Session Title:
Looking for Early Evidence of Implementation: Using the Lens of Propensity Score Matching

MSP Project Name:
Math and Science Partnership in New York City (MSPinNYC)

Presenters:
Howard Everson, Graduate School, CUNY
Sarah Bonner, Hunter College, CUNY
Ally Stevens Thomas, Graduate School, CUNY

Authors:
Howard Everson, Graduate School, CUNY

Project Session

Strand 3

Summary:
Our session relates to the Evaluation, Research, and Implementation strand. MSPinNYC2 restructures high school STEM courses using peer-enabled restructured classrooms (PERC) to improve achievement. In year one 711 students enrolled in Integrated Algebra, Biology, and Chemistry courses across four public high schools in NYC. Trained Teaching Assistant Scholars (n=234), mostly high school sophomores, collaborate with teachers and tutor PERC students in class. Our session will demonstrate the utility of propensity score matching methods for developing early evidence of program impact and efficacy—showing how differences in prior achievement and academic experiences between TAS and non-TAS students contribute to variation in college readiness; and how prior achievement and instructional experiences between PERC and non-PERC students contribute to variation in achievement.

Section 1: Questions framing the session:
Our awardee-led project session will focus on asking how best to develop early evidence of program implementation and impact. Specifically, we are seeking early evidence (after 1 year of implementation) of the effects of *MSPinNYC2* on high school students' achievement—both in terms of course grades and scores on end-of-course tests in three key STEM disciplines: Integrated Algebra, Biology, and Chemistry. Using an evidence-based approach—*propensity score matching*—we ask if, in its early stages, the program is making a difference in students' academic achievement and college readiness? We also ask which program aspects (i.e., course design and domain, the use of TAS in the classroom, the teachers, and the schools) contribute to differences in students' achievement?

Section 2: Conceptual framework:
Our proposed session relates directly to the *Evaluation, Research, and Implementation* conference strand. The *MSPinNYC2* program restructures early high school STEM courses to include 6-8 Teaching Assistant Scholars (TAS) who, along with the teachers, facilitate daily in-classroom group work. Early pilot studies suggested the model of peer-enabled restructured classrooms (PERC) increases student achievement and closes the achievement gap in high school STEM courses. The *MSPinNYC2* program also attempts to promote TAS college readiness by providing mentoring and a supportive pipeline from high school-to-college.

In year one more than 700 students ($n = 711$) were enrolled in three STEM courses—Integrated Algebra, Biology, and Chemistry across four public high schools in NYC. The TAS ($n = 234$), who serve as in class tutors, are average-achieving students who passed the course and the end-of-course exam during the previous year at the same high school. They are trained by the Program staff and in a dedicated TAS class at their schools to work collaboratively with teachers and students to improve achievement.

Our aim in this proposed session is to develop and present a mix of year one descriptive data, and results of statistical matching methods (propensity score matching) designed for use in observational studies, to address the question of how we know if the program is making a difference in terms of student achievement. Propensity score matching (PSM) had its origins nearly three decades ago in biomedical research for reducing estimation bias when comparing non-equivalent groups, and for drawing causal inferences in observational study designs (i.e., studies where random assignment to treatment is not possible). Education researchers have been slow to adopt PSM largely because of a lack of familiarity with the methodology and few examples of how PSM is used in educational program evaluation research, where random assignment to instructional interventions is not feasible. Our presentation will provide an example of how PSM can be used to monitor the efficacy of a large, multi-site instructional intervention like *MSPinNYC2*.

Section 3: Explanatory framework:
In its inaugural year the *MSPinNYC2* project recruited 234 TAS and over 700 students in four New York City public high schools to participate in PERC classes. The students served by the project are not the academically elite. Less than half (48%) the TAS, for example, were proficient in math based on their 8th grade NCLB tests, and only about one-third (36%) were proficient in English. The students enrolled in the PERC classes had a similar profile: about 26% were proficient in math, and roughly one in five was proficient in English. Tables 1 and 2 show the distribution of TAS and PERC students during the first year by subject and high school.

Table 1. Distribution of TAS by Academic Subject & High School.

	Subject			Total
	Biology	**Integrated Algebra**	**Chemistry**	
E. Bronx Academy	0	25	0	25
Harry S.Truman H.S.	51	71	21	143
H.S.of Hospitality Management	17	0	0	17
Hillcrest H.S.	21	28	0	49
Total	89	124	21	234

Table 2. Distribution of PERC Students by Academic Subject & High School.

	Subject					Total
	Biology	**Integrated Algebra**	**Chemistry**	**Biology & Chemistry**	**Biology & Integrated Algebra**	
E. Bronx Academy	0	72	0	0	0	72
Harry S.Truman H.S.	95	109	105	1	179	489
H.S.of Hospitality Management	25	0	0	0	0	25
Hillcrest H.S.	46	22	0	0	57	125
Total	166	203	105	1	236	711

The courses were taught by eleven high school teachers—all trained in the PERC model by the Program's staff. Each course served between 25-30 students who were tutored in-class by 6 or 7 TAS. The Program's goal in the first year was to increase the end-of-course Regents tests' passing rates in all PERC classes above citywide baselines by 10%. The achievement target for the TAS was improved readiness for college, defined as earning *mastery* scores (i.e., > 80) upon re-taking the end-of-course Regents exams in the subjects they tutored (i.e., Integrated Algebra, Biology (Living Environment), and Chemistry). In collaboration with both the NYC Education Department and the *College Now Program* at CUNY, the *MSPinNYC2* Research and Evaluation team are building a longitudinal database to track students' academic progress. This database was used to develop the findings presented next.

Preliminary Findings. We begin by summarizing the achievement outcomes of the TAS, and then turn to the 9th grade PERC students. Of the 124 TAS who served in the *Integrated Algebra* classes, all achieved proficiency (a score of 65 or higher) on the Algebra Regents end-of-course exam, and by year's end, after re-taking the exam, 37% achieved mastery (a score of 80 or higher). The PERC Biology courses were served by 89 TAS, nearly all (95%) were proficient (a score > 64) on the Living Environment (Biology) Regents exam; after re-taking the exam at year's end, 58% achieved a mastery score (80 or higher). The PERC Chemistry course was as a pilot at only one high school.

Twenty-one TAS tutored in the three Chemistry classes. Of them, 19 (90%) achieved a proficient score on the rigorous Regents Chemistry exam.

PERC Students. As noted earlier (see Table 1), 711 students enrolled in *MSPinNYC2* developed PERC courses in Integrated Algebra, Biology (listed as Living Environment), and Chemistry. Roughly one third took at least two PERC courses as 9th graders. (When reviewing end-of-course exam scores for these students, it is important to be mindful that two-thirds of them were classified as less than proficient in *both* English and math as 8th graders.) The pattern of end-of-course exam scores was as follows: (1) 258 students sat for the Living Environment exam (Biology), and 70% were proficient; 362 students took the Integrated Algebra exam and 50% were proficient; and only about one-third of the Chemistry students took the Chemistry Regents exam, and 21% were proficient. These achievement outcomes are impressive, and suggest the instructional interventions developed in the first year of *MSPinNYC2* are having a positive impact. When contrasted with city-wide Regents exam passing rates, which hover around 40% in Living Environment and are even somewhat lower in Integrated Algebra, the passing rates for the students in the first-year of the *MSPinNYC2* appear to be dramatically higher. Because the *MSPinNYC2* did not use an experimental design students were not randomly assigned to these courses. Analyses of student achievement outcomes using PSM are still underway but will be completed in the time for presentation at the conference.

Section 4: Discussion:
When the *MSPinNYC2* was implemented it was not feasible, understandably, to use an experimental design that assigned students randomly to the PERC and non-PERC courses in these key STEM disciplines. Clearly we need to account for this limitation when attempting to make causal inferences about the instructional efficacy of the PERC model and assessing the fidelity of implementation of the initiative. Currently we are creating the matched samples and completing a series of PSM analyses in preparation for publication and dissemination.

With this in mind, our session will demonstrate the utility of *propensity score matching* methods for developing early evidence of program impact and efficacy. Because random assignment was not possible, the PERC students have to be compared or matched with New York City public school students who have similar prior academic achievement profiles. Using the comprehensive, city-wide database described earlier we will draw a matched sample of non-PERC students enrolled in Biology, Algebra and Chemistry courses in highly similar NYC public high schools. The students—TAS and non-TAS and PERC and non-PERC—will be matched on gender, race/ethnicity and 8th grade English and math standardized test scores, all variables presumed to influence Regents exam scores. Using logistic regression we will construct a propensity score that balances and summarizes the student-level information of this set of important covariates. The four groups of students will then be balanced on this propensity score and statistical comparisons will be conducted across the student groups. This PSM approach creates statistically equivalent groups and permits stronger, more defensible inferences about the efficacy of the *MSPinNYC2*. Our presentation will provide an example of how PSM can

be used to monitor the efficacy of a large, multi-site instructional intervention like *MSPinNYC2*.

Section 5: How will you structure this session? What is your plan for participant interaction?

We will illustrate the use of propensity score matching methods using student achievement data from matched samples of TAS ($n = 234$) and non-TAS high school students ($n = 300$), and a matched sample of students enrolled in PERC ($n = 711$) and non-PERC classes ($n = 1000$) in Integrated Algebra, Biology, and Chemistry. Our session will describe the available data set, and provide detail on how we identified the set of covariates used to create the propensity scores using logistic regression analysis. Our stratification approach for matching students will be described, and we will explain the statistical tests used to detect potential mean differences in achievement outcomes between the TAS and non-TAS students, and the PERC and non-PERC students. We intend to make available a modified propensity score matching data set for use by MSP colleagues, and provide participants with a codebook, statistical syntax in SPSS and STATA, and references for further exploration of PSM methods.