Reporting Your Findings with an Eye to a WWC Review

Anne Wolf (Presenter)
Barbara Goodson (Principal Investigator)

Abt Associates
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 i3 grant program and led technical assistance for the First in the World grant program.
Agenda

- Learn *what* information to include in an evaluation report on program effects, including:
  - Intervention and comparison conditions
  - Study design and measures
  - Analytic approach
  - Findings
- Access *additional resources* to support you in findings
Reporting Checklist
## Reporting Checklist

<table>
<thead>
<tr>
<th>Topic</th>
<th>What to Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>✓ Key intervention activities/inputs</td>
</tr>
<tr>
<td></td>
<td>✓ Participant and setting characteristics</td>
</tr>
<tr>
<td>Comparison</td>
<td>✓ What comparison group members received</td>
</tr>
<tr>
<td></td>
<td>✓ What they could not receive</td>
</tr>
<tr>
<td>Study Design</td>
<td>✓ Method of assignment</td>
</tr>
<tr>
<td></td>
<td>✓ Unit of assignment</td>
</tr>
<tr>
<td></td>
<td>✓ Use of blocking</td>
</tr>
<tr>
<td>Measures</td>
<td>✓ Reliability &amp; face validity</td>
</tr>
<tr>
<td></td>
<td>✓ Not over-aligned with treatment</td>
</tr>
<tr>
<td></td>
<td>✓ Consistently collected in both conditions</td>
</tr>
<tr>
<td>Analysis Approach</td>
<td>✓ Statistical model (clustering, blocking, baseline covariates)</td>
</tr>
<tr>
<td></td>
<td>✓ Approach to handling missing data</td>
</tr>
<tr>
<td></td>
<td>✓ Approach to establishing baseline equivalence, if necessary</td>
</tr>
<tr>
<td>Findings</td>
<td>✓ Effects</td>
</tr>
<tr>
<td></td>
<td>✓ Attrition</td>
</tr>
<tr>
<td></td>
<td>✓ Baseline Equivalence</td>
</tr>
<tr>
<td></td>
<td>✓ Representativeness</td>
</tr>
</tbody>
</table>
Reporting Checklist: Intervention

- Key intervention activities/inputs
- Participant and setting characteristics
What to Report about the Intervention

Key Intervention Activities
- Content
- Method of delivery
- Intensity and duration
- Resources
- Training and coaching

Participants and Settings
- Eligibility criteria
- Setting location and characteristics
- When intervention took place
Reporting Checklist: Comparison Condition

- What comparison group members received
- What they could not receive
What to Report about the Comparison Condition

**Services Offered**
- Describe the business-as-usual services or the alternative intervention

**Services Withheld**
- Identify elements of the intervention that were not available to the comparison group
Reporting Checklist: Study Design

- Method of assignment
- Unit of assignment
- Use of blocking
Describe Method of Assignment

Random Assignment

- How assignment occurred randomly
- Probability of assignment
- Any post-random assignment exclusions
- For cluster RCTs, when students identified

Non-Random Assignment

- How units gained access
- How and when units selected
- Matching methods, if used
- Baseline matching characteristics
Study Design – Unit of Assignment: What to Report

Individual Assignment

Cluster Assignment
Study Design – Use of Blocking: What to Report

- **Blocking**
  - Grouping or pairing units based on shared characteristics
  - Units are randomly assigned (RCTs) or matched (QEDs) with blocks
  - Ensures balance across conditions

- **Describe blocks**
  - For example: “For each cohort, students were assigned within schools…”
  - For RCTs
    - Identify probability of assignment within blocks
    - Probabilities may differ across blocks
Reporting Checklist: Measures

- Reliability & face validity
- Not over-aligned with treatment
- Consistently collected in both conditions
What to Report about Measures

- Identify the measure
  - Name of the measure
  - Standardized or not standardized?
  - Modified or developed for the evaluation?

- Reliability
  - Provide an appropriate reliability statistics (not required for standardized measures)

- Face validity
  - Description should show that it measures the intended construct

- Not over-aligned
  - Make it clear that the treatment group does not have an unfair advantage based on practice or prior exposure

- Consistently collected
  - Method and timing of data collection
  - Data collection personnel
Reporting Checklist: Analysis Approach

- Statistical model (clustering, blocking, baseline covariates)
- Approach to handling missing data
- Approach to establishing baseline equivalence, if necessary
What to Report about the Analysis Model

- **Statistical Model**
  - Describe the model
  - *For example, linear regression, HLM, ANCOVA*

- **Clustering Adjustment**
  - Describe adjustment method used
  - *For example, HLM, Huber-White sandwich estimator, Generalized Estimating Equations*

- **Blocking Adjustment**
  - Describe how analysis accounted for blocking
  - *For example, block dummy covariates; estimate effects separately by block*

- **Covariate Adjustment**
  - Describe any pre-intervention measures included as covariates in the analysis models
  - *For example, measures required to establish baseline equivalence*
The hierarchical linear model took the following form:

**Level-1 (student-level):**  
\[ Y_{ij} = \beta_0 + \beta_1'X_{ij} + \varepsilon_{ij} \]

**Level-2 (course-section level):**  
\[ \beta_0 = \gamma_00 + \gamma_{01}T \text{rt}_{ij} + \sum_{m=2}^M \gamma_{0m}C_{\text{O}h\text{o}r\text{t}_j} + \mu_j \]

Where:
- \( Y_{ij} \) = outcome score for student \( i \) in course section \( j \)
- \( \beta_0 \) = outcome score in control course section \( j \)
- \( \beta_1 \) = vector of coefficients for individual characteristics
- \( X_{ij} \) = set of variables for individual characteristics and the measures used to establish baseline equivalence (Accuplacer Elementary Algebra score and a first-generation indicator)
- \( \varepsilon_{ij} \) = error term for student \( i \) in course section \( j \)
- \( \gamma_{00} \) = average outcome score in control courses
- \( \gamma_{01} \) = outcome score difference between treatment course sections and control course sections
- \( T \text{rt}_{ij} \) = treatment status for course section \( j \)
- \( \gamma_{0m} \) = vector of parameters for the effects of cohort
- \( C_{\text{O}h\text{r}\text{t}_j} \) = set of dummy variables indicating cohort A, B, or C for course section \( j \)
- \( \mu_j \) = error term for course section \( j \)
Describe Approach to Handling Missing Data

No Missing Data Imputed
- Complete case analysis

Missing Data Imputed
- Regression imputation
- Maximum likelihood
- Non-response weights
- Dummy variable method (for RCTs only)
What to Report about the Baseline Equivalence Model

- **Approach Used**
  - Simple difference in unadjusted means, OR
  - Model-based approach that reflects design (blocking, unit of assignment, propensity score weights)

- **Statistical Model**
  - Report similar information as for the analysis model
  - Describe the model (such as linear regression, HLM…)
  - RECOMMENDED: Provide the Greek model

- **Describe Any Adjustments**
  - Unequal allocation in random assignment
  - Blocking, weighting, clustering
Reporting Checklist: Findings

- Effects
- Attrition
- Baseline Equivalence
- Representativeness
What to Report About the Effects (for each outcome, by condition)

- **Sample Sizes**
  - Number of individuals in T and C groups for each analytic sample
  - For cluster design, number of clusters in T and C groups

- **Posttest Means**
  - For comparison group: Unadjusted posttest mean
  - For treatment group: Model-adjusted posttest mean (sum of comparison mean and treatment effect)

- **Posttest Standard Deviations**
  - Continuously-scaled outcome measure: Standard deviation (SD) for individuals in the T and C groups
  - Binary outcome measure: Not applicable

- **Treatment Effect**
  - Unstandardized estimate from the analysis model (i.e., coefficient for treatment)
  - Standardized effect size converted using Hedges’ g or Cox’s Index

- **p-value**
  - Exact p-value for the treatment effect, from a two-tailed test
## Findings – Effects: Example Estimates of Program Effects

<table>
<thead>
<tr>
<th>Measure</th>
<th>Treatment Group</th>
<th>Comparison Group</th>
<th>Estimated Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N schools</td>
<td>N students</td>
<td>Posttest Mean</td>
</tr>
<tr>
<td>Math: State Test</td>
<td>30</td>
<td>285</td>
<td>84.4</td>
</tr>
<tr>
<td>ELA: State Test</td>
<td>30</td>
<td>278</td>
<td>95.1</td>
</tr>
</tbody>
</table>

<sup>a</sup>The effect size is the T-C difference divided by the pooled standard deviation (Hedges’ g).
Reporting Checklist: Findings about Attrition
What to Report about Attrition (RCTs only)

**RCTs with Student Assignment**
- Number of individuals
  - Randomly assigned to T and C groups
  - In analytic sample, in each condition

**RCTs with Cluster Assignment**
- Number of clusters
  - Randomly assigned to T and C groups
  - In analytic sample, in each condition
- Number of individuals
  - For non-attrited clusters ONLY
  - Randomly assigned to T and C groups
  - In analytic sample, in each condition
What to Report:
Example Sample Sizes to Assess Attrition in Student RCT

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Comparison Group</th>
<th>Comparison Group</th>
<th>Treatment Group</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># Randomized</td>
<td># Analytic Sample</td>
<td># Randomized</td>
<td># Analytic Sample</td>
</tr>
<tr>
<td>Math standardized test</td>
<td>375</td>
<td>300</td>
<td>450</td>
<td>285</td>
</tr>
<tr>
<td>ELA standardized test</td>
<td>375</td>
<td>297</td>
<td>450</td>
<td>278</td>
</tr>
</tbody>
</table>
### What to Report: Example Sample Sizes to Assess Attrition in Cluster RCT

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Comparison Group</th>
<th>Comparison Group</th>
<th>Comparison Group</th>
<th>Comparison Group</th>
<th>Treatment Group</th>
<th>Treatment Group</th>
<th>Treatment Group</th>
<th>Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># randomized</td>
<td># randomized</td>
<td># randomized</td>
<td># randomized</td>
<td># randomized</td>
<td># randomized</td>
<td># randomized</td>
<td># randomized</td>
</tr>
<tr>
<td>Math test</td>
<td>30</td>
<td>30</td>
<td>375</td>
<td>300</td>
<td>30</td>
<td>30</td>
<td>450</td>
<td>285</td>
</tr>
<tr>
<td>ELA test</td>
<td>30</td>
<td>30</td>
<td>375</td>
<td>297</td>
<td>30</td>
<td>30</td>
<td>450</td>
<td>278</td>
</tr>
</tbody>
</table>

*a* Reported only for cluster-assignment evaluations. Not applicable for individual assignment evaluations.

*b* Number of students in *non-attrited* clusters only.
Reporting Checklist: Findings about Baseline Equivalence
## Findings – Baseline Equivalence: When to Report

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Report on Baseline Equivalence?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student RCT</strong></td>
<td></td>
</tr>
<tr>
<td>Low attrition</td>
<td>No</td>
</tr>
<tr>
<td>High attrition</td>
<td><strong>YES</strong>, for students</td>
</tr>
<tr>
<td><strong>Cluster RCT</strong></td>
<td></td>
</tr>
<tr>
<td>Low cluster attrition, low student attrition, &amp; allowed joiners or no joiners</td>
<td>No</td>
</tr>
<tr>
<td>High cluster attrition</td>
<td><strong>YES</strong>, for students or clusters</td>
</tr>
<tr>
<td>Low cluster attrition &amp; high student attrition</td>
<td><strong>YES</strong>, for students or clusters</td>
</tr>
<tr>
<td>With unallowed joiners</td>
<td><strong>YES</strong>, for students or clusters <em>(or show representativeness)</em></td>
</tr>
<tr>
<td><strong>QED</strong></td>
<td></td>
</tr>
<tr>
<td>Student-level assignment</td>
<td><strong>YES</strong>, for students</td>
</tr>
<tr>
<td>Cluster-level assignment</td>
<td><strong>YES</strong>, for students or clusters</td>
</tr>
</tbody>
</table>
What to Report: Baseline Equivalence (for each measure, by condition)

- **Sample Sizes**
  - Number of individuals in T and C groups for each analytic sample
  - For cluster design, number of clusters in T and C groups

- **Baseline Means**
  - For comparison group: Unadjusted baseline mean
  - For treatment group: Model-adjusted baseline mean (sum of comparison mean and treatment coefficient)

- **Baseline Standard Deviations**
  - Continuously-scaled baseline measure: Standard deviation (SD) for individuals in the T and C groups
  - Binary (or dichotomous) baseline measure: Not applicable

- **Baseline Mean Differences**
  - Coefficient for treatment variable in the baseline model or the difference between the unadjusted means for the T and C groups
  - Standardized baseline effect size (measured in standard deviation units)
Findings – Baseline Equivalence: Example Results from Baseline Equivalence Assessment

<table>
<thead>
<tr>
<th>Pretest Measure</th>
<th>Comparison Group</th>
<th>Treatment Group</th>
<th>Baseline Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample Size</td>
<td>Baseline Mean</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Pretest</td>
<td>350</td>
<td>80.8</td>
<td>16.4</td>
</tr>
<tr>
<td>SAT</td>
<td>378</td>
<td>1000</td>
<td>62</td>
</tr>
<tr>
<td>Low income</td>
<td>378</td>
<td>0.72</td>
<td>n/a</td>
</tr>
</tbody>
</table>

a The standardized difference is the baseline T-C difference divided by the pooled standard deviation (Hedges’ g).
Reporting Checklist: Findings about Representativeness
# Findings – Representativeness: When to Report

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Posttest?</th>
<th>Pretest?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student QED</strong></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Student RCT</strong></td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Cluster QED</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEQ established for individuals</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>BEQ NOT established for individuals</td>
<td><strong>YES</strong></td>
<td><strong>YES</strong></td>
</tr>
<tr>
<td><strong>Cluster RCT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEQ established for individuals</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>BEQ NOT established for individuals</td>
<td><strong>YES</strong></td>
<td><strong>YES</strong></td>
</tr>
<tr>
<td>• High Cluster Attrition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Lower Cluster Attrition, Unallowed Joiners</td>
<td><strong>YES</strong></td>
<td>No</td>
</tr>
<tr>
<td>• Lower Cluster Attrition, High Student Attrition</td>
<td><strong>YES</strong></td>
<td>No</td>
</tr>
</tbody>
</table>
What to Report: Representativeness (for each sample, by condition)

- **Sample Size**
  - Number of individuals contributing to posttest mean, for each condition (and baseline, if needed)

- **Number Enrolled**
  - Number of individuals present in clusters at posttest, for each condition (and baseline, if needed)

- **Posttest Dates**
  - Dates when posttest measured (and baseline, if needed)

- **Enrollment Dates**
  - Dates when enrollment measured
## Findings – Representativeness: Example Posttest Sample Sizes Needed to Assess Representativeness

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Treatment Group</th>
<th>Comparison Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of students contributing to posttest mean</td>
<td>Number of students enrolled in clusters</td>
<td>Dates outcome measured</td>
</tr>
</tbody>
</table>
Resources about Reporting
What Works Clearinghouse (WWC) Resources

- Reporting Guide for Study Authors
  - 6-page document with guidance for report on program effects
  - [Link to WWC Reporting Guide.pdf](#)

- WWC Handbooks
  - WWC Standards Handbook, version 4.1 (see p. 34-44 & Appendix A, if imputed missing data)
  - WWC Procedures Handbook, version 4.1
  - [Link to WWC Handbooks](#)

- WWC Review Protocols
  - [Link to WWC Review Protocols](#)

- WWC Standards and Process Briefs
  - [Link to WWC Briefs](#)
EIR TA Resources

- Using matching methods
  - Guidance document, power point, example datasets and programs in SAS and Stata

- Analysis models – program effects and baseline equivalence
  - Guidance documents, power point, example datasets and programs in SAS, Stata, SPSS, R

- Missing data
  - Power point, excel worksheet

- Reporting
  - Guidance document, power point
And Your TA Liaison in Here for You!
Contact

EIR Evaluation TA Team

EIREvalTA@AbtAssoc.com

www.abtassociates.com