



NEW YORK STATE MATHEMATICS SCIENCE PARTNERSHIP (MSP)

Evaluation and Reporting Guide 2009-2010

Prepared by:

Compass Consulting Group, LLC
5726 Fayetteville Road, Suite 202
Durham, NC 27713
Phone: 919.544.9004
Fax: 919.321.6997

www.CompassConsultingGroup.org



INTRODUCTION

The purpose of this guide is to outline the expectations for MSP project evaluations based on the Federal and New York State Education Department (NYSED) requirements for monitoring and reporting. These expectations include:

- ⦿ **Collecting and Tracking Service Statistics:** Collect and track the number of professional development hours completed by each participant and activity.
- ⦿ **Measuring Teacher Content Knowledge:** Administer pre- and post-content knowledge test to all teachers who participate in MSP activities (and comparison teachers if your evaluation design includes any) to assess changes in teacher content knowledge.
- ⦿ **Assessing Teacher Pedagogical/Teaching Skills:** Use methods that go beyond self-report to assess changes in teacher pedagogical/teaching skills (e.g., observations, Surveys of Enacted Curriculum).
- ⦿ **Assessing Student Achievement:** Obtain and analyze mathematics or science testing data annually for the students in classes of teachers participating in MSP activities and conduct an experimental or quasi-experimental design to assess how MSP activities lead to changes in student achievement.

Compass Consulting Group has been contracted by the NYSED to help projects develop and implement evaluations that fulfill these requirements and to summarize evaluation information across projects. To facilitate this process, each project has been assigned a Compass evaluator with whom they will work over the course of the grant. Compass evaluators will meet with projects once or twice annually and be available via phone and e-mail. For the name and contact information of the Compass evaluator assigned to your project, please locate your MSP below.

- ⦿ **Dr. Anne D'Agostino, anne-d@compasseval.com, 877-652-0189 (toll-free)**
 - Buffalo City SD (math)/Buffalo State College
 - Buffalo City SD (science)/Buffalo State College
 - Morristown CSD (STEM)/Clarkson University
 - Niagara Falls City SD (STEM)/Niagara University
 - NYC CSD 9 (STEM)/NY Institute of Technology
 - NYC CSD 10 (science)/City College of NY
 - NYC CSD 30 (math)/Hunter College
 - NYC CSD 24 (STEM)/Queens College
 - NYC CSD 32 (STEM)/Queens College
 - Rochester City SD (math)/SUNY Brockport

- ⦿ **Dr. Bethany Howell, bahowell@compasseval.com, 877-652-0189 (toll-free)**
 - NYC CSD 6 (math)/Brooklyn College
 - NYC CSD 12 (science)/Queens College
 - NYC CSD 19 (math)/Queens College
 - NYC CSD 24 (science)/Brooklyn College
 - NYC CSD 30 (science)/Brooklyn College
 - Syracuse City SD (math)/Syracuse University
 - Syracuse City SD (science)/Syracuse University
 - Utica City SD (STEM)/SUNY Institute of Technology

SERVICE STATISTICS

Service statistics capture the number of MSP participants and how much and in what ways you served them. You should have already received a copy of the NYSED MSP Participant Tracking Form (if not, it is on the enclosed CD) into which you can enter these data. The NYSED Form requires that you keep track of the MSP professional development hours completed by each participant and by each activity. Figure 1 provides a sample of the NYSED Form. As shown, projects are asked to enter basic information about each participant (e.g., school, name, grade level, role in school) as well as the date and hours completed for each MSP activity in which they participated. Detailed instructions for how to complete this form are provided on a separate spreadsheet within the excel file. Projects may choose to use their own form or database provided it includes the same information.

Figure 1. Sample NYSED MSP Participant Tracking Form

| District Name/Number | School Name/Number | Last Name | First Name | Role in School | Grade Level(s) | Primary Subject(s) | MSP PD Activity 1 | |
|----------------------|--------------------|------------|------------|----------------|----------------|--------------------|-------------------------------|----------|
| | | | | | | | Name of PD: Teacher Institute | |
| | | | | | | | Type of PD: Summer Institute | |
| | | | | | | | # Hours Possible: | 40 |
| | | | | | | | Target Group: | Teachers |
| Date(s) | # Hours Completed | | | | | | | |
| CSD 21 | PS 253, Seaside | Page | Jason | Teacher | 5 | All | July 2007 | 40 |
| CSD 21 | PS 253, Seaside | Howard | Angela | Teacher | 5 | All | July 2007 | 40 |
| CSD 21 | PS 253, Seaside | Gagliano | Meggan | Teacher | 5 | All | July 2007 | 32 |
| CSD 21 | PS 253, Seaside | Smith | Laura | Teacher | 4 | All | July 2007 | 40 |
| CSD 21 | PS 253, Seaside | Rubin | Beth | Teacher | 3 | All | July 2007 | 40 |
| CSD 21 | PS 253, Seaside | James | Abby | Teacher | 4 | All | July 2007 | 24 |
| CSD 21 | PS 253, Seaside | Simpson | Brendan | Teacher | 3 | All | July 2007 | 40 |
| CSD 21 | PS 253, Seaside | Parton | Jean | Teacher | 5 | All | July 2007 | 32 |
| CSD 21 | PS 253, Seaside | Heinemeier | Sarah | Teacher | 3 | All | July 2007 | 40 |
| CSD 21 | PS 253, Seaside | Jimenez | Marisol | Teacher | 5 | All | July 2007 | 40 |
| CSD 21 | PS 253, Seaside | Brown | Valerie | Principal | NA | NA | | |
| CSD 21 | MS 135 | Howell | Bethany | Teacher | 7 | Math | July 2007 | 40 |
| CSD 21 | MS 135 | Thompson | Sheri | Teacher | 8 | Math | July 2007 | 32 |
| CSD 21 | MS 135 | Germuth | Amy | Teacher | 6 | Math | July 2007 | 40 |

The service statistics you collect will be used for two purposes: monitoring and reporting. NYSED requests that you submit your tracking form as part of their monitoring of grant implementation to assess whether your project appears to be on target for meeting the terms of the grant. The grant requires that a minimum number of participants complete 60 hours of MSP professional development each year for three years, for a total of 180 hours of professional development for each participant by the end of the grant. The minimum number to be trained is based on the level of funding you received:

- Grants funded up to \$800,000 must train 200 or more teachers,
- Grants funded up to \$500,000 must train between 100 – 200 teachers, and
- Grants funded up to 200,000 must train 100 teachers or fewer.

With regard to reporting, service statistics will be required as part of both the Federal and NYSED annual performance reports (APR). **FOR BOTH REPORTS, PLEASE USE THE FOLLOWING DEFINITIONS OF ELEMENTARY, MIDDLE, HIGH, AND OTHER:**

- Elementary: Schools with mostly elementary school grades, which are defined as grades k-5 (e.g., Prek-5, Prek-6, k-2)
- Middle Schools: Schools with mostly middle school grades, which are defined as grades 6-8 (e.g., 5-8, 6-8, 6-7, 7-8, 6-9)
- High Schools: Schools with mostly high school grades, which are defined as grades 9-12 (e.g., 9-12, 9-10, 8-12, 9)
- Other Schools: Schools that contain mixed-grades wherein the majority of grades are from two or more grade spans (e.g., k-7, k-8, 6-10, 7-10, 6-12, k-12)

The Federal APR requires that projects report on the following service statistics:

- ⦿ The total number of teachers receiving MSP professional development in math or science.
- ⦿ The number of elementary, middle, and high school participants within each of the following categories: total teachers, regular core content teachers, gifted and talented teachers, special education teachers, teachers of English language learners, non-teaching math/science coaches, paraprofessionals, administrators, and other (you may want to use these exact terms on your tracking form—"Role in School" on the NYSED Form—to make it easier to aggregate for reporting).
- ⦿ An estimate of the number of students taught by participating teachers at the elementary, middle, and high school level.
- ⦿ The number of contact hours provided by the grant for each participant. This should reflect the number of hours that an average participant received of direct professional development.
- ⦿ The types of professional development activities conducted (i.e., summer institute, on-site professional development during academic year, study groups, on-line coursework and distance learning networks, course work at universities, and other), the duration in hours per participant for all activities other than summer institutes or follow-up activities to summer institutes, and a narrative description of the professional development model. (Again, you may want to use the Federal terms for type of professional development on your tracking form—"Type of PD" on the NYSED Form—to make reporting easier.)
- ⦿ The type of content (Math: number and operations, algebra, geometry, measurement, probability and statistics, problem solving, reasoning and proof, calculus, technology, other; Science: scientific inquiry, physical science/physics, chemistry, life science/biology, earth science, technology, other) taught to participants by grade span.

Please include information for schools with mixed grade configurations (see definitions above) in "Other." Specifically, in Section G. Administrators, under Part IV. Professional Development Participants of the Federal report, please list within the "D. Other" box the number of mixed grade participants within each of the following categories: regular core content teachers, gifted and talented teachers, special education teachers, teachers of English language learners, non-teaching math/science coaches, paraprofessionals, and administrators. Report the sum of the mixed-grade participants as the number for "Other".

The NYSED APR requires that projects report on the following service statistics by grade span (i.e., elementary, middle, high, other):

- ⦿ The number of NON-PUBLIC elementary, middle, high, and mixed-grade school participants within each of the following categories: school teachers, gifted and talented teachers, special education teachers, teacher of English language learners, non-teaching math/science coaches, paraprofessionals, and administrators.
- ⦿ Total # of teachers who took part in MSP professional development activities (NOT INCLUDING COLLEGE COURSES).
- ⦿ Average # of MSP professional development hours completed by teachers (NOT INCLUDING COLLEGE COURSES).

- ⊙ Total # of teachers who took college courses through MSP project.
- ⊙ Average # of college course CONTACT hours (i.e., number hours spent in courses) completed by teachers through MSP project.
- ⊙ Average # of college course CREDIT hours (i.e., number college credits received) completed by teachers through MSP project.
- ⊙ Total # of teachers who completed 60 hours through MSP project.
- ⊙ Total # of administrators who participated in MSP professional development activities.
- ⊙ Average # of MSP professional development hours completed by administrators.
- ⊙ Total # of administrators who completed 60 hours through MSP project.
- ⊙ Total # of coaches who participated in MSP professional development activities.
- ⊙ Average # of MSP professional development hours completed by coaches.
- ⊙ Total # of coaches who completed 60 hours through MSP project.

Additionally, the NYSED report requests basic demographic information about the public and non-public schools participating in your MSP project.

- ⊙ Total # of public schools participating in MSP project.
- ⊙ Total # of public schools in need of improvement (SINI) participating in MSP project.
- ⊙ Total # of teachers at public schools eligible to participate in MSP (i.e., all math and/or all science teachers at participating schools).
- ⊙ Total # of students at participating public schools.
- ⊙ Total # of non-public schools participating in MSP project.
- ⊙ Total # of teachers at non-public schools eligible to participate in MSP (i.e., all math and/or science teachers at participating non-public schools).
- ⊙ Total # of students at participating non-public schools.

TEACHER CONTENT KNOWLEDGE

The Government Performance Results Act (GPRA) is a law that was passed in 1993 that requires federally funded agencies to develop and implement an accountability system based on performance measurement. Federal funded program offices (like the MSP federal program office) must outline the long-term and annual performance goals of their program, develop indicators to assess those performance goals, collect and analyze data on the indicators, and report progress toward achieving performance goals based on the data collected for the indicators and from grantees' evaluations. The GRPA and grantee evaluation findings are then used by Congress to inform appropriations. One of the GPRA indicators for the MSP program is: **The percentage of teachers who significantly increase their content knowledge in mathematics or science, as reflected in project-level pre- and post-assessments.**

Measuring changes in teacher content knowledge is also required by NYSED. As stated in the 2007 MSP grant application: The [evaluation] plan should include rigorous indicators and benchmarks that measure the impact of the activities carried out in each program year, including: The change in teacher mathematics or science content knowledge using evaluation methods that go beyond self report measures.

Compass will work with projects to revise their evaluation plan as specified in their application if the plan does not meet Federal and NYSED requirements for measuring change in content knowledge.

- ⊙ **Who should be assessed?** Pre- and post-tests should be administered to ALL teachers who received **direct** MSP professional development (i.e., do not include teachers who received turnkey training) in math and/or science.
- ⊙ **When should assessments be administered?** Pre-tests should be administered prior to the start of the MSP activity/college course and post-tests should be administered after the MSP activity/college course and any follow-up are complete, but within the same school year as the initial MSP activity.
- ⊙ **What assessment should I use?** Because of the variability in the activities projects implement as part of their MSP, the use of one instrument to measure changes in content knowledge across projects is not a viable option. Ideally, each project will use an established instrument(s) that has been field tested and has evidence of reliability and validity. Instruments should be aligned with the content of your MSP professional development activity. The next several pages provide information about some of the instruments that are available, have been used for similar purposes, and have been shown to generate reliable and valid results. **(If you choose not to use an established instrument, you will need to provide evidence of the reliability and validity of the results obtained for the instrument(s) you do use.)**
- ⊙ **How should I analyze these data?** The MSP federal program office has provided an electronic spreadsheet (included on the enclosed CD) onto which you can enter the pre- and post-test scores. The spreadsheet uses a dependent t-test to calculate, with 85% certainty, the number of teachers who showed significant gains on content knowledge tests. You will need to enter each teacher's pre- and post-test score into the spreadsheet. A separate spreadsheet should be used for each content knowledge instrument you use (i.e., if you use different instruments). For the NYSED report, you will need to analyze the data in a way that will allow you to report on the benchmark set in your evaluation plan.

- © **How should I report my results?** For the Federal APR, you will report the number of teachers who received MSP professional development, the number of teachers who have both pre- and post-test scores on a content knowledge assessment, and the number of teachers who showed significant gains. If you use more than one spreadsheet, because you used more than one content knowledge instrument, sum the results (i.e., the number of teachers who showed significant gains) across spreadsheets. For the NYSED APR, you will need to report on when you administered the pre- and post-tests, response rates, and results as compared to the benchmark(s) set on your evaluation plan.

Instruments for Measuring Teacher Content Knowledge in Mathematics

◎ **Content Knowledge for Teaching Mathematics Measures [CKT-M Measures]**

Learning Mathematics for Teaching (LMT) has designed surveys that measure teachers' knowledge for teaching mathematics in several content areas and teaching domains including: Number and operations (K-6, 6-8); Patterns, functions, and algebra (K-6, 6-8); and Geometry (3-8). Items in each category capture whether teachers can not only answer the mathematics problems they assign students, but also how teachers solve the special mathematical tasks that arise in teaching, including evaluating unusual solution methods, using mathematical definitions, representing mathematical content to students, and identifying adequate mathematical explanations.

The measures have each been piloted with the help of over 2000 teachers, yielding information about reliability and item characteristics. They are suitable for use in evaluations of content-focused professional development, studies examining teacher learning from pre-service coursework, new curriculum materials, or experience, and projects exploring the contribution of teacher knowledge to student achievement. Users may elect to use forms that have already been piloted and analyzed, or construct their own from our item pool.

Instruments are available for free but you must attend a free, one-day training in Ann Arbor, MI to get copies. For more information about the instruments and training see the project's website, <http://sitemaker.umich.edu/lmt/about>.

◎ **Diagnostic Mathematics Assessments for Elementary/Middle School Teachers**

Diagnostic Mathematics Assessments for Elementary/Middle School Teachers serve two purposes: (1) to describe the breadth and depth of mathematics content knowledge so that researchers and evaluators can determine teacher growth over time, the effects of particular experiences on teachers' knowledge, or relationships among teacher content knowledge, teacher practice, and student performance and (2) to describe teachers' strengths and weaknesses in mathematics knowledge so that teachers can make appropriate decisions with regard to courses or further professional development.

The assessments for elementary school teachers measure mathematics knowledge in the following four content domains: (1) Whole Number/Computation, (2) Rational Number/Computation, (3) Geometry/Masurement, and (4) Probability/Statistics/Algebra. The assessments for middle school teachers measure mathematics in the following four content domains: (1) Number/Computation, (2) Geometry/Masurement, (3) Probability/Statistics, and (4) Algebraic Ideas. The four types of teachers' knowledge assessed are memorized knowledge, conceptual understanding, problem solving/reasoning, and pedagogical content knowledge. Each assessment is composed of 20 items—10 multiple-choice and 10 open-response—and takes about 1 hour to complete. Six versions of each assessment are available in paper-pencil format so that they can be administered as pre- and post-tests.

No training is required to use the instruments. There is no cost associated with obtaining the instruments; however, for only \$7.00 per teacher CRMSTD staff will score and develop a detailed scoring summary. Instructions on how to obtain the instruments and information about instrument development, including validity and reliability, is available online at <http://louisville.edu/edu/crmstd/>.

Instruments for Measuring Teacher Content Knowledge in Science

☉ **Assessing Teacher Learning About Science Teaching (ATLAST)**

ATLAST is funded as an MSP RETA by the National Science Foundation and operated by Horizon Research, Inc. (HRI) in partnership with Project 2061 of the American Association of the Advancement of Science (AAAS). The project is founded on the premise that professional development for science teachers operates with an implicit theory that is largely untested. The theory asserts that professional development brings about changes in teachers' knowledge, which then lead to changes in classroom practice, ultimately improving student achievement.

ATLAST is developing instruments in three narrowly defined content areas. Each holds a pivotal position in physical science, earth science, and life science; and they represent the extremes in what is known about student thinking. Each is found in both *Benchmarks for Science Literacy* and the *National Science Education Standards* at the middle grades level.

- Force and motion: An unbalanced force acting on an object changes its speed or direction of motion, or both.
- Plate tectonics: The outer portion of Earth-including both the continents and the ocean basins-consists of huge plates of solid rock. The plates move very slowly (a few centimeters per year). Plate movement causes adjacent plates to interact with one another. Interactions between plates result in events and features that are observable on Earth's surface (e.g., earthquakes, volcanoes and mountain ranges); these typically occur along the boundaries between two plates.
- Flow of Matter and Energy in Living Systems: Food provides molecules that serve as fuel and building materials for all organisms. Plants use the energy in light to make sugars out of carbon dioxide and water. This food [sugars] can be used immediately for fuel or materials, or it may be stored for later use. Organisms that eat plants break down the plant structures to produce the materials and energy they need to survive. Then they are consumed by other organisms.

Teacher and student versions of the assessments are available through the website. For more information about the instruments see the project's website: <http://www.horizon-research.com/atlast/>. The contact person for these measures is the project PI, Sean Smith (ssmith@horizon-research.com).

☉ **Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART)**

The MOSART project was funded by the NSF to provide direct assessment support to NSF-funded MSPs. The instrument was developed based on the NRC National Science Education Standards and multiple forms are available for grades K-4 physical science, earth science, and astronomy; grade 5-8 physical science, earth science, and astronomy; and grades 9-12 earth science, astronomy, chemistry, and physics.

The tests have been used to measure gains in teacher content knowledge in professional development settings such as summer institutes. The structure of MOSART test questions are multiple-choice with 5 responses; contain one response that is a strong distracter, based on a specific and widespread misconception about the topic; and were created so that one answer is clearly correct, and no distracters are so trivial they can be easily dismissed. The difficulty level of the items is described as an infinite continuum going from very basic concepts (everyone should know) to very advanced concepts (most scientists should know).

The website has a user's tutorial that explains how the tests were developed and how they can be used. Once the user has completed the online tutorial, the person has free access

to all tests and supporting data as down-loadable PDF files. All tests are currently available on the website. For more information, please see the project's website:
<http://www.cfa.harvard.edu/smgphp/mosart/index.html>.

⊙ **Diagnostic Science Assessments for Middle School Teachers**

Diagnostic Science Assessments for Elementary/Middle School Teachers serve two purposes: (1) to describe the breadth and depth of science content knowledge so that researchers and evaluators can determine teacher growth over time, the effects of particular experiences on teachers' knowledge, or relationships among teacher content knowledge, teacher practice, and student performance and (2) to describe teachers' strengths and weaknesses in science knowledge so that teachers can make appropriate decisions with regard to courses or further professional development.

The assessments for elementary school teachers measure science knowledge in the following three content domains: (1) Physical Science, (2) Life Science, and (3) Earth/Space Science. The four types of teachers' knowledge assessed are memorized knowledge, conceptual understanding, problem solving/reasoning, and pedagogical content knowledge. Each assessment is composed of 25 items—20 multiple-choice and 5 open-response—and takes about 50 minutes to complete. Six versions of each assessment are available in paper-pencil format so that they can be administered as pre- and post-tests.

No training is required to use the instruments. There is no cost associated with the obtaining the instruments; however, for only \$7.00 per teacher CRMSTD staff will score and develop a detailed scoring summary. Instructions on how to obtain the instruments and information about instrument development, including validity and reliability, are available online at <http://louisville.edu/edu/crmstd/>.

Instruments for Measuring Teacher Content Knowledge in STEM

⊙ Determining Educational Technology And Instructional Literacy Skillsets (DETAILS)

DETAILS was designed to explore the current role of technology use in the classroom by measuring three key areas: (1) classroom teachers' Level of Technology Implementation (LoTi), (2) Personal Computer Use (PCU), and (3) Current Instructional Practices (CIP). The LoTi Profile portion assesses classroom teachers' current Level of Technology Implementation based on the LoTi Framework developed by Dr. Christopher Moersch; the PCU Profile portion assesses classroom teachers' comfort and skill level with using a personal computer; and the CIP Profile portion assesses classroom teachers' current instructional practices relating to a subject-matter versus a learner-based curriculum approach.

Six versions of the DETAILS for the 21st Century Questionnaire are available: Preservice Teacher, Inservice Teacher, Instructional Specialist, Media Specialist, Administrator, and Higher Education Faculty. DETAILS for the 21st Century Questionnaire helps determine your current professional development priorities related to technology and instruction and will also enable your educational institution to make better choices regarding staff development and future technology purchases. Individual information remains anonymous and the aggregate information provides various comparisons for your school, school district, regional service agency, and/or state.

NBEA and the NYSED have been working together to determine how LoTi can best be utilized throughout New York. Access to the LoTi Lounge including the online DETAILS for the 21st Century Questionnaire (previously known as the LoTi Questionnaire), individual questionnaire results profiling personalized professional development needs based on five empirically-validated skill sets, and an extensive database of resources catalogued by grade level and subject area are available at no cost to educators in the State of New York. This donation continues the process of schools in New York evaluating their instructional uses of technology for the purposes of improving student achievement tied to the initiatives set forth by the Partnership for 21st Century Skills. For more information about the development of the survey (including information on reliability and validity) as well as content, see the LoTi website: <http://www.loticonnection.com/>.

TEACHER PEDAGOGICAL SKILLS

Measuring changes in teacher pedagogical skills is required by NYSED. As stated in the 2007 MSP grant application: The [evaluation] plan should include rigorous indicators and benchmarks that measure the impact of the activities carried out in each program year, including: The change in teacher pedagogical skills in mathematics or science using evaluation methods that go beyond self report measures. Compass will work with projects to revise their evaluation plan as specified in their application if the plan does not meet NYSED requirements.

- ⊙ **What teachers should be included?** If possible, assess changes in pedagogy skills of all teachers participating in MSP. However, if observations or other time-intensive data collection methods are planned, a sample of teachers may be used. Please use random selection (e.g., stratified random sampling) to identify the teachers for your sample.
- ⊙ **What type of instruments should I use?** Preferably established instruments or instruments that have been used with a similar sample and have evidence of reliability and validity. If you choose to create your own instruments, you will need to provide evidence showing that the results are reliable and valid. Below are the names of a few instruments that have been used for assessing changes in teacher practice which have evidence of reliability and validity.
 - **Surveys of Enacted Curriculum (SEC):** A teacher self-report instrument that can be used to measure changes in the instructional practices and instructional content in mathematics or science. For more information, see http://www.ccsso.org/projects/Surveys_of_Enacted_Curriculum/Products/.
 - **A Coding Rubric for Measuring the Mathematics Quality of Instruction:** A rubric for analyzing videotaped lessons to assess changes in the pedagogical practices of teachers in mathematics. For more information, see http://sitemaker.umich.edu/lmt/faq_about_video_codes.
 - **Reformed Teaching Observation Protocol (RTOP):** An observation instrument that can be used to assess changes in the context, lesson design and implementation, lesson content and classroom culture in mathematics or science. For more information, see http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/index.htm.
 - **Classroom Observation Protocol (COP):** An observation instrument that can be used to measure the quality of an observed K-12 science or mathematics classroom lesson by examining the design, implementation, mathematics/science content, and culture of that lesson. For more information, see: <http://www.horizon-research.com/instruments/clas/cop.php>.
- ⊙ **How do I report my results?** Results will be reported as part of the NYSED APR. Information to be reported includes when instruments were administered, response rates, and results as compared to the benchmark(s) set on your evaluation plan.

STUDENT ACHIEVEMENT

The GPRA indicators for the MSP program related to student achievement are: **(1) The percentage of students in classrooms of MSP teachers who score at the basic level or below in State assessments of mathematics or science and (2) The percentage of students in classrooms of MSP teachers who score at the proficient level or above in State assessments of mathematics or science.**

Measuring student achievement is also required by NYSED. As stated in the 2007 MSP grant application: The [evaluation] plan should include rigorous indicators and benchmarks that measure the impact of the activities carried out in each program year, including: The change in mathematics and science achievement for students of teachers participating in MSP program activities using rigorous achievement measures. Evaluations will be expected to use experimental designs or quasi-experimental designs with comparison groups when measuring change in student achievement.

Compass will work with projects to revise their evaluation plan as specified in their application if the plan does not meet Federal and NYSED requirements for assessing student achievement. The following are reporting guidelines for the Federal APR:

- ⦿ **What teachers and students should be included?** Standardized testing data in mathematics/science should be obtained for ALL students in the mathematics/science classes of ALL MSP teachers in grade levels/courses that have standardized tests. Do not include students who were taught by MSP teachers in other subject areas.
- ⦿ **What does basic level or below mean?** Students in grades 3-8 scoring at levels 1 or 2 on New York State mathematics or science tests have scored at or below the basic level. Students in grades 9-12 scoring below a 65 on the NY State Regents examination have scored at or below the basic level.
- ⦿ **What does proficient level or above mean?** Students in grades 3-8 scoring at levels 3 or 4 on New York State mathematics or science tests have scored at or above the proficient level. Students in grades 9-12 scoring at or above a 65 on the NY State Regents examination have scored at or above the proficient level.
- ⦿ **How do I analyze these data?** (1) Calculate the number of students who scored at the basic level or above in the classes of each participating teacher and then sum across classes. (2) Calculate the number of students who scored at the proficient level or above in the classes of each participating teacher and then sum across classes.
- ⦿ **How do I report my results?** Results will be reported as part of the Federal APR. You will need to report the total number of students in the mathematics/science classes of participating teachers, the total number of students with student testing data in mathematics/science, the total number of students who scored at basic or below on the mathematics/science test, and the total number of students who scored at proficient or above on the mathematics/science test.

As stated above, the NYSED requires that evaluations use experimental designs or quasi-experimental designs with comparison groups when measuring change in student achievement. Additionally, there are two GPRA indicators for the MSP program related to experimental and quasi-experimental designs: **(1) The percentage of MSP projects that report using an experimental or quasi-experimental design for their evaluations and (2) The percentage of MSP projects that use an experimental or quasi-experimental design**

for their evaluations that are conducted successfully and that yield scientifically valid results. Based on these GPRA and NYSED requirements, MSP projects are to develop and implement a study to assess the impact of MSP project activities on student achievement.

- ⊙ **What should my study focus on?** Studies should focus on the effects of MSP activities on student achievement in mathematics/science. Because MSP indirectly affects student achievement, studies should consider changes in teacher pedagogical skills. Therefore, it is recommended that you include the teachers for which data on pedagogy are being collected in your sample.
- ⊙ **What type of evaluation design should I use?** The two evaluation designs that satisfy the NYSED requirements and MSP GPRA indicators are experimental and quasi-experimental. Experimental designs are considered more rigorous because they involve random assignment. In other words, people from a common pool are randomly assigned into two groups. One group would receive MSP professional development and the other group would not. This method yields two equivalent groups and, assuming both groups are treated the same with the exception of the MSP professional development, any differences observed in outcomes (measured via pre- and post-tests) would be attributable to the only difference between the two groups: MSP.

Two commonly used quasi-experimental designs are the nonequivalent groups design and the regression-discontinuity design. The nonequivalent groups design generally requires a pre- and post-test for a treatment and comparison group. It is identical to the experimental design except that groups are not created through random assignment. Instead, statistical controls or matching techniques are used to make the comparison group similar to the treatment group (e.g., number of years of teaching experience, grade levels taught, ethnicity, etc.). Use of a rigorous matching method such as propensity score matching is highly recommended.

The regression-discontinuity (RD) design in its most traditional form is a pre-test-post-test, treatment-comparison group approach. The characteristic that sets RD designs apart from other pre-post group designs is that participants are assigned to a group on the basis of a cutoff score on a pre-intervention measure. The effect of the treatment is measured by differences (“discontinuity”) in the regression line on either side of the cutoff point.

The evaluation design you choose will depend to some extent on the design of your program as well as the feasibility of conducting one design over the other. It is usually quite difficult to successfully implement an experimental design in an educational setting. If it appears that using any type of control or comparison group will not be possible, please contact your Compass evaluator to discuss the barriers, brainstorm possible solutions, and, if all else fails, to come up with an alternative design.

- ⊙ **How big should the sample be?** Ideally, you will want to perform a power analysis to determine how big your sample should be. However, if the sample size yielded through the power analysis exceeds what is possible given your resources, we suggest including no fewer than 35 treatment and 35 comparison/control teachers (be sure to consider minimum sample size requirements for any statistical tests to be used).
- ⊙ **How do I report my results?** Results will be reported as part of the Federal APR and NYSED report. The Federal APR requests information about the design, how control or comparison groups were created, sample size, data collection measures and process, and type of analysis. The NYSED report will request similar information but will ask for more detail with regard to data collection and results.

