-science center partnership for science teacher preparation. *The New Educator*, *6*, 280–206.

Short, J. B., Elgendy, S., Roditi, H., & Holmes, J. (2012). *History and development of the Urban Advantage program*. Presented at the 2012 Annual Meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, *37*, 963–980.

The White House, Office of the Press Secretary. (2011, January 25). Fact sheet: The state of the Union: President Obama's plan to win the future. Retrieved from http://www.whitehouse.gov/the-press-office/2011/01/25/fact-sheet-state-union-president-obamas-plan-win-future

Thompson, C. L., & Zeuli, J. S. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession. Handbook of policy and practice* (pp. 341–375). San Francisco, CA: Jossey-Bass.

University of the State of New York, State Education Department. (2005, February). Overview of performance in English language arts, mathematics, and science and analysis of student subgroup performance for public schools in New York State. Retrieved from http://www.p12.nysed.gov/repcrd2004/statewide/2004statewideoverview.pdf

Weinstein, M., Whitesell, E. R., & Leardo, M. (2013). *Successful schools: How school-level factors influence success with Urban Advantage* (Working Paper #01–14). New York, NY: Institute for Education and Social Policy.

Author Biographies

Meryle Weinstein is Executive Director at New York University's Institute for Education and Social Policy and Research Assistant Professor of Education Policy at NYU's Steinhardt School. Her research interests are in the areas of the evaluation of education and social programs, including the role of out-of-school time and community resources. Recent topics include summer school programs, summer youth employment, and school food policy.

Emilyn Ruble Whitesell is a doctoral candidate at New York University's Wagner Graduate School of Public Service and a graduate research fellow at NYU's Institute for Education and Social Policy. Her research focuses on education policy and program evaluation. Current topics include school accountability, stakeholders' perceptions of schools, and formal-informal educational partnerships.

Amy Ellen Schwartz is Daniel Patrick Moynihan Professor of Public Affairs at Syracuse University's Maxwell School and Director of the Institute for Education and Social Policy at NYU's Steinhardt School. Her research uses applied econometrics to explore issues in education policy, urban policy, and public finance. Recent topics include the effects of neighborhoods, housing, mobility, summer employment, and school reforms on student performance, childhood obesity and academic success.