DEFINITIONS

**Essential Questions:** Key questions that raise important conceptual issues, provoke and sustain engaged inquiry, stimulate debate, raise further questions, and require problem solving.

**Formative Assessment:** “Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes. Formative assessment is not an adjunct to teaching but, rather, integrated into instruction and learning with teachers and students receiving frequent feedback.” (The Council of Chief State School Officers (CCSSO) http://www.ccsso.org/publications/details.cfm?PublicationID=362)

**High-Need District:** For the purpose of the MMSP, Level 3 & 4 districts are considered high need districts.

- **Level 3** districts with one or more schools among the lowest-performing 20% based on quantitative indicators.
- **Level 4** districts identified by quantitative and qualitative indicators through a district review; districts with one or more schools among the lowest-performing and least improving 2% based on quantitative indicators.
- **Level 5** is a “chronically underperforming” school and is both low performing and not showing signs of substantial improvement over time. It is the most serious category in Massachusetts’ accountability system, representing receivership.

**Instructional Leadership:** The efforts of educational leaders (Superintendents, Principals, Central Office Staff, Curriculum Coordinators, Instructional Coaches, Mentor Teachers) to support the implementation of standards-based curriculum and instruction and ensure the school’s/district’s instructional systems are properly aligned.

**Inquiry-Based Learning Experiences:** Problem-based activities that actively engage learners in understanding science content and its application to the natural and human-made world. Inquiry-based learning experiences engage the learner in using analytical methods (making observations, posing questions, formulating hypotheses, designing and conducting scientific investigations/experiments, analyzing and interpreting data or other evidence, and communicating and applying the results of scientific investigations/experiments) to construct an accurate understanding of scientific concepts, ideas, and principles.

**Professional Development:** Instructional activities that:

1. improve and increase teachers’ knowledge of the academic subjects they teach;
2. are sustained, intensive, and classroom-focused in order to have a positive and lasting impact on classroom instruction and the teacher’s performance in the classroom;
3. enable teachers to become licensed in the subject area(s) and grade level(s) they teach; and
4. are based on scientifically based research and state academic content standards, student academic achievement standards, and assessments.

**Scientifically Based Research** and **Rigorous Summative Evaluation:** Research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs. Includes research that:

1. employs systematic, empirical methods that draw on observation or experiment and involve rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;
2. relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators;
3. is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs, or activities are assigned to different conditions, with appropriate controls to evaluate the effects of the condition of interest and with a preference for random-assignment experiments or other designs to the extent that those designs contain within-condition or across-condition controls;
4. ensures that experimental studies are presented in sufficient detail and clarity to allow for replication or, at minimum, to offer the opportunity to build systematically on their findings; and
5. has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.

*Please see Table 1: Criteria for Quasi-Experimental Design

**Supplemental Activities:** Activities following the 45 hour course that facilitate teachers’ application of their learning in the professional development course to their classroom instruction.
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Key Recommendations for meeting Quasi-Experimental Design criteria</th>
<th>Meeting the Criteria for Assessing whether Quasi-Experimental Designs are Conducted Successfully and Yield Scientifically Valid Results</th>
</tr>
</thead>
</table>
| A. Baseline Equivalence of Groups | 1. Complete and report pre-testing for differences between groups on key outcomes.  
2. Provide full information (e.g., sample size, mean, standard deviation, test results such as t-statistics) about the pre-test. | There are no significant pre-intervention differences between treatment and comparison group participants on variables related to the study’s key outcomes; or adequate steps are taken to address the lack of baseline equivalence in the statistical analysis. |
| B. Sample size¹              | 1. Clearly report sample sizes for all groups and subgroups for key outcomes.                                                      | Sample size is adequate (i.e., based on power analysis with recommended significance level=0.05, power=0.8, minimum detectable effect size informed by the literature or otherwise justified).  
Alternatively, meet or exceed “rule of thumb” threshold sample sizes: Teacher Outcomes: 12 schools or 60 teachers  
Student Outcomes: 12 schools or 18 teachers or 130 students                                                                 |
| C. Quality of the Measurement Instruments | 1. Note the validity and reliability of all instruments used.  
2. Use full sub-scales when taking items from existing instruments, where possible.  
3. Test for validity and reliability when creating a new instrument. | The study uses existing data collection instruments that had already been deemed valid and reliable to measure key outcomes; or data collection instruments are developed specifically for the study and are sufficiently pre-tested with subjects who are comparable to the study sample. |
| D. Quality of the Data Collection Methods | 1. Collect data from the comparison and treatment groups at the same time in a systematic fashion. | The methods, procedures, and timeframes used to collect the key outcome data from treatment and comparison groups are the same. |
| E. Data Reduction Rates (i.e., Attrition Rates, Response Rates) | 1. Report initial and final sample sizes for all groups and subgroups.  
2. Note the number of students in the classrooms and the number of students who transfer in and out over the course of the evaluation. | (1) The study measures the key outcome variable(s) in the post-tests for at least 70% of the original study sample (treatment and comparison groups combined) or there is evidence that the high rates of data reduction are unrelated to the intervention, AND (2) the proportion of the original study sample that is retained in follow-up data collection activities (e.g., post-intervention surveys) and/or for whom post-intervention data are provided (e.g., test scores) is similar for both the treatment and comparison groups (i.e. less or equal to a 15-percent difference), or the proportion of the original study sample that is retained in the follow-up data collection is different for the treatment and comparison groups, and sufficient steps are taken to address this differential attrition are not taken in the statistical analysis. |
| F. Relevant Statistics Reported | 1. Describe the sample sizes, the means, and the standard deviations as well as the statistical tests, used to analyze results. | The final report includes treatment and comparison group post-test means, and tests of statistical significance for key outcomes; or provides sufficient information for calculation of statistical significance (e.g., mean, sample size, standard deviation/standard error). |


¹ The critical sample size here is related to the unit of grouping. For example, if the grouping is made at the school level, the relevant sample size is the number of schools involved.