

Understanding Evaluation-Related Performance Measures

Rachel McCormick | Anne Wolf (Presenters)

Abt Associates

Rachel McCormick



Role: Evaluation Technical Assistance Project Director



Background: Rachel is a Senior Associate in Abt Associates' Social and Economic Policy division, with 15 years of experience running large-scale education research projects. She is currently the Project Director for the evaluation technical assistance efforts in ED's i3 and EIR grant programs, and was previously Project Director for the First in the World grant program's evaluation technical assistance efforts.





Agenda

- Understanding the EIR performance measures and why they matter
- Evaluation-related performance measures, criteria, and assessment results
 - Performance measure 1: Evidence of effectiveness
 - Performance measure 2: Key practices and approach of the project
 - Performance measure 3: Performance feedback
 - Performance measure 4: Cost-effectiveness
- Wrap-up & Knowledge Check

Understanding the EIR Performance Measures and Why They Matter

Important Terminology

- EIR Performance Measures are a set of metrics assessed annually and at the end of the grant period (cumulative) that are used to report on the extent to which grantees individually and overall are meeting program goals.
- An Annual Performance Report (APR) provides an EIR grant's most current performance and financial expenditure information. Grantees submit APRs to ED's Office of Elementary and Secondary Education (OESE).
- An Evaluation Design Plan describes how an independent evaluator will measure the effect of a funded EIR intervention/program. Evaluators submit Evaluation Design plans to the Evaluation Technical Assistance (TA) Team.

Evaluation is Critical to EIR's Program Goals



Develop and expand innovative practices that improve student achievement or seem



Build the research base on what works in different contexts: EIR's contribution to the field depends on the quality of evidence the EIR evaluations produce



Disseminate findings to allow others to benefit from ED's grant investments

Annual EIR Performance Measures Align with Program Goals

- Targeted number of students served annually
- Targeted number of high-need students served annually
- Evaluations of effectiveness that are well-designed & independent
- Evaluations that provide information about key elements or practices & approach
- Evaluations designed to provide performance feedback
- Evaluations that provides information on cost-effectiveness
- Cost per student served by the grant

Performance Measures on Evaluation Quality Establish Clear Expectations

- Targeted number of students served annually
- Targeted number of high-need students served annually
- 1. Evaluations of effectiveness that are well-designed & independent
- 2. Evaluations that provide information about key elements or practices & approach
- 3. Evaluations designed to provide performance feedback
- 4. Evaluation that provides information on cost-effectiveness
- Cost per student served by the grant

Three Groups Contribute to the EIR Performance Measures



EIR Project Teams

Provide key information on the performance of their projects and evaluations:

- Report on project-specific measures, students served & costs in APR
- Submit Evaluation Design Plan to the Evaluation TA Team
- Report evaluation findings to OESE and publicly in ERIC



Evaluation TA Team

Provide information on the rigor and quality of evaluations to grantees and OESE:

- Review Evaluation Design Plans from evaluators
- Provide feedback memos on evaluation designs
- Risk reporting on threats to rigor/quality of evaluations



OESE

Assess information on project and evaluation-related performance measures using:

- Grant-specific APRs
- Evaluation feedback memos
- Risk reporting from Eval TA team
- WWC review of final reports

Multiple Data Sources Inform EIR Performance Measures

Data Source

 Annual Performance Report (APR)

2. Feedback memos on Evaluation Design Plan

3. Evaluation risk reports

4. Public evaluation findings reports & their WWC reviews

Type of EIR Performance Measure Informed

Project-Specific Performance Measures

Annual Measures

Cumulative Measure

Performance measures on students served and cost

Annual Measures

Cumulative Measure

Performance measures on evaluation

Annual Measures

Cumulative Measure

Today's Session Focuses on Annual Evaluation-Related Measures

Data Source

Type of EIR Performance Measure Informed

2. Feedback memos on Evaluation Design Plan

3. Evaluation risk reports

Performance measures on evaluation

Annual Measures

Knowing Where to find Evaluation TA Team's Review Results

Feedback memo on **Draft**Evaluation Design Plan

- Initial primary source
- Shared with grantee, evaluator, and OESE program officer

Feedback memo on <u>Final</u> Evaluation Design Plan

- Final primary source
- Shared with grantee, evaluator, and OESE program officer

Evaluation risk reports

- Aligned with feedback memos at the time memos are sent
- Reports on changes after the draft/final evaluation design plan memo sent

Your Participation Supports Evidence Based Decision Making

- Submitting your APRs and Evaluation Design Plan and posting your findings publically is critical to giving policymakers consistent, high-quality data on the EIR program
- Thank you for your commitment to supporting evidence-based decision making



Evaluation-Related Performance Measures, Criteria, & Assessment Results

Anne Wolf



Role: TA Leader & TA Liaison



Background: Anne is a Senior Scientist in Abt Associates' Social and Economic Policy division. She has more than 15 years of experience in designing and conducting experimental and quasi-experimental evaluations of education, nutrition, and human service programs. She provides evaluation technical assistance for the EIR and i3 grant program and led technical assistance for the First in the World grant program.



Evaluation-Related Performance Measures

Performance Measure	Number of Criteria	Grant Types Affected
1. Well-designed, independent evaluation that provides evidence of effectiveness at improving student outcomes	5	All grants
2. Evaluation that provides information about key	All grants for 4 criteria	
practices and the approach of the project to facilitate testing, development, or replication in other settings	5	Mid-phase & Expansion grants for 1 criterion
3. Evaluation that provides performance feedback to inform project design	1	Early-phase grants
4. Evaluation that provides information on cost- effectiveness of key practices to identify potential obstacles to and success factors in scaling	2	Mid-phase & Expansion grants

Feedback Memo on Evaluation Design Plan

```
Based on our review of your Evaluation Design Plan:
```

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Potential WWC Rating (Meets WWC Standards with or without Reservations)

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

ESSA level of evidence (Strong evidence or Moderate evidence)

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets / Does Not Meet EIR expectations for independence

Meets / Does Not Meet EIR expectations for alignment of evaluation sample to intervention population

Meets / Does Not Meet EIR expectations for registration of planned impact analyses

Meets / Does Not Meet EIR expectations for the logic model

Meets / Does Not Meet EIR expectations for the fidelity measure

Meets / Does Not Meet EIR expectations for alignment of fidelity sample to intervention sites

Meets / Does Not Meet EIR expectations for the scale-up logic model

Meets / Does Not Meet EIR expectations for performance feedback

Meets / Does Not Meet EIR expectations for measuring cost-effectiveness

Performance Measures, Criteria, & Assessment Results

Evaluation Performance Measure 1:

Well-designed, independent evaluation that provides evidence of effectiveness at improving student outcomes

All Grant Types

Evaluation Performance Measure 1: Well-designed, independent evaluation that provides evidence of effectiveness

Defined by 5 criteria:

- 1a. Evaluation is independent
- 1b. Evaluation design meets WWC standards with or without reservations
- 1c. Evidence is based on sample that is large and multi-site
- 1d. Evidence is based on sample that overlaps with populations & settings that receive the intervention
- 1e. Evaluation has scientific integrity

1a. Evaluation Is Independent: Criteria

- Main activities of evaluation are designed and carried out independent of, but in coordination with, entities responsible for developing and implementing the intervention
- Includes independent conduct of random assignment, data collection of outcomes, analysis, and reporting of findings

1a. Evaluation Is Independent: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual.

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

1b. Evaluation Design Meets WWC Standards: Criteria

Criteria varies by your EIR grant type:

- Early-phase grants
 - At least one analysis that meets WWC standards with or without reservations
- Mid-phase & Expansion grants
 - At least one analysis that meets WWC standards without reservations

1b. Evaluation Design Meets WWC Standards: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

1c. Sample Is Large & Multi-Site: Criteria

- Large
 - Analytic sample includes at least 350 individuals, regardless of design or clustering
- Multi-site
 - Study has multiple schools (at least 2) in the intervention group and in the comparison group

1c. Sample is Large & Multisite: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual.

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for

Strong evidence

Our assessment of other EIR expectations for the evaluation

Meets EIR expectations for independence

Meets EIR expectations for alignment of evalua

Meets EIR expectations for registration of plann

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sa •

Meets EIR expectations for the scale-up logic mod •

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

ESSA level of evidence is based on the combination of the WWC rating and the scale of the sample.

Strong evidence =

- Meets WWC standards without reservations
- AND sample is LARGE & MULTI-SITE

Moderate evidence =

- Meets WWC standards with reservations
- AND sample is **LARGE & MULTI-SITE**

1d. Sample Overlaps with Intervention Populations & Settings: Criteria

The impact evaluation sample includes either:

- All or a random sample of the schools, teachers, and students offered the EIR intervention
- A nonrandom sample of these schools, teachers, or students that excludes ≤ 25% of schools and ≤10% of individuals

1d. Sample Overlaps with Intervention Populations & Settings: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

1e. Evaluation Has Scientific Integrity: Criteria

The evaluation should:

- Pre-specify the confirmatory analyses prior to collecting outcome data, and then reports findings from these planned analyses
- Be pre-registered in the Registry of Efficacy and Effectiveness Studies (REES)

1e. Evaluation Has Scientific Integrity: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

Performance Measures, Criteria, & Assessment Results

Evaluation Performance Measure 2:

Provides information about *key practices and the approach* of the project, to facilitate testing, development, or replication in other settings

All Grant Types

Evaluation Performance Measure 2: Provides information about key practices to facilitate testing, development, or replication

Four criteria apply to all grant types:

- 2a. Intervention logic model specifies key components, mediators, and student outcome domains
- 2b. Fidelity measure specifies thresholds to determine if for the entire sample each key component of the intervention was implemented with fidelity
- 2c. Fidelity is measured periodically for the entire implementation sample, and fidelity findings are reported
- 2d. Sample on which fidelity is measured represents the entire implementation sample or a random subsample for each year of measurement

Evaluation Performance Measure 2: Provides information about key practices to facilitate testing, development, or replication

One criterion applies to only EIR Mid-phase & Expansion grants:

2e. Scale-up implementation process shows goals and mechanism(s) through which the intervention is expected to reach scale-up goals

Mid-phase & Expansion Grants Only

2a. Intervention Logic Model: Criteria

- There is a logic model for the intervention
- Logic model clearly articulates three elements:
 - Key components of the intervention
 - Mediator(s) through which the intervention is expected to have its intended outcomes
 - Student outcome domain(s)

2a. Intervention Logic Model: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

2b. Implementation Fidelity Measure: Criteria

- There is a measure of implementation fidelity for <u>each key component</u> of the intervention
- The measure <u>specifies a threshold</u> for determining whether each key component of the intervention was implemented with fidelity at the sample level

2c. Periodic Measurement of Implementation Fidelity: Criteria

- Implementation is <u>measured periodically</u> (at least twice) during the grant
- Fidelity is <u>assessed and reported</u> for each key component at the sample level

2b & 2c. Implementation Fidelity Measure & Periodic Measurement of Implementation Fidelity: Feedback Memo

Based on our review of your evaluation plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

2d. Fidelity Measured on Entire Implementation Sample: Criteria

In each year that fidelity is measured, the fidelity evaluation sample includes either:

• All units that received the i3 or EIR intervention

<u>or</u>

A random sample of the units that received the i3 or EIR intervention

2d. Fidelity Measured on Entire Implementation Sample: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

2e. Scale-up Implementation Process: Criteria

Evaluation identifies:

- At least one challenge to scale-up that is being addressed in current evaluation
- At least one scale-up mechanism to address each challenge
- Goal for each scale-up mechanism
- Plan for documenting and reporting extent to which each scale-up goal has been met

2e. Scale-up Implementation Process: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Mid-phase & Expansion

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

Performance Measures, Criteria, & Assessment Results

Evaluation Performance Measure 3:

Evaluation that provides *performance feedback* to inform project design

Early-phase Grants Only

Evaluation Performance Measure 3: Provides performance feedback to inform project design

Defined by 1 criterion:

 Evaluation does or does not include a pilot study to provide performance feedback to inform project design

Evaluation design plan should describe:

 Whether or not the evaluation will include a pilot study prior to the impact evaluation

Evaluation Performance Measure 3: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

Early-phase only

Performance Measures, Criteria, & Assessment Results

Evaluation Performance Measure 4:

Evaluation provides information on *cost-effectiveness* of key practices to identify potential obstacles or success factors in scaling

Mid-phase & Expansion Grants Only

Evaluation Performance Measure 4: Provides information on costeffectiveness and identifies potential obstacles and success factors in scaling

Defined by 2 criteria:

- 4a. Evaluation is expected to measure and report on the **cost-effectiveness** of the intervention
- 4b. Evaluation is expected to identify whether cost appears to be an obstacle to scaling, and to identify any other actual or potential barriers to and success factor(s) in scaling that appear to be present

4a. Cost-Effectiveness of Intervention: Criteria

- Annual per student cost is reported as part of APR
- An impact on at least one student outcome is estimated in effect size units

4a. Cost-Effectiveness of Intervention: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

Mid-phase & Expansion

4b. Barriers & Success Factors in Scaling: Criteria

 The evaluation satisfies the expectations for documenting the scale-up implementation process (criterion 2e)

4b. Barriers & Success Factors in Scaling: Feedback Memo

Based on our review of your Evaluation Design Plan:

Your impact evaluation design is: DESIGN (e.g., RCT or QED)

Your impact evaluation will compare INTERVENTION NAME versus business-as-usual

Our assessment of the highest potential WWC rating for your evaluation, as currently planned, is:

Meets WWC Standards without Reservations

Our assessment of the highest potential level of evidence for your evaluation, as currently planned, is:

Strong evidence

Our assessment of other EIR expectations for the evaluation, as currently planned:

Meets EIR expectations for independence

Meets EIR expectations for alignment of evaluation sample to intervention population

Meets EIR expectations for registration of planned impact analyses

Meets EIR expectations for the logic model

Meets EIR expectations for the fidelity measure

Meets EIR expectations for alignment of fidelity sample to intervention sites

Meets EIR expectations for the scale-up logic model

Meets EIR expectations for performance feedback

Meets EIR expectations for measuring cost-effectiveness

Same as criterion 2e:

Mid-phase & Expansion

Check – What Should I Understand Now?

- ✓ What are performance measures, and why does it matter for my grant?
- ✓ What are the Evaluation-related performance measures? When and how are they reviewed?
- ✓ Where do I find information on the results of my short-term evaluation-related performance review?

What are My Key Steps Regarding Evaluation-Related Performance Measures?

Evaluators

- Submit Draft and Final Evaluation Design Plans to the Evaluation TA Team (as soon as feasible)
- Write and publically post a Findings Report
- Evaluators & Grantees
 - Review feedback memos from Evaluation TA Team
 - Note implications of feedback memos for evaluation-related performance measures
 - Notify evaluation TA Team about changes to evaluation designs
 - Note any risks your TA Liaisons indicates will be shared with OESE

Questions



Contact

EIR Evaluation TA Team

EIREvalTA@AbtAssoc.com



www.abtassociates.com