Abstract Title: Effects of a Teacher Professional Development Program on the Mathematics Achievement of Middle Level Students

MSP Project Name: Rocky Mountain – Middle School Math Science Partnership (RM-MSMSP)

Author(s): Sample McMeeking, Laura B., Orsi, Rebecca, Cobb, R. Brian, & Basile, Carole

Presenter(s): Orsi, Rebecca, Basile, Carole, & Cobb, R. Brian

120 word summary:

We developed and offered 20 integrated STEM content-pedagogy professional development courses to teachers in seven districts. Students who received instruction from a participating teacher *after* the teacher took one mathematics course had 1.3 times greater odds of at least a proficient outcome on our state math assessment than students who received instruction from the same teacher *before* the teacher took a math PD course. Such effects were even stronger for teachers took *two or more* mathematics PD courses. This finding is important: teaching mathematics teachers deeper content and how to use an inquiry-based approach to deliver that content translates into greater student proficiency in mathematics. Likewise, doing more of such PD with teachers confers greater effects on their students.

- Section 1: **Questions for dialogue at the MSP LNC**.

1. What are the research design implications inherent in large-scale professional development interventions such as the MSP's when the PD treatment is given at the teacher level and the PD achievement effects are measured at the student level?

2. For funded MSP research projects, how should the roles and functions of the researchers be distinguished from, and integrated with, the roles and functions of the evaluators? How "external" can/should the evaluation actually be?

- Section 2: **Conceptual framework**.

In the original approved proposal of the RM-MSMSP, we chose to use the Colorado Student Assessment Program (CSAP) test as a means of measuring the program's effects on student achievement. The CSAP is a standards-based assessment designed to measure Colorado student performance relative to the Colorado State Model Content Standards in the content areas that are assessed. CSAP tests contain both structured response and multiple-choice items. After taking the test, students receive a total scale score (CTB McGraw-Hill, 2007), which is then divided into four categories that represent the proficiency level of each student: Unsatisfactory, Partially Proficient, Proficient, and Advanced.

Scale scores present a fundamental problem when used for analysis because the range of scale scores associated with a specific proficiency level depends on grade. Thus, a student's scale scores can rise when proficiency does not. Since the student observations in the study span

multiple grades within each student observation, we chose to work with the four ordinal proficiency ratings in our analysis instead of with the scale scores in order to improve measurement reliability and statistical conclusion validity (Shadish, Cook & Campbell, 2002). The ordinal scores are more stable from year-to-year. Furthermore, they are the most relevant scores when considering whether schools have met their goals under the No Child Left Behind accountability law.

Our student success measurement design required collapsing the four-level proficiency outcome variable (unsatisfactory, partially proficient, proficient, and advanced) into a dichotomous outcome variable for two reasons. First, the most critical outcome for schools in Colorado under No Child Left Behind is whether or not a child achieves a "proficient" (or above) rating. Thus it is relevant to estimate whether or not teacher PD increases the odds of students achieving any outcome considered proficient or higher. Second, in order to enable the model to estimate the effect on the outcome variable while holding constant levels of the covariate(s), it was necessary that there be sufficient observations for each category of the outcome variable in combination with each category of the covariate. Because there were virtually no students who had a covariate value of "unsatisfactory" (the pre-test value) and an outcome of "advanced" (and vice versa) for a post-test value, we could not reliably fit a model using all four levels for both the covariate pre-test and the post-test outcome. Collapsing the outcome into two categories enabled a better fit of the model to the data. Our student success analytic design, then, involved comparing a PD teacher's students' proficiency ratings prior to the teacher receiving the PD treatment, with the teacher's students' proficiency ratings after receiving the PD treatment. We represented our findings in terms of comparing the relative odds of students being proficient or greater prior to their teacher receiving PD with the odds of such status after receiving the PD.

- Section 3: **Explanatory framework**.

We found that students who received instruction from a participant teacher after the teacher took one mathematics PD course had odds of at least a proficient outcome which were approximately 1.3 times higher than students who received instruction before the teacher took a math course. Students who received instruction after the teacher took two or more mathematics courses had odds of at least a proficient outcome which were approximately two times higher than students who received instruction before the teacher took any math courses. The second result is of great enough magnitude to be considered statistically, as well as practically, significant.

The methodological lessons of the study are important as well for future MSP work involving professional development. First, the professional development literature is replete with recommendations that effects of work with teachers be estimated through the learning of their students, but the methodological literature is much less robust with examples of how to do it. Our adaptation to the traditional cohort control design is now in the methodological literature (Sample McMeeking, Cobb & Basile, 2010), and we look for others to adapt it further in similar kinds of efforts. Second, we think our use of a logistic generalized linear mixed model (Agresti, 2007) offers a simple, straightforward, and policy-aligned way of representing the results of our work. We recognize that by eschewing traditional significance testing we may be pushing the boundaries for some in how we add our work to existing theory. Nonetheless, by expressing our findings in an "odds of improvement" metric, we believe we are offering a sense of

interpretability to policymakers and decision makers in educational environments that more traditional significance testing cannot deliver.

Third, from a greater design perspective, theory-based evaluation (which integrates evaluation and research) proved to be extremely effective in aligning the implementation and research aspects of the project. Finally, from a research perspective, however, the mobility of teachers impacted the research in a significant way. Even though we affected hundreds of teachers each summer through our professional development opportunities, tracking teachers proved to be extremely difficult through subsequent and following academic years thereby decreasing the number of teachers (and ultimately their students) in the sample. In addition, teachers were able to take courses in any sequence and/or across disciplines therefore making the fidelity of treatment tenuous.