



**2020 Education Innovation and Research (EIR) Project Directors
and Evaluators Technical Assistance Meeting**

Virtual Meeting | October 21–23, 2020

Reporting Your Findings with an Eye to a WWC Review

Anne Wolf (Presenter)

Barbara Goodson (Principal Investigator)

Abt Associates



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 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

Topic	What to Report
Intervention	<ul style="list-style-type: none">✓ Key intervention activities/inputs✓ Participant and setting characteristics
Comparison	<ul style="list-style-type: none">✓ What comparison group members received✓ What they could not receive
Study Design	<ul style="list-style-type: none">✓ Method of assignment✓ Unit of assignment✓ Use of blocking
Measures	<ul style="list-style-type: none">✓ Reliability & face validity✓ Not over-aligned with treatment✓ Consistently collected in both conditions
Analysis Approach	<ul style="list-style-type: none">✓ Statistical model (clustering, blocking, baseline covariates)✓ Approach to handling missing data✓ Approach to establishing baseline equivalence, if necessary
Findings	<ul style="list-style-type: none">✓ Effects✓ Attrition✓ Baseline Equivalence✓ Representativeness

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*

Analysis Approach – Statistical Model: Example Model for a Cluster Design

EXAMPLE

The hierarchical linear model took the following form:

Level-1 (student-level):
$$Y_{ij} = \beta_{0j} + \beta_1' X_{ij} + \varepsilon_{ij}$$

Level-2 (course-section level):
$$\beta_{0j} = \gamma_{00} + \gamma_{01} TRT + \sum_{m=2}^M \gamma_{0m} Cohort_j + \mu_j$$

Where:

Y_{ij} = outcome score for student i in course section j

β_{0j} = outcome score in control course section j

β_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)

ε_{ij} = error term for student i in course section j

γ_{00} = average outcome score in control courses

γ_{01} = outcome score difference between treatment course sections and control course sections

Trt_j = treatment status for course section j

γ_{0m} = vector of parameters for the effects of cohort

$Cohort_j$ = set of dummy variables indicating cohort A, B, or C for course section j

μ_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

Measure	Treatment Group				Comparison Group				Estimated Effect			Effect Size ^a
	N schools	N students	Posttest Mean	Std. Dev.	N schools	N students	Posttest Mean	Std. Dev.	T-C Difference	Std. Error	P-value	
Math: State Test	30	285	84.4	17.5	30	300	80.8	16.4	3.62	0.22	0.017	.21
ELA: State Test	30	278	95.1	15.9	30	297	90.6	16.2	4.54	0.28	0.003	.28

^a 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

	Comparison Group	Comparison Group	Treatment Group	Treatment Group
	#	#	#	#
Outcome Measure	Randomized	Analytic Sample	Randomized	Analytic Sample
Math standardized test	375	300	450	285
ELA standardized test	375	297	450	278

What to Report: Example Sample Sizes to Assess Attrition in Cluster RCT

	Comparison Group	Comparison Group	Comparison Group	Comparison Group	Treatment Group	Treatment Group	Treatment Group	Treatment Group
	Clusters ^a	Clusters ^a	Students ^b	Students ^b	Clusters ^a	Clusters ^a	Students ^b	Students ^b
Outcome Measure	# randomized	# analytic sample	# randomized	# analytic sample	# randomized	# analytic sample	# randomized	# analytic sample
Math test	30	30	375	300	30	30	450	285
ELA test	30	30	375	297	30	30	450	278

^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*

Study Design	Report on Baseline Equivalence?
Student RCT	
Low attrition	No
High attrition	YES , for students
Cluster RCT	
Low cluster attrition, low student attrition, & allowed joiners or no joiners	No
High cluster attrition	YES , for students or clusters
Low cluster attrition & high student attrition	YES , for students or clusters
With unallowed joiners	YES , for students or clusters <i>(or show representativeness)</i>
QED	
Student-level assignment	YES , for students
Cluster-level assignment	YES , for students or clusters

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

Pretest Measure	Comparison Group			Treatment Group			Baseline Difference	
	Sample Size	Baseline Mean	Standard Deviation	Sample size	Baseline Mean	Standard Deviation	T-C Mean Difference	Standardized Difference ^a
Pretest	350	80.8	16.4	420	84.4	17.5	3.62	0.22
SAT	378	1000	62	432	990	58	-10.00	-0.17
Low income	378	0.72	n/a	432	0.77	n/a	0.05	0.16

^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*

Study Design	Posttest?	Pretest?
Student QED	No	No
Student RCT	No	No
Cluster QED	No	No
BEQ established for individuals	No	No
BEQ NOT established for individuals	YES	YES
Cluster RCT	No	No
BEQ established for individuals	No	No
BEQ NOT established for individuals	YES	YES
<ul style="list-style-type: none"> • High Cluster Attrition 	YES	YES
<ul style="list-style-type: none"> • Lower Cluster Attrition, Unallowed Joiners 	YES	No
<ul style="list-style-type: none"> • Lower Cluster Attrition, High Student Attrition 	YES	No

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

	Treatment Group				Comparison Group			
Outcome	Number of students contributing to posttest mean	Number of students enrolled in clusters	Dates outcome measured	Dates enrollment measured	Number of students contributing to posttest mean	Number of students enrolled in clusters	Dates outcome measured	Dates enrollment measured
Student engagement survey	589	616	May 2016	Oct. 2015	691	739	May 2016	Oct. 2015
School climate survey	596	616	May 2016	Oct. 2015	700	739	May 2016	Oct. 2015

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 is Here for You!

Questions?



Contact

EIR Evaluation TA Team

EIREvalTA@AbtAssoc.com



**BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT**

www.abtassociates.com