



## Example Evaluation Plan for a Cluster Randomized Controlled Trial

The Evaluation Plan Template identifies the key components of an evaluation plan and provides guidance about the information typically included in each section of a plan for evaluating both the effectiveness and implementation of an intervention. Evaluators can use this tool to help develop their plan for a rigorous evaluation, with a focus on meeting *What Works Clearinghouse*<sup>TM</sup> evidence standards. The template can be used in combination with the Contrast Tool, a tool for documenting each impact that the evaluation will estimate to test program effectiveness.

This document provides an example of a detailed evaluation plan for evaluating the effectiveness of an intervention. Developed using the Evaluation Plan Template, the plan is for a randomized controlled trial (RCT) in which clusters (i.e., class sections, in this example) are randomly assigned to an intervention or a control condition. This example illustrates the information that an evaluator should include in each section of an evaluation plan, as well as provides tips and highlights key information to consider when writing an evaluation plan for a cluster RCT. Accompanying this example evaluation plan is the Example Contrast Tool for a Cluster RCT, which lists each impact that the example evaluation will estimate to test program effectiveness. The example Evaluation Plan and the example Contrast Tool can be reviewed side-by-side.

### Prepared by:

**Abt Associates**

Anne Wolf

Michael Frye

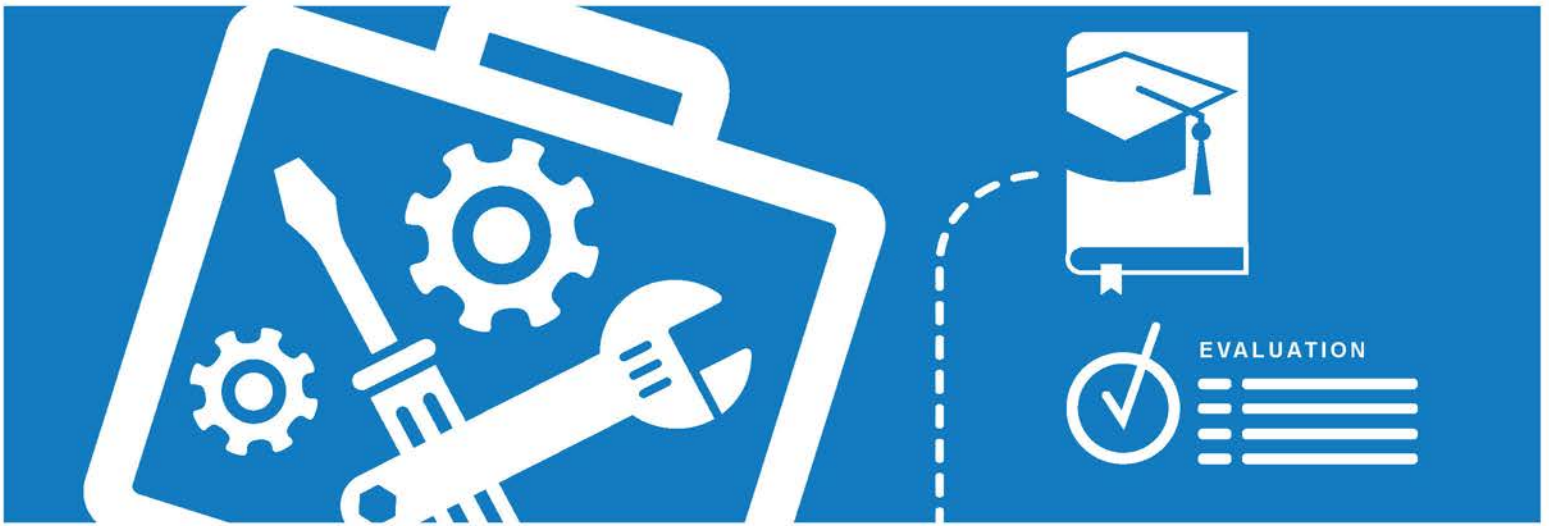
Barbara Goodson

Cristofer Price

Beth Boulay



August 2016



**The Institute of Education Sciences has made this tool publicly available as a courtesy to evaluators. However, the content of this tool does not necessarily represent IES’s views about best practices in scientific investigation.**

This tool was developed under U.S. Department of Education Institute of Education Sciences (IES) contracts ED-IES-10-C-0064 with Abt Associates and ED-ODS-12-A-0019/0031 with AEM Corporation and its subcontractor Abt Associates. These contracts provided evaluation technical assistance for evaluations of interventions funded by the Investing in Innovation and First in the World programs. Tools, webinars, and other materials were developed to help grantees engage in good scientific practice and produce evaluations that meet *What Works Clearinghouse*<sup>TM</sup> evidence standards.





# CONTENTS

## Grantee: Southeast Technical University (SETU)

<b>1. Evaluator Information</b>	<b>1</b>
<b>1.1 Contact Information</b>	<b>1</b>
<b>1.2 Independence</b>	<b>1</b>
<b>1.3 Confidentiality Protections</b>	<b>1</b>
<b>2. Summary of Interventions</b>	<b>2</b>
<b>3. Impact/Effectiveness Evaluation</b>	<b>3</b>
<b>3.1 Research Questions</b>	<b>3</b>
<b>3.2 Comparison Condition</b>	<b>4</b>
<b>3.3 Study Sample and How Intervention and Comparison Groups are Selected/Assigned</b>	<b>4</b>
Eligibility and Recruitment of Study Participants	4
Random Assignment	5
Expected Sample Sizes	6
Progression of Students and Cohorts Across Multiple Years	7
<b>3.4 Key Measures and Plan for Obtaining Data</b>	<b>9</b>
Data Collection	9
Analytic Measures	9
<b>3.5 Statistical Analysis of Impacts</b>	<b>13</b>
Impact Analysis Model	14
Handling Missing Data	15
Adjusting for Multiple Comparisons	16
Cross-Overs/No-Shows	17
<b>3.6 Attrition (RCTs only)</b>	<b>17</b>
<b>3.7 Baseline Equivalence Testing (QEDs and RCTs with High Attrition)</b>	<b>18</b>
<b>4. References</b>	<b>21</b>



# EXAMPLE EVALUATION PLAN

## 1. Evaluator Information

### 1.1 Contact Information

World's Best Evaluators (WBEval)  
123 Main Street  
Washington, DC 20001

Principal Investigator: John Doe  
[jdoe@wbeval.com](mailto:jdoe@wbeval.com)  
(202) 888-1111

### 1.2 Independence

WBEval will be the independent evaluator of the *FACT* intervention. In this role, WBEval will independently conduct all key aspects of the evaluation. Specifically, WBEval will be responsible for executing the random assignment, collecting and analyzing student outcome data, and estimating and reporting program impacts on the student outcomes. The evaluation findings will not be subject to the approval of the project developer/grantee before being reported.

### 1.3 Confidentiality Protections

WBEval and Southeast Technical University (SETU) both have internal IRB processes that uphold rigorous standards related to the protection of human subjects. Both IRBs have reviewed and approved the research activities associated with *FACT* evaluation.

WBEval and SETU will securely store and handle any materials containing sensitive data; these materials will be limited to staff working directly on the projects. Whenever possible, data will be recorded in unidentified ways. No individuals will be identified in any reports.



## 2. Summary of Interventions

The Flipped + App Classroom Teaching (*FACT*) approach to introductory math (the Applied Mathematics course) is designed to: (1) equip students with the skills they need to succeed in college-level math, (2) develop the ability of students to solve complex STEM-related problems, and (3) foster students' interest in STEM careers and issues. The intervention is intended to address barriers that discourage college students from pursuing and successfully completing a STEM-related degree. Even among students who place into college-level math in their first semester at community college, a substantial portion perceive themselves as lacking the skills to succeed in math courses at the college level. Further, they lack experience in tackling complex problems related to math and other STEM disciplines. By increasing opportunities to work on complex problems, students will develop better problem-solving skills. These enhanced skills are vital, particularly for students who wish to pursue a STEM related degree. In addition, it is hypothesized that as students gain experience and develop skills with STEM-related problems, they will become more interested in pursuing STEM careers, even students who have previously been uninterested.

The *FACT* intervention involves two key strategies – flipped classroom instruction and a mobile app to support an out-of-class online learning community. *FACT* courses utilize the flipped classroom format – class time is used for students to complete assignments; course content is presented through online videotaped lectures and reading done outside of class, on the students' own time. By flipping the traditional instructional format of in-class lectures and out-of-class assignments, students are able to engage in active thinking and problem solving when they are together with peers and the professor. Lectures and readings completed outside of class form the foundation for active learning in class. During class, students work on their own and with each other to complete assignments. Assignments rely on problem-based-learning strategies to engage students and encourage them to collaborate on problems and “think outside the box.” The flipped classroom format is designed to increase student learning, engagement, and interest in STEM related fields. The content of the lectures and in-class assignments was developed by SETU professors from the Science and Education schools.

As part of the *FACT* intervention, the flipped classroom is supplemented by a mobile app. The *FACT* app has the following features:

- Brief questionnaires that allow students to check their own understanding of the lecture content. These questionnaires are optional and do not contribute to students' grades. However, students have a profile for the app and they can “level up” if they answer a certain number of questions correctly.
- Systems for students to (a) be part of a class forum where students can post questions and pose ideas that can be viewed by the class and the professor, and (b) to connect with other students and the professor through private messages (similarly to other social networking apps).

The theory behind the app is that by allowing students to connect via their mobile devices, the out-of-classroom lectures will be more engaging than the standard online video format of flipped classroom lectures. The app is being designed by an independent technology developer, who will coordinate with the professors who developed *FACT* to produce the app content.



## 3. Impact/Effectiveness Evaluation

WBEval will conduct an impact evaluation of the *FACT* model at Southeast Technical University (SETU) over the 4 years of the grant. The evaluation will use a multi-cohort, cluster-randomized trial design to examine effects of *FACT* on student achievement in mathematics, interest in STEM, and enrollment in STEM courses. The *FACT* intervention will be implemented in the one-semester Applied Mathematics courses. Applied Mathematics is required of all students who are not in Mathematics or Engineering programs or in need of math remediation (Calculus or Remedial Math are required of the other two respective groups). Students will then be followed through their fourth semester of enrollment. Teachers who have at least two sections of Applied Mathematics per semester will be recruited to participate in the study. For all interested teachers, their Applied Math course sections will be randomly assigned – half to the treatment condition and half to the control condition. Consequently, each participating teacher will have at least one treatment course section and one control course section. Assignment of course sections will occur after students have enrolled in the Math course and been assigned to sections. The students who are already enrolled in the course sections at the time of random assignment will comprise the student sample. Students who enroll in any of the study sections after random assignment will not be included in the study. However, all students who enroll in a *FACT* section will be able to participate in *FACT* even though they are not in the study sample.

### 3.1 Research Questions

The evaluation will address the outlined below. Research questions are also listed in the accompanying contrast tool.

#### TIP!

In your evaluation plan...

- Outline specific, narrowly defined research questions that will be addressed by the study.
- Have a research question for each specific test of the intervention effect.

In this example there are 7 such tests—even though there are only 4 different outcome measures, because the effect of the intervention is tested on some outcome measures at multiple points in time.

The first two questions examine effects on *math achievement and problem-solving skills*.

- For three cohorts of entering college freshmen required to take Applied Math, what is the effect of a *FACT* course section on earning a *course grade of C or higher in Applied Math* compared to a business-as-usual course section?
- For three cohorts of entering college freshmen required to take Applied Math, what is the effect of a *FACT* course section on *math problem-solving skills* at the end of the end of the semester (measured by the Math Applications & Concept Inventory) compared to a business-as-usual course section?



# EXAMPLE EVALUATION PLAN

One question examines the effect on *interest in STEM*.

- For three cohorts of entering college freshmen required to take Applied Math, what is the effect of a *FACT* course section on *interest in STEM* at the end of the semester compared to a business-as-usual course section?

Four questions examine effects on *credit accumulation*.

- For three cohorts of entering college freshmen required to take Applied Math, what is the effect of a *FACT* course section on *total credits earned in STEM courses* at the end of students' 2<sup>nd</sup> semester in community college compared to a business-as-usual course section?
- For two cohorts of entering college freshmen required to take Applied Math, what is the effect of a *FACT* course section on *total credits earned in STEM courses* at the end of students' 3<sup>rd</sup> semester in community college compared to a business-as-usual course section?
- For two cohorts of entering college freshmen required to take Applied Math, what is the effect of a *FACT* course section on *total credits earned in STEM courses* at the end of students' 4<sup>th</sup> semester in community college compared to a business-as-usual course section?
- For two cohorts of entering college freshmen required to take Applied Math, what is the effect of a *FACT* course section on *declaring a STEM major* by the end of students' 4<sup>th</sup> semester in community college compared to a business-as-usual course section?

## 3.2 Comparison Condition

The comparison condition will consist of business-as-usual instruction in the Applied Mathematics classes. In business-as-usual classrooms, teachers will deliver instruction in the traditional lecture format, with homework, readings, and group assignments to be conducted on the students' own time. The students in the business-as-usual classrooms will not have access to the *FACT* app.

## 3.3 Study Sample and How Intervention and Comparison Groups are Selected/Assigned

The study design is a multi-cohort cluster-randomized trial. Each fall semester for three consecutive years, 2017, 2018, and 2019, all class sections of interested Applied Math faculty will be randomly assigned to either the treatment or control condition. Information about planned analytic samples is also shown in the accompanying contrast tool, on the “samples” tab.

## Eligibility and Recruitment of Study Participants

**Applied Mathematics faculty.** Teachers will be recruited to participate if they teach at least two sections of Applied Mathematics in the same semester. Recruitment will happen in late spring and early summer of 2016. Teacher participants will be trained in summer 2016 to use the *FACT* model. We estimate 8-10 teachers will participate. We expect the same teachers to continue to participate each year and do not plan to recruit additional teachers in years 2 and 3. Teachers will not receive additional training after summer 2016. Study teachers who leave the school or drop out of the study will not be replaced.



# EXAMPLE EVALUATION PLAN

**Applied Mathematics class sections.** All sections of Applied Mathematics taught by participating teachers will be included in the evaluation.

**Students.** At SETU, all entering freshmen who are not in need of math remediation and are not majoring in Mathematics or Engineering are required to take Applied Mathematics, an introductory math course that is only offered in the fall semester. Freshmen entering SETU in fall 2016, fall 2017, or fall 2018 will be included in the evaluation if they are required to take Applied Mathematics and are enrolled in an Applied Mathematics section taught by a trained FACT teacher. The student sample for each course section will be defined as students who have enrolled by one week before the course starts. Students who enroll in study class sections after the start of the semester (e.g., during “add/drop” period) will be excluded from the study sample.

## TIP!

- ☑ Provide information about when the baseline sample of students – from which attrition will be measured – will be identified relative to the timing of random assignment. For example, is the sample defined as students in clusters at the time of random assignment, students enrolled on the first day of class (after randomization), students enrolled after the add/drop period (after randomization).
- ☑ Be clear about whether the sample will include students who join clusters after random assignment. The WWC may consider students to be joiners unless there is clear evidence that the students were already members of the cluster before randomization (e.g., based on enrollment lists for a time prior to random assignment).

In this example, the sample is defined as students enrolled in course sections one week before classes start, when random assignment will occur, so the sample will not include joiners. If the sample includes all students who joined clusters after random assignment, the WWC may consider there to be *risk of bias in the sample due to the presence of joiners*.

When the sample includes students who join clusters after randomization, it may be possible for students to knowingly self-select into clusters (i.e., sections) based on the assignment condition. If so, it is not possible to determine whether differences in outcomes between the treatment and control groups are due to the intervention alone, or due to differences between students introduced after random assignment.

The determination of whether the sample includes joiners and the timing of their joining has ramifications for the WWC evidence rating a study has the potential to receive. For more information, see [WWC Revised Standards for the Review of Cluster Design Studies](#) for planned revisions to the standards.

## Random Assignment

Random assignment will be conducted in the fall semester (the semester when Applied Mathematics is offered) for three consecutive cohorts – fall 2016, fall 2017, and fall 2018. Each fall semester, random assignment will occur approximately one week before courses start. All class sections taught by study teachers will be randomly assigned to either the treatment or control condition with equal probability using the following procedure:





# EXAMPLE EVALUATION PLAN

1. Study teachers will be randomly sorted.
2. For each study teacher, class sections will be randomly sorted.
3. For each study teacher, the first class section in the randomly-sorted list of class sections will be randomly assigned to treatment or control (50/50 chance of being in either condition), and
4. The subsequent sections in the randomly-sorted list for that teacher will be assigned in alternating fashion (e.g. treatment, control, treatment, control, treatment).

For teachers with an even number of class sections, half will be randomly assigned to treatment and half to control. For teachers with an odd number of class sections, the extra class section will have an equal probability of being randomly assigned to the treatment or control condition. For example, if a teacher has three sections, there will be a 50% probability of two control sections and one treatment section and a 50% probability of one control section and two treatment sections.

Within each of three fall semesters (or “cohorts”), random assignment of course sections will be conducted within teacher. With 3 cohorts and 8 to 10 teachers, there will be a total of 24 to 30 randomization blocks.

## Expected Sample Sizes

Exhibit 1 presents a summary of the number of course sections and students, by cohort, by condition, and overall. Assuming 8 to 10 teachers and an average of four class sections per teacher, there will be an estimated 32 to 40 sections randomized each fall: 16 to 20 treatment sections and 16 to 20 control sections. Across the three fall semesters, there will be a total of 96 to 120 class sections randomized, with about 48 to 60 sections in each condition.

With approximately 20 students enrolled in each Applied Mathematics section, the student sample size for analyses including all three cohorts is expected to be between 1,920 (8 teachers x 4 sections x 20 students x 3 fall semesters) and 2,400 (10 teachers x 4 sections x 20 students x 3 fall semesters) students, with about 960 to 1,200 students in each condition. For analyses based on two cohorts of students, the total sample size is expected to be 1280-1600 students, with approximately 640-800 students in each condition.

### KEEP IN MIND...

Students may need to provide consent to:

1. Participate in the intervention,
2. Participate in random assignment, and
3. Participate in data collection.

Consent for each activity may be linked or kept separate. It's important to know which of these activities require consent and for each, when consent will happen relative to randomization.

In this example, student consent is not required for participation in any of these three activities. Students consent to the instructional approach (i.e., the intervention) by enrolling in the class. Student consent is not required for random assignment, because class sections are randomized. Student consent is not required for participation in data collection, because students are not asked to participate in any data collection activities outside of usual classroom practice.

### TIP!

- If a study includes multiple cohorts, be sure to specify when random assignment occurs for each cohort.



# EXAMPLE EVALUATION PLAN

**Exhibit 1. Expected number of faculty, course sections, and students in the evaluation**

	Faculty	Course Sections (4 per teacher)	Students (20 per section)
Cohort A	8-10	32-40	640-800
Cohort B	n/a	32-40	640-800
Cohort C	n/a	32-40	640-800
Full Sample	8-10	96 - 120	1,920 – 2,400
Treatment Group		48 - 60	960 – 1,200
Control Group		48 - 60	960 – 1,200

## Progression of Students and Cohorts Across Multiple Years

Exhibit 2 shows the semester-by-semester progression of students over time, for the three study cohorts that will be included in the evaluation of *FACT*. Although the intervention lasts one semester, the evaluation will follow students and measure the effects of the intervention on outcomes through each student’s fourth semester of enrollment in community college.

The three cohorts of students will be combined for analyses, and the effects of *FACT* will be examined at the end of students’ first semester, second semester, third semester, and fourth semester in community college. The timing of each outcome data collection will vary by cohort, as shown in Exhibit 2. The last round of data collection is spring 2019. This schedule means that the study will be able to follow the first two cohorts through their fourth semester of college, but will only be able to follow the third cohort of students through their third semester in college. Therefore, only the first two cohorts will be able to contribute to the analysis of outcomes at the end of the fourth semester. The samples for the analyses of each outcome are as follows:

- *1<sup>st</sup> semester outcomes*: Fall 2016 for cohort A, fall 2017 for cohort B, and fall 2018 for cohort C (cells A1, B1, and C1).
- *2<sup>nd</sup> semester outcomes*: Spring 2017 for cohort A, spring 2018 for cohort B, and spring 2019 for cohort C (cells A2, B2, and C2).
- *3<sup>rd</sup> semester outcomes*: Fall 2017 for cohort A and fall 2018 for cohort B (cells A3 and B3). Cohort C will not reach semester 3 before the end of the evaluation.
- *4<sup>th</sup> semester outcomes*: Spring 2018 for cohort A and spring 2019 for cohort B (cells A4 and B4).

### TIP!

- ☑ If an intervention spans multiple semesters or multiple years, and/or if the study includes multiple cohorts, clearly describe the progression of students (and cohorts) over time.
- ☑ Indicate when students will receive the intervention, and when data on outcomes, baseline measures, and covariates will be collected.
- ☑ Include a chart, table or other graphic (like Exhibit 2) to clearly show how students (and cohorts) progress from year to year (or semester to semester) relative to the timing of the intervention and collection of outcome data.
- ☑ State when impacts will be assessed in relation to (a) the amount of intervention exposure/length of follow-up; (b) student “grade,” and/or (c) how long the intervention has been in place.



# EXAMPLE EVALUATION PLAN

Also, see the “samples” tab in the accompanying contrast tool for a summary of detailed information about the planned analytic samples.

## TIP!

Clearly state whether the cohorts will be combined or analyzed separately.

Combining cohorts will increase the sample size and improve statistical power for detecting intervention effects. However, if there are differences in the intervention for different cohorts, you may want to analyze cohorts separately. But be aware, the WWC may adjust for multiple comparisons if cohorts are analyzed separately. For more information, see [WWC Procedures and Standards Handbook \(version 3.0\).pdf](#), p. 25-26 and Appendix G.

**Exhibit 2. Progression of Students over Time, by Cohort**

Academic Year	Semester	Cohort A	Cohort B	Cohort C
2015-16	Pre-intervention (Year 0)	A0 Pre-intervention No exposure		
	Fall 2016 (Semester 1)	A1* 1 <sup>st</sup> semester (Freshman Year) 1 semester exposure		
2016-17	Spring 2017 (Semester 2)	A2* 2 <sup>nd</sup> semester (Freshman Year) 1 semester post-intervention	B0 Pre-intervention No exposure	
	Fall 2017 (Semester 3)	A3* 3 <sup>rd</sup> semester (Sophomore Year) 2 semesters post-intervention	B1* 1 <sup>st</sup> semester (Freshman Year) 1 semester exposure	
2017-18	Spring 2018 (Semester 4)	A4* 4 <sup>th</sup> semester (Sophomore Year) 3 semesters post-intervention	B2* 2 <sup>nd</sup> semester (Freshman Year) 1 semester post-intervention	C0 Pre-intervention No exposure
	Fall 2018 (Semester 5)		B3* 3 <sup>rd</sup> semester (Sophomore Year) 2 semesters post-intervention	C1* 1 <sup>st</sup> semester (Freshman Year) 1 semester exposure
2018-19	Spring 2019 (Semester 6)		B4* 4 <sup>th</sup> semester (Sophomore Year) 3 semesters post-intervention	C2* 2 <sup>nd</sup> semester (Freshman Year) 1 semester post-intervention

**Note: Cell labels indicate the cohort (A, B, or C) and students’ semester in community college (1-4). Asterisks (\*) indicate the timing of outcome data collection for each cohort. Shading indicates the timing of the intervention for each cohort.**



# EXAMPLE EVALUATION PLAN

## 3.4 Key Measures and Plan for Obtaining Data

In this section, we describe data collection and the variables that will be examined in the analysis.

### Data Collection

There are five student outcomes: grade in Applied Math, math problem-solving skills, number of STEM credits earned, major in a STEM discipline, and interest in STEM. With the exception of the measure of student interest in STEM, which will be collected via a student survey administered for the study, outcome data will be collected from SETU administrative files on students.

**Administrative data.** The outcome data from SETU administrative files will be transferred to the evaluator annually each summer during the study – summer 2017, summer 2018, and summer 2019. Note that all data from SETU administrative records will be transmitted to WBEval via SETU’s secure file transfer protocol (FTP) site. Data files will include administrative data for all students in study course sections. Using rosters of students enrolled in study course sections at the time of random assignment, evaluators will identify those students to be included in the evaluation sample. Students not on rosters at the time of random assignment will not be included in the evaluation sample.

Administrative files transferred to the evaluator will include data on students’ grade and completion status for Applied Mathematics, students’ score on the applied mathematics final exam (the MACI), number of STEM credits attempted and earned each semester, students’ major, student demographic characteristics, and academic performance prior to matriculation.

**Student survey.** All students in the study course sections will be administered the STEM Interest Survey Instrument (SISI) as part of the Applied Mathematics course. The participating teachers will administer the SISI survey to their students twice – once in the first two weeks of the course and once in the last two weeks of the semester.

### Analytic Measures

Below we describe the outcome measures, baseline measures, and other independent variables that will be used in analyses of the impacts of *FACT*. Information about the planned analytic measures is also provided in the accompanying contrast tool, on the “outcomes” and “baseline measures” tabs.

**Outcome measures.** The evaluation will examine intervention effects on five outcome measures – four of which will be obtained from university administrative data. One will be constructed from the STEM Interest Survey Instrument administered to students at the end of the Applied Mathematics course.



## EXAMPLE EVALUATION PLAN

The following four variables will be constructed from administrative data files:

- **Applied Mathematics course grade of C or better:** Students' final course grade in Applied Mathematics will be classified into two categories – (1) a course grade of C or higher or (2) a course grade of D or lower, including withdrawal. Both students who withdraw from the course and remain at SETU and students who withdraw from the university before completing the semester will be coded as not achieving a grade of C or better, instead of being assigned a missing value for the outcome.
- **Math problem solving:** Math problem-solving skills will also be measured using the Math Applications and Concepts Inventory (MACI; Smith & Jones, 2009), which serves as the final exam for the Applied Mathematics course. A student's score on the MACI comprises 40% of his/her grade in the course. The MACI assessment was developed by content and measurement experts. It gauges the application of concepts in calculus, algebra, geometry, and statistics and probability, including mastery of skills to solve applied math problems. The MACI assessment is designed to focus the ability to apply concepts, rather than the theoretical underpinnings. The MACI produces a total score, which has demonstrated reliability – the test-retest reliability estimate is  $R = 0.92$ . Students who withdraw from the course or the university before taking the MACI will be assigned a missing value for the outcome. Students who do not take the MACI for other reasons will be assigned a score of 0.
- **Number of credits earned in STEM courses:** The cumulative number of credits earned in STEM courses will be measured for students in their second, third, and fourth semester in community college. Students can elect to take a variety of STEM-related courses and may vary considerably in the number they choose to take. For students who graduate from SETU (a 2-year community college) prior to their 4<sup>th</sup> semester, the total number earned as of graduation will be counted for any subsequent semesters. Students who withdraw from SETU or transfer to another college before the end of the course will be assigned a missing value for any semester after they leave SETU.

### TIP!

- ☑ Clearly define outcome measures for students who leave a study institution before the outcome data are collected, especially for studies of postsecondary interventions.

In this example, the evaluator distinguishes between when outcome data will be treated as missing and when the outcome will be defined based on the last semester of enrollment.

- ☑ Treat data as missing for outcome measures that would have a different value for students that leave a study institution if data are available from sources outside the study institution (or attempt to obtain the data from other sources).

In this example, a student who leaves SETU after his/her third semester may go on to major in a STEM-related discipline at another institution – without such data, his/her outcome for “major in a STEM discipline” would be missing.



## EXAMPLE EVALUATION PLAN

- **Major in STEM discipline.** Each student's major in their 4<sup>th</sup> semester in college will be classified as being either in a STEM discipline or in a non-STEM discipline. Students who are still enrolled at SETU but have not declared a major three semesters after their Applied Mathematics course (i.e., in their 4<sup>th</sup> semester in college) will be coded as not having a STEM major. For students who have graduated from SETU (a 2-year community college) or are known to have transferred to a 4-year college prior to the third semester after completing Applied Math, their major at the time of graduation or transfer will be classified as STEM or non-STEM. Students who have withdrawn from SETU before their 4<sup>th</sup> semester will be coded as missing this outcome.

One outcome measure will be constructed from a student survey.

- **Interest in STEM** will be measured by the STEM Interest Survey Instrument (SISI; Johnson, 2011). The SISI was developed by educational psychologists, and is intended to measure student interest in STEM topics and STEM related career paths. The SISI is comprised of 20 items scored on a Likert-type scale, which are combined to create a single scale score. The reported internal consistency estimate for the SISI total score is  $\alpha = .85$ .

Exhibit 3 summarizes the domains, outcomes, and measurement timeline. In addition, Exhibit 3 lists, for each outcome, anticipated baseline measures to be used to assess the equivalence of treatment and control students prior to the start of the intervention, in case of high attrition from the initial sample.

### TIP!

When using surveys...

- ☑ Describe what scale(s) will be constructed from the survey and used in the analysis, not just the survey.
- ☑ Report reliability data for the actual scales that will be used as analytic variables.
- ☑ Use published reliability data for existing measures, if available.
- ☑ Calculate reliability using study data when the outcome measure is a newly-developed measure or comprised of only a subset of items from an existing scale, as published reliability data may only be available for the full measure (i.e., not for the particular subset).
- ☑ There's no need to report reliability for scales that will not be analyzed.

### KEEP IN MIND...

Although you may choose to define your outcome domains differently than the WWC does, it's important to be aware of how your outcomes will be classified by the WWC, because the WWC will apply multiple comparisons adjustments for multiple impacts estimated in the same domain.

For more information, see the relevant [WWC Topic Area Review Protocol](#).



# EXAMPLE EVALUATION PLAN

**Exhibit 3. Outcome domains, measures, timing of data collection, and associated baseline measures**

Domain	Outcome Measure	Timing of Measurement	Baseline Measure(s)
Math Achievement	Applied Math course grade of C or better	End of 1st semester (cohorts 1, 2, 3)	Baseline math test (SAT or placement test) Pell grant eligibility
	Math Applications & Concept Inventory (MACI)	End of 1st semester (cohorts 1, 2, 3)	
Credit Accumulation	Total number of credits earned in STEM courses	End of 2nd semester (cohorts 1, 2, 3) End 3rd & 4th semester (cohorts 1, 2)	Baseline math test (SAT or placement test) Pell grant eligibility
	Major in STEM discipline	End of 4th semester (cohorts 1, 2)	
Interest in STEM	STEM Interest Survey Instrument (SISI)	End of 1st semester (cohorts 1, 2, 3)	STEM Interest Survey Instrument (SISI)

**Baseline measures.** We will assess the equivalence of the intervention and control groups on the relevant baseline measures (as shown in Exhibit 3). For outcomes in both the math achievement and credit accumulation domains, we will assess baseline equivalence on the same two measures: a baseline measure of academic achievement and a baseline measure of socioeconomic status. The baseline measure of socioeconomic status is student Pell grant eligibility status. Students will be classified as eligible or not eligible for a Pell grant.

The baseline measure of math achievement will be constructed from either students' math placement test, which is taken upon entry into SETU, or students' math SAT, which was taken in high school. There will be entering students who have not taken the placement exams and those who have not taken the SAT in high school, but all students will have either a math placement test or math SAT score. For students who have taken both tests, the math placement test will be used. For each test, standardized z-scores will be constructed so that a single measure of baseline math achievement can be constructed from the two tests – SAT math and math placement exam. As a result, all students who took either the SAT or the math placement test will have a baseline measure of academic achievement.

## TIP!

- ☑ Be sure to have a baseline measure for each student in the analytic sample.
- ☑ Consider combining different tests that measure the same construct (e.g., math placement test and math SAT, in this example), if all students do not have the same baseline measure. Measures will have to be rescaled to a common metric in order to be combined.
- ☑ Think carefully about what measures can be combined sensibly. Ask yourself whether they are measuring the same thing.



# EXAMPLE EVALUATION PLAN

## KEEP IN MIND...

For cluster designs that must establish baseline equivalence (i.e., cluster RCTs with high attrition or with student joiners), it may be possible to use *cluster* means to establish baseline equivalence instead of using baseline measures for the individual students in the analytic sample. This may be a good option if baseline data on individual students are not available. For more information, see the [WWC Procedures and Standards Handbook \(version 3.0\)](#), p. 16. Also, see [WWC Revised Standards for the Review of Cluster Design Studies](#) for planned revisions to the standards, which may also require studies to demonstrate that the students contributing to cluster means are representative of the cluster.

In addition, student's interest in STEM will be assessed using the same SISI scale used as an outcome measure. As noted above, the SISI is administered at the start of the Applied Mathematics course.

**Independent variables.** The other independent variables to be included in analyses are:

- *Treatment indicator:* A variable indicating the group to which a course section was randomly assigned (0 = control group, 1 = *FACT*).
- *Randomization blocks:* Dummy variables to account for teacher and cohort. For 8-10 teachers and 3 cohorts, there are a total of 24-30 randomization blocks.
- *Gender:* A dummy variable indicating whether a student is male or female.
- *Race:* A categorical variable indicating whether student is white, black or African American, Asian, or other.
- *Ethnicity:* A dummy variable indicating whether a student is Hispanic or not Hispanic.
- *First-time student status:* A dummy variable indicating whether a student is a first-time freshman or transfer student upon entering SETU.

## 3.5 Statistical Analysis of Impacts

We will use hierarchical linear models (HLM) to estimate the impact of the *FACT* intervention on math achievement, STEM interest, credits earned in STEM courses, and STEM major, adjusting for randomization blocks and baseline covariates. The impact of *FACT* will be estimated at the course section level, which is the level of assignment. See the accompanying contrast tool for information about each impact of the intervention's effect that will be estimated to address the study research questions, which is shown in the "contrasts" tab.

### TIP!

- ☑ Account for cluster assignment in the analytic model. Outcomes for students grouped together in the same cluster (e.g., class, school) are likely to be correlated. If models do not adjust for clustering, standard errors may be underestimated.
- ☑ Use methods such as multilevel modeling (HLM), Huber-White Sandwich estimator, or GEE (e.g., Stata's "cluster" option) to adjust standard errors for clustering. If you do not, the WWC will apply a post-hoc correction to the standard error of your impact estimate, which will likely be more conservative (i.e., resulting in a larger p-value) than the adjustment you apply based on your sample data.





# EXAMPLE EVALUATION PLAN

## Impact Analysis Model

The model specified below will be used to estimate the impact of the *FACT* intervention.

Level-1 (student-level):

$$Y_{ij} = \beta_{0j} + \sum_{k=1}^K \beta_{kj} X_{ij} + \varepsilon_{ij}$$

Level-2 (course-section level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Treatment_j + \sum_{m=2}^M \gamma_{0m} Block_j + \mu_j$$

Where:

- $Y_{ij}$  = outcome score for student  $i$  in course section  $j$
- $\beta_{0j}$  = covariate-adjusted outcome score in control course section  $j$
- $\beta_{kj}$  = vector of parameters for the effects of student-level covariates, including baseline measures of SES, achievement, and interest in STEM
- $Treatment_j$  = treatment status for course section  $j$
- $\gamma_{00}$  = average covariate-adjusted outcome for control course sections
- $\gamma_{01}$  = estimated treatment impact
- $\gamma_{0m}$  = vector of parameters for the effects of randomization blocks
- $Block_j$  = block status for course section  $j$ ; blocks indicate teacher x cohort combination
- $X_{ij}$  = covariates for characteristics and baseline performance of student  $i$  in course section  $j$
- $\varepsilon_{ij}$  = error term for student  $i$  in course section  $j$
- $\mu_j$  = error term for course section  $j$

The parameter estimate,  $\gamma_{01}$ , provides a covariate-adjusted estimate of the impact of *FACT*. The hypothesis test for  $\gamma_{01}$  will determine whether or not the intervention has a statistically significant impact on the given outcome. A standardized effect size will be calculated by dividing the impact estimate ( $\gamma_{01}$ ) by the pooled standard deviation derived from the unadjusted sample standard deviations for the outcome in the intervention and comparison groups.

We will estimate this HLM model for all five student outcomes – for those on a binary scale as well as those on a continuous scale. For both binary and continuous outcomes, the linear model yields unbiased estimates of the intervention impact.

### TIP!

- ☑ Include terms in the analysis model to represent randomization blocks to improve the precision of the impact estimate.
- ☑ Blocking terms are necessary to properly account for unequal assignment probabilities (i.e., if they differ for different blocks, which is not the case in this example) and adjust for unequal rates of attrition across blocks. The WWC requires RCTs to account for unequal assignment probabilities in order to meet evidence standards.

In this example, randomization blocks correspond to each unique teacher and cohort combination. With 8-10 teachers and 3 cohorts, there will be a total of 24-30 randomization blocks.

### KEEP IN MIND...

Evaluators may choose to analyze impacts on binary outcomes using multilevel logistic regression or multilevel linear regression. Either approach will yield unbiased estimates of the intervention impact. Logistic regression models are designed for binary data. However, linear models may be simpler to estimate and interpret, and they yield standard error estimates that are approximately correct even when the underlying data generating process is nonlinear (Judkins & Porter, 2015).



## EXAMPLE EVALUATION PLAN

The contribution of covariates for student characteristics and baseline performance will be assessed for inclusion in the model. Covariates include: gender, race, ethnicity, first-time student status, interest in STEM, Pell grant eligibility, and a baseline measure of academic achievement (standardized score generated from either placement test scores or SAT score).

### TIP!

- ☑ Develop criteria for which covariates to include/exclude from the analysis model. The WWC rating of the study will not be affected by the approach used to include/exclude covariates – as long as you are careful not to include any covariates that could have been affected by the intervention.
- ☑ Use literature in the field to guide the selection of covariates. There may be covariates that should be included based on theory or prior empirical research, leading you to include certain covariates regardless of p-value or any other criteria.
- ☑ Consider backward selection, or another empirically-based approach, if you do not have a substantive basis for selecting covariates.

If the coefficient term for a covariate has a p-value less than  $p = 0.20$ , we will consider that covariate to be contributing to the precision of the impact estimate, and will include it in the model. Research has demonstrated that this approach is effective for identifying covariates to retain and those to drop, in order to minimize the standard error on the impact estimate (Budtz-Jorgensen et al, 2001; Maldonado & Greenland, 1993; Price et al, 2007). The block dummy variables will be included in the model regardless of coefficient significance, in order to account for the assignment of class sections within teachers and cohorts.

### Handling Missing Data

Students with missing outcome data will be excluded. However, we will handle missing covariate data using dummy variable adjustments, which is an effective method for handling missing data in RCTs (Puma, Olsen, Bell, & Price, 2009). All missing values will be imputed with a constant value (e.g., 0), and a dummy variable will be created for each covariate to indicate whether a missing value was imputed for the observation or not. In the event that attrition is high, missing data for variables used to establish baseline equivalence (see Section 3.4.2 and section 3.7) will not be imputed.



# EXAMPLE EVALUATION PLAN

## KEEP IN MIND...

The WWC only considers imputation of missing data acceptable in RCTs with low attrition that use one of the following approved methods (for handling missing outcome and baseline data):

- Complete case analysis
- Maximum likelihood methods
- Multiple imputation
- Non-response weights

The WWC does not prescribe acceptable methods for imputing missing data for covariates. For additional information on imputation of missing data, see:

- WWC Procedures and Standards Handbook (version 3.0), p. 18-19.
- Puma, Olsen, Bell, & Price (2009). What to do when data are missing in group randomized controlled trials (NCEE 2009-0049).
- Deke & Puma (2013). Coping with missing data in randomized controlled trials. Evaluation Technical Assistance Brief for OAH & ACYF Teenage Pregnancy Prevention Grantees.

## Adjusting for Multiple Comparisons

Two of the outcome domains of interest in the study – math achievement and credit accumulation – will be tested using multiple outcome measures. Math achievement is measured by the applied math course grade and the MACI. Similarly, credit accumulation is measured by the total number of STEM credits and STEM major. Within each outcome domain, we will apply Benjamini-Hochberg adjustments to any statistically significant findings. This approach is consistent with WWC practice, and is intended to account for inflated chance of a Type I error (i.e., finding a statistically significant effect in the sample when one does not exist in the population).

### TIP!

- Adjust for multiple comparisons to lower the chance of a false positive finding.

## KEEP IN MIND...

The WWC will determine whether or not multiple comparisons adjustments are necessary, and they will independently calculate any such corrections. For more information, see [WWC Procedures and Standards Handbook \(version 3.0\)](#), p. 25-26 and Appendix G.

For most outcomes under the WWC postsecondary education review protocol, the longest follow-up period available for a variable will be selected as the primary outcome. For the access and enrollment domain, the WWC privileges the earliest time point. For more information, see [WWC Postsecondary Education Review Protocol v3.1](#), p. 5.

Following the WWC Postsecondary Education review protocol, which indicates that the longest follow-up period should be treated as primary, we will apply the Benjamini-Hochberg correction for the impacts on the longest follow-up period for outcomes in the same domain: (1) at the end of the first semester Applied Mathematics course for the math achievement domain, and (2) at the end of students' 4<sup>th</sup> semester in community college for the credit accumulation domain. We will not adjust for tests of impacts in earlier semesters.



## Cross-Overs/No-Shows

The study will use an intent-to-treat approach to analysis. All course sections and students will be analyzed in their assigned condition, regardless of whether the teacher provides instruction consistent with the assigned condition and regardless of whether the student remains in the assigned condition. Students who drop out of their initial Applied Mathematics course and/or enroll in a different Applied Mathematics section (of the same or opposite instructional condition) will be analyzed in their assigned condition, as long as outcome data are available for the students.

### 3.6 Attrition (RCTs only)

We will examine attrition of course sections and students. In this section, we describe potential sources of attrition at each level, as well as how attrition rates will be calculated.

**Cluster-level attrition.** We expect little to no attrition of course sections. The only reasons that entire sections would be lost from the study would be if the course teacher unexpectedly leaves the school during the semester or the teacher is unable to provide outcome data from one or more sections. At SETU, teacher turnover from semester to semester is low, and essentially non-existent, within semesters. Further, all teachers who participate will be given clear guidance about data collection procedures. Therefore, we do not expect either issue to affect attrition of course sections.

In cluster RCTs, overall and differential attrition should be calculated at the cluster and the student levels for each outcome separately. Attrition at the cluster (course section) level will be calculated as the proportion of course sections that were randomly assigned to the treatment condition or to the control condition (i.e., the baseline sample) that do not contribute data to the analytic sample for an outcome. We will provide information on the numbers of treatment and control sections randomized and the number with data for each outcome. This sample size information will enable overall and differential attrition of course sections to be adequately assessed.

### TIP!

- ☑ Conduct an intent-to-treat analysis. Keep *no-shows* (students randomly assigned to the treatment group who fail to participate in the intervention) and *cross-overs* (students assigned to the control group who participate in the intervention) in the analysis sample in their originally assigned condition.

If cross-overs and no-shows are excluded, the WWC may view the study as having comprised random assignment.

- ☑ If interested in analyzing the treatment-on-the-treated, include the planned TOT analysis as supplementary to, and distinct from, the intent-to-treat analysis.

### TIP!

- ☑ Include the information necessary for the WWC to calculate overall and differential attrition for each analytic sample, when you report study findings.

For a cluster RCT...

- ☑ Report the total number of clusters randomized to each condition, and report the total number of clusters in each condition for which outcome data are non-missing.
- ☑ For clusters that have not been lost through attrition, report the total number of students at the time of random assignment in both conditions, and report the total number of students for which outcome data are non-missing.
- ☑ Remember that sample sizes can vary from one contrast to another.



## EXAMPLE EVALUATION PLAN

***Student-level attrition.*** At the student level, for the outcomes obtained at the end of the Applied Math course, (i.e., course grade of C or better; MACI test scores, STEM interest survey), students will be lost from the analytic sample only if they exit the treatment or control course during the semester. We expect that most student-level attrition from the Applied Mathematics course will occur in the first two weeks of the semester, when mobility is high. Because this early mobility is related to students shifting their schedules, we anticipate that attrition will be relatively even across treatment and comparison groups. Given this anticipated low differential attrition, even with a fair amount of overall attrition, we expect the attrition of students to be low when compared to the WWC liberal attrition standards.

On the STEM Interest Survey, students may also have missing data on the outcome measurement because they decline to complete the survey. Students will be offered repeated opportunities to complete the survey in an effort to minimize non-response.

Leaving the school is another potential source of student attrition for all of the outcomes. Students will have missing outcome data if they withdraw from SETU or transfer to another college during the semester in which the outcome is measured (with the exception of the STEM major outcome, which will not be missing for transfer students if they declared a major prior to transfer). For outcomes measured at multiple time points (i.e., cumulative number of credits earned in STEM, measured in semesters 2, 3, and 4), data will also be missing for any subsequent semester after students leave the school.

For each analytic sample (i.e., for each outcome measure at each measurement time point), we will assess the overall and differential attrition of students for non-attrited course sections. We will report the number of students who were enrolled at the time of random assignment in non-attrited treatment course sections and in non-attrited control course sections (i.e., the baseline student sample). Also, we will report the number of treatment students from the baseline sample (i.e., in non-attrited sections) who have outcome data and the number of control students from the baseline sample (i.e., in non-attrited sections) who have outcome data (i.e., the analytic student sample sizes).

### ***3.7 Baseline Equivalence Testing (QEDs and RCTs with High Attrition)***

Because this study uses random assignment, we assume that any differences between treatment and control group students on observable and unobservable variables that exist at baseline occur purely by chance. However, in the event of high attrition, the balance between the treatment and control groups resulting from random assignment may no longer hold. Therefore, we will assess the equivalence of the treatment and control students at baseline for each analytic sample. If attrition is high, the analytic sample will be defined as students without a missing outcome and without missing data for baseline measures of math achievement and SES (or without a missing baseline score on the SISI for analyses of impacts on interest in STEM). Analytic samples for each outcome may vary slightly, given differences in missing data; therefore, baseline equivalence will be assessed for each analytic sample. The study findings will report the mean and standard deviation of each baseline measure, along with the sample size for each group at baseline.



## EXAMPLE EVALUATION PLAN

If there is high attrition, we will test the baseline equivalence of the analytic sample using a regression model reflecting the structural features of the design (i.e., the blocking used in random assignment). Specifically, we will use a modified version of the model described previously for testing intervention impacts. However, we will move the baseline measure to the left-hand side of the model, retain the treatment indicator and blocking variables on the right-hand side, and omit all other covariates. The parameter estimate for the treatment variable ( $\gamma_{01}$ ) will provide an estimate of the magnitude of the baseline mean difference between the treatment and comparison students in the scale of the baseline measure.

Level-1 (student-level):

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

Level-2 (course-section level):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Treatment_j + \sum_{m=2}^M \gamma_{0m}Block_j + \mu_j$$

Where:

$Y_{ij}$	= baseline score for student $i$ in course section $j$
$\beta_{0j}$	= baseline score in control course section $j$
$Treatment_j$	= treatment status for course section $j$
$\gamma_{00}$	= average baseline score in control courses
$\gamma_{01}$	= baseline score difference between treatment course sections and control course sections
$\gamma_{0m}$	= vector of parameters for the effects of randomization blocks
$Block_j$	= block status for course section $j$ ; blocks indicate teacher x cohort combination
$\varepsilon_{ij}$	= error term for student $i$ in course section $j$
$\mu_j$	= error term for course section $j$

For continuously-scaled measures (e.g., baseline math achievement), we will calculate the standardized baseline difference (Hedges'  $g$ ) by dividing the parameter estimate (i.e.,  $\gamma_{01}$ ) by the pooled standard deviation derived from the unadjusted sample standard deviations for the intervention and comparison groups.

### TIP!

- ☑ Be prepared to assess baseline equivalence for the analytic sample (or samples) if there is high attrition or if there are students who join clusters after randomization.
- ☑ Do not include any student who is missing the outcome measure in tests of baseline equivalence. In cluster RCTs with joiners and in RCTs with high attrition, the WWC requires that baseline equivalence be assessed for the sample of students that have both non-missing baseline data and non-missing outcome data.
- ☑ Assess baseline equivalence for each analytic sample with high attrition. Remember that the analytic sample may differ from one contrast to another, depending on what data are missing.
- ☑ Be aware that the analytic sample may differ from the sample at the time of random assignment. You may wish to compare the baseline characteristics of the treatment and control groups at the time of random assignment. Note, however, that this comparison is not a proper test of baseline equivalence in the analytic sample.

### KEEP IN MIND...

In this example, baseline equivalence is assessed using a statistical model, accounting for the structural features of the design (i.e., randomization blocks). The WWC will also accept a comparison of unadjusted baseline sample means for the intervention and comparison group to establish baseline equivalence.



## EXAMPLE EVALUATION PLAN

For binary measures (e.g., Pell grant eligibility status), we will report the percentage of students in the control group who are eligible for Pell grants at baseline (i.e., in their first year of enrollment). Using the same modified model described above for estimating the magnitude of the baseline difference, we will calculate and report the model-adjusted percentage of students in the treatment group who are Pell grant eligible. Both of these percentages, as well as the number of students in each condition, can be used to calculate a Cox index (an effect size for binary measures) instead of Hedges'  $g$ .

$$\text{Cox Index} = \left[ \ln\left(\frac{p_t}{1-p_t}\right) - \ln\left(\frac{p_c}{1-p_c}\right) \right] / 1.65$$

Where,  $p_t$  is the probability that a student in the treatment group is eligible for a Pell grant, and  $p_c$  is the probability that a student in the control group is eligible for a Pell grant.

The treatment and control students will be considered to be equivalent on a given measure if the baseline difference is  $\leq 0.25$ , because we will control for the baseline measure in the impact analysis model (regardless of the p-value for the coefficient).

Also, unadjusted control group means, adjusted treatment group means, and standard deviations at baseline will be reported for the following variables:

- SISI
- Placement test score
- SAT score from high school
- Gender
- Race/ethnicity
- Pell grant eligibility

### TIP!

- ☑ For cluster designs using cluster means to establish baseline equivalence, be sure to provide the following information:
  - Number of individuals enrolled in the cluster at (or around) the time of baseline measurement
  - Number of individuals with values on the baseline measure that contribute to the cluster mean at baseline.

The WWC will use this information to assess whether the number of students used to calculate the cluster mean is large enough to be *representative* of the cluster.



## 4. References

Budtz-Jorgensen E, Keiding N, Grandjean EM, Weihe P (2007). *Confounder selection in environmental epidemiology*. Assessment of health effects of prenatal mercury exposure. *Annals of Epidemiology*; 17:27-35.

Deke, J. & Puma, M. (2013). Coping with missing data in randomized controlled trials. Evaluation Technical Assistance Brief for OAH & ACYF Teenage Pregnancy Prevention Grantees.

Johnson, Z. (2011). *The STEM Interest Survey Instrument (SISI)*. Note that this is a fictional reference for fictional instrument used for illustrative purposes only.

Judkins, D. R., & Porter, K. E. (2015). *Robustness of ordinary least squares in randomized clinical trials*. *Statistics in Medicine*. Advance online publication. doi:10.1002/sim.6839

Maldonado, G., Greenland, S., (1993). *Simulation study of confounder-selection strategies*. *American Journal of Epidemiology* 138(11); 923-936.

Price, C., Goodson, B., Stewart G. (2007). *Infant Environmental Exposure to Thimerosal and Neuropsychological Outcomes at Ages 7 to 10 Years Technical Report Volume II*. Bethesda: Abt Associates, Inc.; Prepared for Immunization Safety Office Centers for Disease Control and Prevention.

Puma, M., Olsen, R., Bell, S., & Price, C. (2009). What to do when data are missing in group randomized controlled trials (NCEE 2009-0049).

Smith, Y. & Jones, X. (2009). *Math Applications and Concepts Inventory (MACI)*. Note that this is a fictional reference for fictional instrument used for illustrative purposes only.

What Works Clearinghouse (March, 2014). *WWC Procedures and Standards Handbook, Version 3.0*. [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf)

What Works Clearinghouse (March, 2016). *WWC Cluster Design Standards*. [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_cluster\\_standards\\_030416.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_cluster_standards_030416.pdf)

What Works Clearinghouse (March, 2016). *WWC Reviewer Guidance for Use with the Procedures and Standards Handbook, Version 3.0*. [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_reviewer\\_guidance\\_030416.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_reviewer_guidance_030416.pdf)

What Works Clearinghouse (August, 2015). *WWC Review Protocol for Individual Studies in the Postsecondary Education Topic Area, Version 3.1*. [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_pe\\_protocol\\_v3.1.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_pe_protocol_v3.1.pdf)