



**Government
Performance and
Results Act (GPRA)
Indicators for the
MSP Program,
Performance Period
2010**

**Analytic and
Technical Support
for Mathematics
and Science
Partnerships**

Contract # ED-CFO-10-A-0024

July 11, 2012

Prepared for:
**Patricia O'Connell Johnson
Valerie Randall**

Submitted by:
Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02135

Government Performance and Results Act (GPRA) Indicators for the MSP Program, Performance Period 2010

The U.S. Department of Education's Mathematics and Science Partnerships (MSP) Program creates partnerships between high-need school districts and mathematics, science, and/or engineering departments at institutions of higher education for the purpose of providing intensive content-rich professional development to teachers and thus improving student achievement in mathematics and/or science. The program requires projects to evaluate the impact of participation in MSP professional development on gains in teacher content knowledge and on student achievement.

Under the Government Performance and Results Act (GPRA), all federal agencies are required to develop indicators in order to report to the U.S. Congress on federal program impacts and outcomes. For the MSP Program, the following indicators have been developed:

- ***Teacher Knowledge***
 1. The percentage of MSP teachers who significantly increase their content knowledge as reflected in project-level pre- and post-assessments.
- ***Student Achievement***
 2. The percentage of students in classrooms of MSP teachers who score at the basic level or below in state assessments of mathematics or science.
 3. The percentage of students in classrooms of MSP teachers who score at the proficient level or above in state assessments of mathematics or science.
- ***Evaluation Design***
 4. The percentage of MSP projects that report using an experimental or quasi-experimental design for their evaluations.
 5. The percentage of MSP projects using an experimental or quasi-experimental design for their evaluations whose evaluations are conducted successfully and yield scientifically valid results.
- ***Timeliness***
 6. The percentage of state education agencies (SEAs) that submit complete and accurate data on MSP performance measures in a timely manner.

Data on each of these GPRA indicators for the MSP Program for Performance Period 2010 (PP10) are presented in the sections below. Report data were analyzed from 566 MSP projects, serving a total of 43,755 participants, including elementary, middle, and high school teachers, coaches, paraprofessionals, and administrators.

Teacher Knowledge

1) The percentage of MSP teachers who significantly increase their content knowledge, as reflected in project-level pre- and post-assessments.

As part of their evaluations, MSP projects are required to assess teachers' content knowledge in mathematics and/or science via pre- and post-tests during the years in which they receive intensive professional development. Projects reported the number of MSP teachers who significantly increased their content knowledge in mathematics and/or science topics on project pre- and post-assessments. Exhibit 1 presents data for those teachers for whom gains in content knowledge were assessed. Among the teachers assessed, 65 percent of teachers showed significant gains in mathematics content knowledge and 74 percent of teachers showed significant gains in science content knowledge.

Exhibit 1: Percent of Teachers with Significant Gains In Content Knowledge, Among Teachers with Pre-Post Content Assessments, Performance Period 2010

Content Area	Total number of teachers served	Percent of teachers with content assessments	Percent of assessed teachers with significant gains
Mathematics	25,344	53%	65%
Science	19,562	60	74

Note: Individual teachers who received multiple professional development courses may be counted multiple times

Student Achievement

2) The percentage of students in classrooms of MSP teachers who score at the basic level or below in state assessments of mathematics or science.

3) The percentage of students in classrooms of MSP teachers who score at the proficient level or above in state assessments of mathematics or science.

Projects also reported the number of students served, the number of students assessed, and the number of students scoring at the basic level or below and at the proficient level or above in state assessments in both mathematics and science. In Exhibit 2, it can be seen that nearly 1.3 million students were taught by teachers who received MSP professional development in mathematics, and approximately 900 thousand students were taught by teachers who received MSP professional development in science. State assessment data were reported for nearly two-thirds of students (64 percent) in mathematics and for 39 percent of students in science. In both mathematics and science, approximately two-thirds of students scored at proficient or above (65 percent in mathematics and 67 percent in science). The remaining students scored at basic or below. These numbers were aggregated across all grade levels and all schools with teachers in the MSP project.

Exhibit 2: Percent of Students Scoring at Proficient Level or Above, Among Students Taught by MSP Teachers And Assessed In Each Content Area, Performance Period 2010

Content Area	Total number of students taught by MSP teachers	Percent of students with assessment data	Percent of assessed students at <i>basic</i> or below	Percent of assessed students at <i>proficient</i> or above
Mathematics	1,280,438	64%	35%	65%
Science	903,788	39	33	67

Evaluation Design

4) The percentage of MSP projects that report using an experimental or quasi-experimental design for their evaluations.

Exhibit 3 presents the percentages of MSP projects that reported using various types of evaluation designs in PP10. Three percent of projects reported that they implemented an experimental design, which is the most rigorous research design for testing the impact of an intervention, wherein schools, teachers, or students are randomly assigned to treatment or control groups.

Forty-five reported using a quasi-experimental design to compare the effects of the MSP Program on participating teachers and/or their students to non-participating teachers and/or students. Just over one-fourth of projects (27 percent) used a matched comparison group design, which attempts to show causality by demonstrating equivalence between groups at baseline or adjusting for any initial differences between groups. Eighteen percent of projects reported using a non-matched comparison group.

The remaining projects reported using a one-group design with no comparison group, a qualitative design, or a mixed methods design. Thirty-eight percent of projects reported using pre-tests and post-tests to assess the gains of the teachers served by MSP (a “one-group” design). Twelve percent of projects reported using primarily qualitative methods, and three percent of projects reported using a mix of quantitative and qualitative methods.

Exhibit 3: Types of Evaluation Designs Used by Projects, Performance Period 2010

Evaluation Design	Percent of Projects
Random assignment design (experimental)	2%
Quasi-experimental design	45
<i>Matched comparison groups</i>	27
<i>Non-matched comparison groups</i>	18
One-group design	38
Qualitative / descriptive design	12
Mixed methods	3

5) The percentage of MSP projects using an experimental or quasi-experimental design for their evaluations whose evaluations are conducted successfully and yield scientifically valid results.

Every MSP project is required to design and implement an evaluation and accountability plan that allows for an assessment of its effectiveness. The requirement for evaluation of MSP projects is specified in the program’s enabling legislation in the No Child Left Behind Act. In order to ensure that projects are providing high-quality information on program outcomes, a rubric was developed as part of the Data Quality Initiative (DQI) through the Institute for Education Sciences (IES) at the U.S. Department of Education. The rubric was revised in 2011 and 2012 in order to bring it further into alignment with the What Works Clearinghouse Standards. The four criteria that comprise the rubric, as shown below, specify the conditions for projects using experimental and quasi-experimental designs to be deemed successful evaluations that yield scientifically valid results.

1. **Data reduction rates (i.e., attrition rates, response rates).** This criterion was not relevant to quasi-experimental designs that present evidence of baseline equivalence of the analysis sample. Experimental designs were required to meet two criteria:
 - Overall attrition is less than 30 percent; AND
 - The difference in attrition rates between the groups being compared is 15 percent or less.

2. **Baseline equivalence of groups.** Experimental designs that met the attrition criteria above were not required to establish baseline equivalence. Experimental designs that had high attrition (i.e., did not meet the attrition criteria above) were required to meet the same criteria as quasi-experimental designs, as specified below.
 - Baseline equivalence is established when:
 - The mean difference in baseline measures in the groups being compared is less than or equal to 5 percent of the pooled sample standard deviation; OR
 - The mean difference in baseline measures in the groups being compared is more than 5 percent but less than or equal to 25 percent of the pooled sample standard deviation and the differences between the groups are adjusted for in the analysis.
 - Baseline equivalence should be established in the analysis sample on pre-treatment measures of the outcome variable, or on other variables that are highly correlated with the outcome variable;¹ OR

¹ WWC Quick Review Protocol, Version 2.0:
<http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=28&tocId=1>.

- If the data required for establishing baseline equivalence in the analysis sample are missing (and there was evidence that equivalence was tested), then baseline equivalence for quasi-experimental designs can be established in the baseline sample provided the data reduction standards in the first criterion are met as well.
3. **Quality of the measurement instruments.** Evaluations met this criteria if they met one of three criteria:
 - Data collection instruments that had already been deemed valid and reliable to measure key outcomes; OR
 - Data collection instruments developed specifically for the study were sufficiently pre-tested with subjects who were comparable to the study sample; OR
 - Data collection instruments contained selected items from validated and reliable instrument(s) and the resulting instrument included at least 10 items and at least 70 percent of the items were from validated and reliable instrument(s).
 4. **Relevant statistics reported.** Evaluations met this criteria when the final report included treatment and comparison group post-test means and tests of statistical significance for key outcomes, sufficient information for calculation of statistical significance (e.g., mean, sample size, standard deviation/standard error), or results from statistical models that have been clearly specified.

The evaluations of all final year MSP projects that reported using an experimental or quasi-experimental design with a comparison group were reviewed by a team of reviewers according to this rubric to determine the number of projects that conducted successful evaluations yielding scientifically valid results.²

Most evaluations of MSP projects included multiple evaluations of various outcomes. The review considered outcomes of teacher content knowledge, teacher practices, and student achievement. Projects had a passing evaluation if any of their evaluations of one of these three outcome domains passed all four criteria listed above.

The review team focused primarily on the information contained in the final evaluation reports. Information was supplemented with the evaluation data in the annual performance reports (APRs), as well as information provided directly by projects, in an attempt to fill in missing information and to verify consistent reporting of measures.

Two hundred and eighteen projects reported submitting a final year report in PP10. These were screened for projects that used experimental or quasi-experimental evaluation designs in one of the three designated outcome domains and fifty-nine projects were reviewed and evaluated on the criteria.

² Projects that reported using unspecified “other” designs were screened in order to determine whether they used experimental or comparison group designs. Those that did were also included in the group of projects reviewed.

Fifteen³ of these projects conducted an evaluation that met all of the design criteria specified in the rubric. Ten projects successfully employed quasi-experimental designs that included comparison groups, and three projects successfully implemented experimental designs.

Exhibit 4: Final Year Projects

Projects	Quasi-experimental design	Experimental design	Total
Conducted an experimental or quasi-experimental design with a comparison group in a designated outcome domain	54	5	59
Included at least one evaluation that passed all rubric criteria	12	3	15

Among the fifteen projects that conducted successful experimental or quasi-experimental evaluations, thirteen projects successfully studied their program’s impacts on teacher content knowledge, four projects successfully studied impacts on classroom practices, and five projects successfully studied impacts on student achievement.

Timelines

6) The percentage of SEAs that submit complete and accurate data on MSP performance measures in a timely manner.

Submission guidelines for APRs were developed as a basis for the timeliness calculation. MSP State Coordinators were responsible for ensuring that all projects within their state submitted complete and accurate data by this date. APRs for PP10 were accepted until February 29, 2012. Projects that informed the Department of Education that they would not receive teacher and/or student data in time were given an extension on the due date of their reports. All states submitted the required APR reports on time. Thus, 100 percent of states submitted complete and accurate data on MSP performance measures in a timely manner.

³ In PP09, 16 projects conducted at least one evaluation that passed all rubric criteria. However, there were fewer projects funded in PP10 and thus fewer final-year reports submitted. Among projects that used an RCT or QED design with an appropriate comparison group, the percent of passing projects in both PP09 and PP10 was 25%.