Criteria for Classifying Designs of MSP Evaluations[1]

❒ **Experimental study—**the study measures the intervention's effect by randomly assigning individuals (or other units, such as classrooms or schools) to a group that participated in the intervention, or to a control group that did not; and then compares post-intervention outcomes for the two groups

❒ **Quasi-experimental study—**the study measures the intervention's effect by comparing post-intervention outcomes for treatment participants with outcomes for a comparison group (that was not exposed to the intervention), chosen through methods other than random assignment. For example:

- *Comparison-group study with equating*—a study in which statistical controls and/or matching techniques are used to make the treatment and comparison groups similar in their pre-intervention characteristics

- *Regression-discontinuity study*—a study in which individuals (or other units, such as classrooms or schools) are assigned to treatment or comparison groups on the basis of a "cutoff" score on a pre-intervention non-dichotomous measure

❒ **Other**

- The study uses a design other than a randomized controlled trial, comparison-group study with equating, or regression-discontinuity study, including *pre-post* studies, which measure the intervention's effect based on the pre-test to post-test differences of a single group, and comparison-group studies without equating, or non-experimental studies that compare outcomes of groups that vary with respect to implementation fidelity or program dosage.

[1] To be used for addressing following MSP GPRA measure: *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.*

Criteria for Assessing whether *Experimental Designs* Were Conducted Successfully and Yielded Scientifically Valid Results

A. Sample size[2]

- ❒ **Met the criterion**—sample size was adequate (i.e. based on power analysis with recommended significance level=0.05, power=0.8, and a minimum detectable effect informed by the literature or otherwise justified).

- ❒ **Did not meet the criterion** —the sample size was too small

- ❒ **Did not address the criterion**

B. Quality of the Measurement Instruments

- ❒ **Met the criterion**—the study used existing 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

- ❒ **Did not meet the criterion** —the key data collection instruments used in the evaluation lacked evidence of validity and reliability

- ❒ **Did not address the criterion**

C. Quality of the Data Collection Methods

- ❒ **Met the criterion**—the methods, procedures, and timeframes used to collect the key outcome data from treatment and control groups were the same

- ❒ **Did not meet the criterion**—instruments/assessments were administered differently in manner and/or at different times to treatment and control group participants

D. Data Reduction Rates (i.e. Attrition Rates, Response Rates)

- ❒ **Met the criterion**—(1) the study measured the key outcome variable(s) in the post-tests for at least 70% of the original study sample (treatment and control groups combined) or there is evidence that the high rates of data reduction were unrelated to the intervention, AND (2) the proportion of the original study sample that was retained in follow-up data collection activities (e.g., post-intervention surveys) and/or for whom post-intervention data were provided (e.g.,

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

test scores) was similar for both the treatment and control groups (i.e. less or equal to a 15-percent difference), or the proportion of the original study sample that was retained in the follow-up data collection was different for the treatment and control groups, but sufficient steps were taken to address this differential attrition in the statistical analysis

❒ **Did not meet the criterion**—(1) the study failed to measure the key outcome variable(s) in the post-tests for 30% or more of the original study sample (treatment and control groups combined), and there is no evidence that the high rates of data reduction were unrelated to the intervention; OR (2) the proportion of study participants who participated in follow-up data collection activities (e.g., post-intervention surveys) and/or for whom post-intervention data were provided (e.g., test scores) was significantly different for the treatment and control groups (i.e. more than a 15-percent difference) and sufficient steps to address differential attrition were not taken in the statistical analysis

❒ **Did not address the criterion**

E. Relevant Statistics Reported

❒ **Met the criterion**—the final report includes treatment and control 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)

❒ **Did not meet the criterion**—the final report does not include treatment and control group post-test means, and/or tests of statistical significance for key outcomes; or provide sufficient information for calculation of statistical significance (e.g., mean, sample size, standard deviation/standard error)

❒ **Did not address the criterion**

Criteria for Assessing whether *Quasi-Experimental Designs* Were Conducted Successfully and Yielded Scientifically Valid Results

A. Baseline Equivalence of Groups

- ❒ **Met the criterion**—there were no significant pre-intervention differences between treatment and comparison group participants on variables related to the study's key outcomes; or adequate steps were taken to address the lack of baseline equivalence in the statistical analysis

- ❒ **Did not meet the criterion**—there were statistically significant pre-intervention differences between treatment and comparison group participants on variables related to the study's key outcomes; and no steps were taken to address lack of baseline equivalence in the statistical analysis

- ❒ **Did not address the criterion**

B. Sample size[3]

- ❒ **Met the criterion**—sample size was 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)

- ❒ **Did not meet the criterion** —the sample size was too small

- ❒ **Did not address the criterion**

C. Quality of the Measurement Instruments

- ❒ **Met the criterion**—the study used existing 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

- ❒ **Did not meet the criterion** —the key data collection instruments used in the evaluation lacked evidence of validity and reliability

- ❒ **Did not address the criterion**

D. Quality of the Data Collection Methods

[3] 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.

❒ **Met the criterion**—the methods, procedures, and timeframes used to collect the key outcome data from treatment and comparison groups were the same

❒ **Did not meet the criterion**—instruments/assessments were administered differently in manner and/or at different times to treatment and comparison group participants

E. Data Reduction Rates (i.e. Attrition Rates, Response Rates)

❒ **Met the criterion**—(1) the study measured 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 were unrelated to the intervention, AND (2) the proportion of the original study sample that was retained in follow-up data collection activities (e.g., post-intervention surveys) and/or for whom post-intervention data were provided (e.g., test scores) was 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 was retained in the follow-up data collection was different for the treatment and comparison groups, and sufficient steps were taken to address this differential attrition were not taken in the statistical analysis

❒ **Did not meet the criterion**—(1) the study failed to measure the key outcome variable(s) in the post-tests for 30% or more of the original study sample (treatment and comparison groups combined), and there is no evidence that the high rates of data reduction were unrelated to the intervention; OR (2) the proportion of study participants who participated in follow-up data collection activities (e.g., post-intervention surveys) and/or for whom post-intervention data were provided (e.g., test scores) was significantly different for the treatment and comparison groups (i.e. more than a 15-percent) and sufficient steps were not taken to address differential attrition in the statistical analysis

❒ **Did not address the criterion**

F. Relevant Statistics Reported

❒ **Met the criterion**—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)

❒ **Did not meet the criterion**—the final report did not include treatment and comparison group post-test means, or tests of statistical significance for key outcomes; or provide sufficient information for calculation of statistical significance (e.g., mean, sample size, standard deviation/standard error)

❒ **Did not address the criterion**