



Example Evaluation Plan for a Student-Level Randomized Controlled Trial

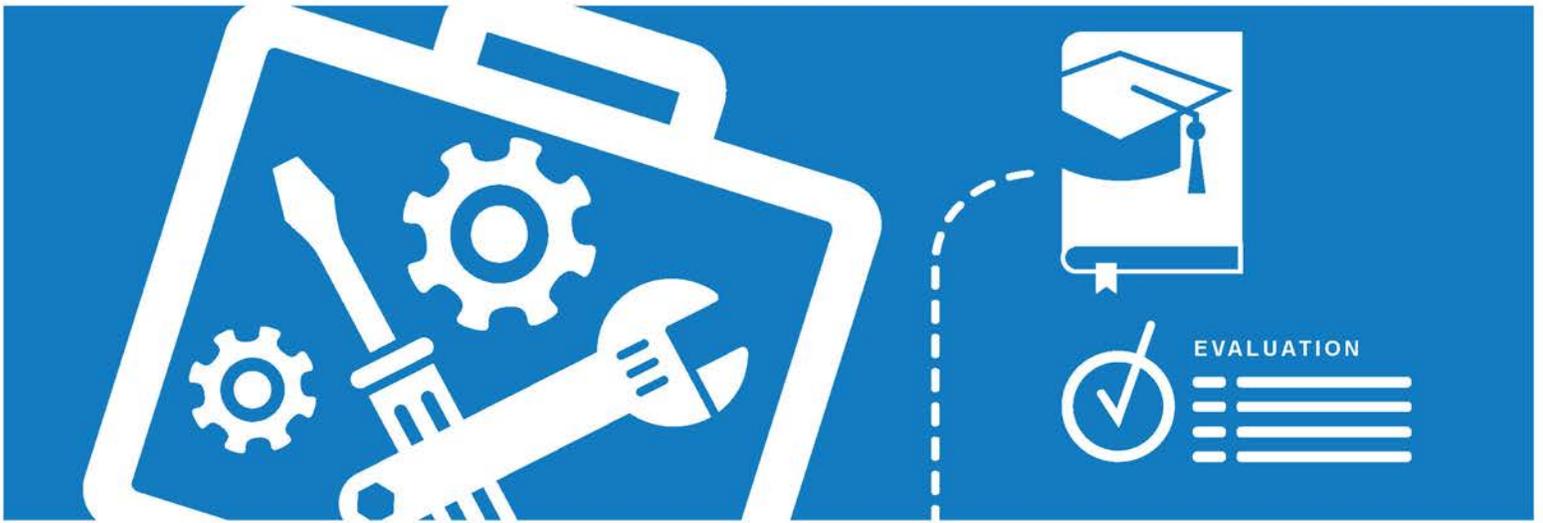
The Evaluation Plan Template identifies the key components of an evaluation plan and provides guidance about the information typically included in each section of a plan for evaluating both the effectiveness and implementation of an intervention. Evaluators can use this tool to help develop their plan for a rigorous evaluation, with a focus on meeting *What Works Clearinghouse*TM 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 students 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 student-level RCT. Accompanying this example evaluation plan is the Example Contrast Tool for a Student-Level 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
Carter Epstein
Barbara Goodson
Cristofer Price
Beth Boulay



August 2016



The Institute of Education Sciences (IES) 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*TM evidence standards.





Grantee: Springdale University

1. Evaluator Information	1
1.1 Contact Information	1
1.2 Independence	1
1.3 Confidentiality Protections	1
2. Summary of Intervention	2
3. Impact/Effectiveness Evaluation	3
3.1 Research Questions	5
3.2 Comparison Condition	7
3.3 Study Sample and How Intervention and Comparison Groups are Selected/Assigned	8
Student Eligibility.....	8
Random Assignment Procedures.....	9
Expected Sample Sizes.....	11
Progression of Students and Cohorts across Multiple Years.....	11
3.4 Key Measures and Plan for Obtaining Data	13
Data Collection.....	13
Analytic Measures	14
3.5 Statistical Analysis of Impacts	17
Impact Analysis Model.....	18
Treatment of Missing Data.....	19
Adjusting for Multiple Comparisons.....	20
Cross-Overs/No-Shows	20
3.6 Attrition	21
3.7 Baseline Equivalence Testing	22
4. References	23



EXAMPLE EVALUATION PLAN

1. Evaluator Information

1.1 Contact Information

Evaluator: Outlook Consulting, LLC
100 Main Street
Anytown, PA 01234

Lead Investigator: Rachel Dorris
rdorris@outlookconsulting.com
(888) 333-1111

1.2 Independence

The lead investigator, Rachel Dorris of Outlook Consulting, LLC, is an independent consultant and is not affiliated with Springdale University; played no role in the development or implementation of Springdale University's "Coach4College" initiative, and will conduct all aspects of the evaluation described in this document. This includes conducting random assignment, the collection of administrative and non-administrative outcome data (Outlook Consulting will hire and train proctors for the administration of the Sense-of-Belonging Scales, SOBS), analysis and preparing written and oral reports of study findings. The findings will not be subject to approval from Springdale University, but will be prepared, reported, and disseminated by Outlook Consulting.

1.3 Confidentiality Protections

IRB approvals have been obtained for implementation and impact evaluation. Chester IRB Services is serving as the IRB. Outlook Consulting and Springdale University will comply with all Federal and Pennsylvania privacy laws including FERPA. For all individual level data, the university will assign an identification number for the study that is unique to each student, and will maintain a crosswalk of these identification codes and personally identifiable information (PII) on university servers. The evaluation team will use this identification number instead of student names to exchange data with the university. In addition, the evaluation team and university will exchange data via a secure file sharing portal (SFTP) and not via email or other non-secured transmission methods. Although Outlook Consulting will have only de-identified data on individual students, Outlook Consulting will treat the data as sensitive and confidential. For administration of the Sense of Belonging Scale (SOBS), Outlook Consulting will send each student a unique link to an online version of the scale. There are numerous safeguards in place to protect student course enrollment, course grade, credits earned, and test data. These protections include de-identifying the data, password-protected computers, locked files, etc. These detailed procedures for each type of interview, observation, and level of school personnel involved are outlined carefully in our IRB application.



2. Summary of Intervention

This study will investigate the impact of a *Coach4College* peer support community provided across the first year of college to first-time, students at Springdale College who are at risk of failing to complete a bachelor's degree within four years. The main goal of the program is to increase the rate of successful completion of college among students who enter college with barriers to on-time progress and graduation. These barriers include: unfamiliarity with the milestones needed to make on-time progress toward college graduation; lack of exposure to strategies for time management, planning and organization; and feeling a sense of isolation or lack of belonging to the campus community. This last factor can be particularly pronounced among students who are members of groups that have historically lower rates of college completion, such as African-American students, Hispanic or Latino students, first-generation-to-college students, and students from low-income households. To achieve these goals, the *Coach4College* "College Squad" program provides a suite of resources and services to incoming freshman over the course of their first year of college enrollment. The emphasis on earning a bachelor's degree within four years is not only to ensure that students attain the degree, but also accumulate the least loan debt.

Coach4College is designed based on research showing that a cohort model, in which a group of students share a common set of activities and experiences, can improve student retention by enhancing their sense of belonging. Academic advisors should not only provide accurate and timely information about academic requirements, procedures, and resources; they should also engage students in a collaborative process that acknowledges student strengths and desire for growth and helps students create educational plans that fulfill personal goals. Moreover, especially for students at-risk of college dropout, proactive advising, in which the advisor takes the initiative to reach out to students and convey that they care about the student's success, is often a more successful way to connect students with advisors or mentors than the traditional passive advising model, in which the responsibility for seeking help falls to students. Finally, research shows that how students respond to disappointment or frustration, particularly with regard to academic performance, can affect whether they persist or gradually withdraw from college engagement. A mindset that emphasizes belief in one's ability to improve (i.e. change is possible), and the importance of hard work and practice (as opposed to innate talent) in academic improvement, is associated with a more proactive response to setbacks and yields improved performance on tests and assignments.

The intervention will target incoming first-time, first-year students who are African American, Hispanic/Latino, first-generation to college students, or have one or more risks for not completing college (i.e. Pell grant eligible; SAT reading or math scores < 500; high school GPA \leq 3.0; admissions essay score \leq 3 out of 5 points). These criteria are based on interviews with students, faculty, and student affairs staff, a review of retention and graduation rates of various subgroups of students over the past 10 years, and a review of recent literature on postsecondary trends in college completion. Together, these data sources suggest that the primary causes of student drop-out at Springdale are (a) failure to complete graduation requirements in a timely manner, and (b) the perception among particular groups of students that they lacked some quality or qualities needed to successfully complete college.



EXAMPLE EVALUATION PLAN

The students offered *Coach4College* will be assigned to a “College Squad” of approximately 15-18 students, and will have the opportunity to participate in cohort-based common courses, a biweekly seminar on strategies for thriving in college, and individualized mentoring. The specific features of the intervention that will be offered to students are listed below.

Intervention Services: *Coach4College* College Squad

- Assignment of students into a “squad” with common enrollment across two of the college’s four required core courses;
- A designated student affairs “squad leader” assigned to work with a group of 15 to 18 students, who follows them as they progress through their first year of college;
- A telephone call from the squad leader one week prior to student arrival on campus;
- Invitation to 10-day College Squad “kickoff” summer program immediately before the 4-day New Student Orientation required of all new students prior to the start of the Fall semester;
- Biweekly small group “Thriving in College” seminar facilitated by the squad leader focusing on the development of skills related to college persistence, such as time management, stress management, coping with challenges/frustration (60 minutes per week during 8 of the 15 weeks of the first semester);
- Biweekly one-on-one meetings with the squad leader (30 minutes, during weeks of the first semester when the Thriving in College seminar is not meeting); and
- Three squad leader check-ins with each student’s core course instructors throughout each semester to monitor student attendance, assignment completion, and quality of the student’s work vis-à-vis expectations for college level academic performance.

3. Impact/Effectiveness Evaluation

The impact study of *Coach4College* will employ a randomized controlled trial (RCT) design, with assignment at the student level. Eligible students entering the university as first-time, first-year students will be grouped based on risks for not completing college and randomly assigned within group to either the control condition (access to the existing academic advising and campus support services, available to all enrolled students) or to the intervention condition (the *Coach4College* College Squad services).

The *Coach4College* intervention services span August through May of students’ first year of college. The evaluation will examine immediate effects on student outcomes during their first year and longer-term effects on student achievement of milestones necessary for graduation, academic performance, and sense-of-belonging in later years.

This intervention will be implemented and evaluated at Springdale University, a mid-sized private, nonprofit four-year college serving approximately 5,100 full-time enrolled undergraduates per year. The average size of the incoming freshman class between 2013 and 2015 was 1,850.



EXAMPLE EVALUATION PLAN

Student demographics include:

- Approximately 55 percent students are eligible to receive Pell grants;
- Approximately 15 percent are first generation to college students;
- 11 percent of students identify as African-American; and
- 13 percent identify as Hispanic.

Springdale has a set of required milestones that all students must reach in order to graduate.

Students must complete:

- Four core courses (required of all students to graduate with a Bachelor's degree, equivalent to 12 credits);
- A set of distributional courses (a minimum of two courses in each of the college's three divisions: humanities and art; social sciences; science and engineering, for a total of 18 credits); and
- Major discipline courses (a minimum of 8 courses in a major field of study, for a total of 24 credits).

In addition, to graduate, all Springdale University students must complete an Interdisciplinary Mentored Project (IMP), designed in consultation with a committee of at least two faculty members, each from a different academic department. Most students spend their junior year developing a formal IMP proposal, which must be approved by their faculty committee before the project work begins. Springdale University considers students who have received approval of their IMP proposal by the end of their junior year to be making on-time progress toward graduation. Students who receive approval later than the end of their junior year often face difficulty completing their project by the end of their senior year.

KEEP IN MIND...

In most evaluation plans, the participating school(s), district(s), or college(s) would be described in the section on the study sample. However, in this example, the intervention is being implemented in a *single* university, and some of the evaluation outcomes focus on completion of university-specific milestones. Therefore, a description of the university context appears before the research questions to provide context for understanding the outcome measures.



3.1 Research Questions

The evaluation will address the research questions 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 14 such tests—even though there are only 7 different outcome measures, the effect of the intervention is tested on some outcome measures at multiple points in time.

The first four questions examine effects on *total credits earned*:

- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad at the end of the first semester of college on total credits earned relative to the business-as-usual academic and campus support services available college wide?
- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad at the end of the second semester of college on total credits earned relative to the business-as-usual academic and campus support services available college wide?
- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad at the end of the fourth semester of college on total credits earned relative to the business-as-usual academic and campus support services available college wide?
- For one cohort of entering college freshman (entering fall 2016), what is the effect of Coach4College College Squad at the end of the sixth semester of college on total credits earned relative to the business-as-usual academic and campus support services available college wide?

One question examines the effect on *completion of four core courses*:

- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad, at the end of the second semester of college, on successful completion of all four required core courses relative to the business-as-usual academic and campus support services available college wide?

One question examines the effect on *persistence in college*:

- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad, at the end of the fourth semester of college, on persistence in college relative to the business-as-usual academic and campus support services available college wide?



EXAMPLE EVALUATION PLAN

One question examines the effect on *declaration of a college major*:

- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad, at the end of the fourth semester of college, on on-time declaration of a college major relative to the business-as-usual academic and campus support services available college wide?

One question examines the effect on approval of the Interdisciplinary Mentored Project:

- For one cohort of entering college freshman (entering fall 2016), what is the effect of Coach4College College Squad, at the end of the sixth semester of college, on on-time approval for the Interdisciplinary Mentored Project relative to the business-as-usual academic and campus support services available college wide?

Four questions examine effects on *cumulative GPA*:

- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad, at the end of the first semester of college on cumulative GPA relative to the business-as-usual academic and campus support services available college wide?
- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad, at the end of the second semester of college on cumulative GPA relative to the business-as-usual academic and campus support services available college wide?
- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad, at the end of the fourth semester of college on cumulative GPA relative to the business-as-usual academic and campus support services available college wide?
- For one cohort of entering college freshman (entering fall 2016), what is the effect of Coach4College College Squad, at the end of the sixth semester of college on cumulative GPA relative to the business-as-usual academic and campus support services available college wide?

Two questions examine effects on students' *sense of belonging*:

- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad, at the end of the first semester of college, on sense-of-belonging relative to the business-as-usual academic and campus support services available college wide?
- For two cohorts of entering college freshman (entering fall 2016 and fall 2017), what is the effect of Coach4College College Squad, at the end of the second semester of college, on sense-of-belonging relative to the business-as-usual academic and campus support services available college wide?



EXAMPLE EVALUATION PLAN

3.2 Comparison Condition

During their first year at Springdale, students in the control group will be able to participate in or receive the usual academic advising and support services available campus-wide. Springdale University has a robust academic advising program, as well as offices and staff dedicated to supporting undergraduate student health and wellbeing in and outside of academic matters. The college assigns each first-year student a faculty academic advisor who is responsible for helping the student plan their academic schedule to balance completion of requirements with a reasonable workload per semester. In addition, first-year students (as well as students transferring to Springdale University) are required to attend a four-day New Student Orientation that occurs in August on the four days prior to the start of classes.

During this orientation, students learn about college policies, academic requirements and the advising program, campus activities and clubs, and meet representatives from each of the college's key student offices including: Admissions and Financial Aid Services; Academic Affairs; the Writing Center (for one-on-one and small group assistance with research papers, compositions, and other writing assignments); Mathematics Lab (for one-on-one and small group assistance with quantitative course assignments); Health and Wellness Center; Fitness and Recreational Services; the Academic Computer and Information Technology Center; Multicultural Student Affairs Office and the Gay, Lesbian, Bisexual or Transgender Student Center (GLBT Center). Specifically, the **business-as-usual** academic advising and campus support services include:

- An assigned faculty “first-year advisor” for all first-year students with whom they are required to meet twice per year (once at the start of each semester) and more often if the student chooses;
- Attending an on-campus New Student Orientation in the four days prior to the start of classes for all incoming students (first-year and transfer students);
- Several on-campus services designed to support students throughout their enrollment at the college (students may self-refer or faculty members may suggest a visit):
 - The Student Academic Support Center, which provides individual and group support for students needing help with coursework, as well as referral to tutoring services;
 - Campus Health Center, including support for stress reduction, mental health counseling and referral services, nutrition, exercise and wellness programming;
 - A range of other campus support centers, including the Campus Writing Center, Math Laboratory, Financial Aid Office, and Career Readiness Center.
- Students may attend and/or participate in cultural and social programming, as well as seek guidance from designated resident advisors, from the Multicultural Student Affairs Office and/or the GLBT Center.



EXAMPLE EVALUATION PLAN

Exhibit 1: Contrast between Coach4College (intervention group) and Business-as-usual (control group) services

	Coach4College College Squad (Intervention Group)	Business-as-Usual First Year Student Experience (Control Group)
Nurturing a sense-of-belonging	10-day "College Kickoff" program	
	New Student Orientation	New Student Orientation
	Common registration to 2 of 4 core courses in first semester	Standard course registration (students register for core classes independently)
	Bi-weekly "Thriving in College" seminar with squad (first semester)	
	Access to campus student groups, activities	Access to campus student groups, activities
Connecting students to advising	Summer telephone call from squad leader	
	Bi-weekly one-on-one meetings with squad leader (first semester)	Traditional faculty academic advising
	Squad leader monitoring of student academic performance 3x per semester in first year	
	Squad-leader (or faculty) referral (or self-referral) to academic/health support services	As-needed faculty referral (or student self-referral) to academic/health support services
Building students' resilience	10-day "College Kickoff" program	
	Bi-weekly "Thriving in College" seminar with squad (first semester)	
Responding promptly to early signals of difficulty	Squad leader meets with student and/or refers student to support services if monitoring of student academic performance suggests student is at-risk	No formal monitoring; faculty may refer student to support services on own initiative or student may self-refer on own initiative

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

Information about study samples is described in this section and also shown in the accompanying contrast tool, on the "samples" tab.

Student Eligibility

Students who are eligible for the intervention include all incoming first-time, first-year students to Springdale College in Fall of 2016 or Fall of 2017 in one of two groups:

- *Group 1:* African-American, Hispanic/Latino, and/or first-generation to college students; OR
- *Group 2:* Students who are not in Group 1 but who are identified by Admissions or Academic Support staff as having one or more risks for not completing college, based on the following criteria:
 - Pell grant eligible; and
 - SAT Critical Reading and/or SAT Math score < 500 and/or high school GPA of 3.0 or lower and/or college admissions essay received an average score of 3 or lower (highest possible score=5).



EXAMPLE EVALUATION PLAN

Based on administrative data from the prior three academic years, approximately 13% of students fall into Group 1, and an additional 20% of students fall into Group 2, for a total of approximately 33% of first-time, first year students who are eligible for the *Coach4College* intervention and for the evaluation. Students in either Group 1 or Group 2 are eligible for the intervention.

TIP!

- ☑ Plan to obtain consent before conducting random assignment.
- ☑ Clearly specify the timing of consent relative to random assignment.

If consent is obtained after random assignment, students who do not consent will count toward attrition and may contribute to high overall attrition.

In addition, any student who meets the criteria for Group 1 or Group 2, but chooses to opt-out of the evaluation, will be excluded from randomization and will not have the opportunity to participate in the intervention. That is, prior to random assignment, all incoming first year students (and parents of students under the age of 18 by August 1st of the year they begin college) will receive an information letter about the intervention, random assignment, and data collection for the evaluation.

Students (and parents of minor students) will have the opportunity to opt-out of possible participation in the College Squad intervention, random assignment, and data collection. These letters will be included with student Welcome Packages, which are mailed to all first-year students upon receipt of their enrollment deposit (i.e., signaling their intent to attend Springdale University in the fall). Enrollment deposits are due July 1 each year. The letter will emphasize that eligible students will be randomly assigned to receive Springdale's existing first year support services or the College Squad intervention. Students who do not wish to participate in the College Squad intervention (i.e., if they were randomly assigned to that condition), may opt out of the study by returning the opt-out form by August 15th. Students who have opted-out of random assignment by August 15th will not be eligible for the intervention, but can still receive the college's business-as-usual first-year advising and support services.

Random Assignment Procedures

Once student rosters are final and all opt-out forms have been processed (expected by August 15th each year), the evaluator will identify the students in Groups 1 and 2 (defined in the previous section), who will be eligible for the intervention and the study. The evaluator will randomly assign all eligible students within each block (i.e., where Group 1 is one block and Group 2 is another), using different proportions: in Group 1, 70 percent will be assigned to the intervention and 30 percent to the control group; in Group 2, 50 percent will be assigned to the intervention and control groups.

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. In this example, students who do not agree to randomization and data collection will not be offered the opportunity to participate in the intervention (and will be excluded from the impact study). In some cases, it may be possible to allow students to opt-out of data collection (and, therefore, the evaluation), but still allow them to participate in the intervention via randomization. For example, students who do not consent to data collection could be randomized separately from consenters and excluded from the impact study.



EXAMPLE EVALUATION PLAN

Given that there will be two cohorts (cohorts A and B), study participants will be randomized to treatment or control conditions within one of the following four randomization blocks: “Cohort A, Group 1”, “Cohort A, Group 2”, “Cohort B, Group 1”, and “Cohort B, Group 2.”

Random assignment will occur the third week of August (in 2016 for Cohort A and 2017 for Cohort B). This will allow random assignment to be based on nearly final student rosters of incoming first-year students, while still allowing the university sufficient time to implement the squad based common course registration in two core courses for students assigned to the intervention condition.

Ineligible students, including first-time, first year students who do not fall into Groups 1 or 2, transfer students, and incoming first-time, first-year students who enroll after random assignment will not receive the intervention. Ineligible students will be offered the usual first-year student advising and support services and will not be included in the evaluation sample.

TIP!

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

Following the completion of random assignment, the list of students randomized to *Coach4College* and to business-as-usual services will be given to the study liaison in Springdale’s Office of the Registrar. University staff will notify intervention students to arrive for the 10-day college kickoff program and new student orientation. Control students, and any other first-year students that were not

randomized to either condition, will be notified to arrive for new student orientation, as is traditionally done at Springdale.

Careful checks will be made to ensure intervention delivery to the correct group of students. Students randomly assigned to the control condition may not participate in College Squad activities, and all squad leaders will be required to adhere to this restriction. Springdale’s study liaison will work with the evaluation team to confirm that only students assigned to the intervention group participate in the college kick-off program. Similarly, the university study liaison, together with the evaluation team, will examine rosters of students enrolled in the common core courses to confirm that only treatment students, and no control students or non-study students, are enrolled in common core courses. Any treatment students enrolled in other (non-common) core course sections and any control or non-study students enrolled in common core courses will be moved to an appropriate course assignment prior to the start of the semester.



EXAMPLE EVALUATION PLAN

Expected Sample Sizes

The target sample sizes for the study are based on the estimated number of study-eligible students in each of the 2016 and 2017 incoming first-year student classes. Exhibit 2 shows the anticipated sample sizes for each cohort and eligibility group.

Exhibit 2: Expected Sample Sizes by Cohort

Cohort	# of Entering First Year Students	Estimated # of Eligible Incoming First-Year Students	Number Assigned to Treatment	Number Assigned to Control
1: Fall 2016	1,850	600		
	Group 1	250	175	75
	Group 2	350	175	175
	Total, Cohort A	600	350	250
2: Fall 2017	1,850	600		
	Group 1	250	175	75
	Group 2	350	175	175
	Total, Cohort B	600	350	250
	TOTAL	1,200	700	500

The impact analyses will use the total sample (Cohorts A and B combined) when estimating intervention impacts. The expected sample size with both cohorts combined is approximately 1200 students—700 intervention students and 500 control students.

Progression of Students and Cohorts across Multiple Years

Exhibit 3 illustrates the semester-by-semester progression of the two study cohorts that will be included in the evaluation of *Coach4College* College Squad intervention. Although the intervention will last one year, the evaluation will measure the effects of the intervention on outcomes through students' third year of enrollment.

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 3) 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.



EXAMPLE EVALUATION PLAN

Exhibit 3. Semester-by-semester Progression of Two Cohorts of Students in the Sample

Academic Year	Semester	Cohort A	Cohort B
2015-16	Pre-intervention (Year 0)	A0 High School No exposure	
2016-17	Fall 2016 (Semester 1)	A1* 1st semester (Freshman Year) 1 semester exposure	B0 High School No exposure
	Spring 2017 (Semester 2)	A2* 2nd semester (Freshman Year) 2 semesters exposure	
2017-18	Fall 2017 (Semester 3)	A3 3rd semester (Sophomore Year)	B1* 1st semester (Freshman Year) 1 semester exposure
	Spring 2018 (Semester 4)	A4* 4th semester (Sophomore Year) 1 year post-intervention	B2* 2nd semester (Freshman Year) 2 semesters exposure
2018-19	Fall 2018 (Semester 5)	A5 5th semester (Junior Year)	B3 3rd semester (Sophomore Year)
	Spring 2019 (Semester 6)	A6* 6th Semester (Junior Year) 2 years post-intervention	B4* 4th semester (Sophomore Year) 1 year post-intervention
End of Data Collection			
2019-20	Fall 2019	A7 7th semester (Senior Year)	B5 5th semester (Junior Year)
	End of Grant		

Note: Cell labels indicate the cohort (A or B) and students' semester in college (1-6). Asterisks (*) indicate the timing of data collection for each cohort. Shading indicates the timing of the intervention for each cohort.

The two cohorts will be combined for analyses, and the effects of the intervention will be examined at the end of students' first semester, second semester, fourth semester, and sixth semester. As shown in Exhibit 3, the timing of outcome measurement will vary by cohort. Data collection for the evaluation will end in spring 2019, which corresponds to the end of three years of college for Cohort A and two years of college for Cohort B.

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.



EXAMPLE EVALUATION PLAN

The samples for the analyses of each outcome are as follows:

- Outcomes at the end of the first semester will draw from Fall 2016 for Cohort A and Fall 2017 for Cohort B (cells A1 and B1).
- Outcomes at the end of the second semester will draw from Spring 2016 for Cohort A and Spring 2017 for Cohort B (cells A2 and B2).
- Outcomes at the end of the fourth semester will draw from Spring 2018 for Cohort A and Spring 2019 for Cohort B (cells A4 and B4).
- Outcomes at the end of the sixth semester will draw from Cohort A only, Spring 2019 (cell A6).

TIP!

- ☑ 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.

In this example, the focus is on student “grade” (e.g., first semester freshman, end of freshman year, etc.), which also aligns with amount of exposure/length of follow-up. Samples are not based on the length of time the intervention has been in place. See the accompanying contrast tool for these dimensions for each contrast.

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

Data will be collected from three sources – administrative data from Springdale University, the National Student Clearinghouse, and student surveys.

1. **Administrative data.** Most measures will come from Springdale University administrative data. Springdale will transfer relevant data to the evaluation team for students in the study sample (i.e., any student ever randomly assigned, regardless of how many semesters of enrollment for each such student) using a secure file transfer protocol and study-specific ID numbers assigned to each student at the onset of the study. Administrative data include information routinely compiled by the university each semester – cumulative number of credits earned, completion status for four required core courses, enrollment status, cumulative grade point average, declaration of major, and approval status of IMP proposal. The university will transfer administrative data to the evaluation annually in summer 2017, summer 2018, and summer 2019.
2. **National Student Clearinghouse (NSC) data.** Data from the National Student Clearinghouse will be obtained on student college enrollment through spring 2018 (cohort A) and spring 2019 (cohort B). These data will supplement Springdale University data on enrollment status, and will be used to determine whether students who have left Springdale University enrolled in another college. In other words, these data will be used to distinguish students who transfer from those who drop out.



EXAMPLE EVALUATION PLAN

- 3. Student surveys.** All students in the study will complete a Sense of Belonging Scale (SOBS) three times: at the start of their first semester (a baseline measure); at the end of the first semester; and again at the end of the second semester. The first administration of the SOBS will be incorporated into a New Student Survey administered online during New Student Orientation. The SOBS items will be administered only to students in the study sample (in both the intervention and control groups), and not to other first-year students. In addition, at the end of their first semester, all first-year students in the study will be given a link to complete the SOBS scale online. Finally, the SOBS will be incorporated into two rounds of Springdale University's annual end-of-year Student Experience Survey: spring of 2017 (to capture second semester SOBS outcome for Cohort A) and again in the spring of 2018 (to capture second semester SOBS outcome for Cohort B). The end-of-year online survey will include SOBS items only for students in the study (i.e., students not enrolled in the study will not receive SOBS items).

Analytic Measures

Below we describe the outcome measures, baseline measures, and other independent variables that will be used in analyses of the impacts of *Coach4College*. 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 seven outcome measures – six of which will be obtained from university administrative data and one will be constructed from the Sense of Belonging Scale administered at the end of the first and second semesters to students in the evaluation sample. The timing of measurement – first semester, second semester, fourth semester, and sixth semester – corresponds to the timeline presented in Exhibit 3. The following six variables obtained from administrative data will be constructed:

- Total credits earned is the cumulative number of credits earned through a given semester of enrollment, and will be measured as of students' first, second, fourth, and sixth semester after entering Springdale. For students who leave the university, the outcome, will be missing for any semester subsequent to the last semester the student completed at Springdale.

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 data point available (i.e. the last semester completed).

- ☑ 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 Springdale University after his/her third semester may go on to earn credits toward completion of a bachelor's degree at another institution – without such data, his/her outcome for “total credits earned at the end of the fourth semester of college” would be missing.



EXAMPLE EVALUATION PLAN

- Completion of the four required core courses will be coded as “yes” for students who complete all four core courses by the end of their first year or “no” for students who do not. Students who leave Springdale University prior to the relevant semester will not be treated as missing, but will be coded based on data in the last semester of enrollment at Springdale. For example, a student who completed three of four core courses by the end of the first semester and then withdrew from Springdale will be coded as not having completed all four courses by the end of the second semester.
- Persistence in college for four semesters will be coded as “yes” for students who are continuously enrolled in Springdale or at any other four-year college (drawing on data from NSC) for four semesters or “no” for all others.
- Declaration of major by end of 4th semester will be coded as “yes” for students who have declared a major and “no” for students who have not. For students who have left the university, this outcome will be missing.
- Approval of IMP proposal by end of 6th semester will be coded as “yes” for students who have received approval and “no” for students who have not. Because approval of the IMP proposal is only possible for students enrolled at Springdale, students who have left the university before obtaining approval of an IMP proposal will be coded as “no,” and not as missing.
- Cumulative GPA is the cumulative grade point average as of a given semester, and will be measured as of students’ first, second, fourth, and sixth semester after entering Springdale University. For students who leave the university, the outcome will be missing for any semester from that point forward.

The Sense of Belonging Scale (SOBS; Hoffman et al., 2003) consists of 16 items across three subscales: perceived peer support (6 items); perceived faculty understanding/comfort (7 items); and perceived classroom comfort (3 items). Together, all 16 items create an overall measure of *sense of belonging*. The total scale has a Cronbach’s alpha of .90 across 16 items; (Tovar & Simon, 2010). The total scale score will be calculated for each student and will be used as the outcome measure.

Exhibit 4 summarizes the domains, outcomes, and measurement timeline for each cohort in the study. In addition, Exhibit 4 lists, for each outcome, anticipated baseline measures to be used 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 the particular subset).
- ☑ There’s no need to report reliability for scales that will not be analyzed.



EXAMPLE EVALUATION PLAN

Exhibit 4. Outcome Domains, Measures, Timing of Data Collection and Associated Baseline Measures

Domain	Outcome & Measure	Timing of Data Collection*	Baseline Measure(s)
Credit accumulation	Total credits earned (continuous)	End of 1 st , 2 nd , 4 th , and 6 th semesters	SAT-Critical Reading + Math combined score Pell grant eligibility status
	Completion of core courses (dichotomous)	End of 2 nd semester	
	Persistence in college (dichotomous)	End of 2 nd semester	
	Declaration of major by end of 4 th semester (dichotomous)	End of 4 th semester	
	Approval of IMP proposal by end of 6 th semester (dichotomous)	End of 6 th semester (cohort A only)	
Academic achievement	Cumulative GPA (continuous)	A: End of 1 st , 2 nd , 4 th , and 6 th semesters B: End of 1 st , 2 nd , and 4 th semesters	SAT-Critical Reading + Math combined score Pell grant eligibility status
Sense of belonging	Sense of Belonging Scale (SOBS): Total Score (continuous)	End of 1 st and 2 nd semesters	Sense of Belonging Scale Total Score

***For Cohort A, 1st semester ends December, 2016; 2nd semester ends May, 2017; 4th semester ends May, 2018; 6th semester ends May, 2019. For Cohort B, 1st semester ends December, 2017; 2nd semester ends May, 2018; 4th semester ends May, 2019; 6th semester for Cohort B ends after the end of the grant period (no 6th semester outcomes will be collected for Cohort 2).**

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](#).

Baseline measures. The evaluation will collect one baseline measure of academic achievement and one baseline measure of socioeconomic status. The baseline measure of academic achievement is students' combined SAT scores for Critical Reading and Math. The baseline measure of socioeconomic status is student Pell grant eligibility status.

In addition, we will assess students' sense of belonging at baseline using the same Sense of Belonging Scale used as an outcome measure. As noted above, the scale is administered to students at the New Student Orientation, prior to the start of their first semester.



EXAMPLE EVALUATION PLAN

We will assess the equivalence of the intervention and control groups on the relevant baseline measures (as shown in Exhibit 4). For outcomes in the credit accumulation and academic achievement domains, we will assess baseline equivalence on a baseline measure of academic achievement (SAT reading and math combined score) and a baseline measure of socioeconomic status (Pell grant eligibility). For Sense of Belonging, we will assess baseline equivalence on the baseline measure of the same scale used as an outcome measure.

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

- Treatment indicator: a variable indicating the group to which a student was randomly assigned (0=control group, 1=Coach4College group).
- Baseline measures: for the credit accumulation and academic achievement outcomes, students' prior achievement scores on the SAT Combined (Critical Reading and Math) and Pell-grant-eligibility status will be included as covariates. For the sense of belonging outcome, the corresponding August pre-first year score on the SOBS will be included as a covariate.
- Randomization blocks: dummy variables to account for the randomization group (1 or 2) and cohort—Group 1/cohort A; Group 2/cohort A; Group 1/cohort B; Group 2/ cohort B.
- Other covariates: Each model will also include covariates for (a) gender and (b) ethnicity, where gender is defined as 0=male, 1=female, and ethnicity is defined as 0=not Latino/Hispanic and 1=Latino/Hispanic.

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) 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 Group A and Group B and cohorts 1 and 2, creating four randomization blocks.

3.5 Statistical Analysis of Impacts

Analyses will estimate, for each outcome, the difference between students randomly assigned to the *Coach4College* intervention group and students randomly assigned to the business-as-usual group, adjusting for randomization blocks, baseline measures, and other student characteristics, such as gender and ethnicity. As an RCT with low (expected) attrition, it is not necessary to control for baseline measures in the analysis model, but their inclusion can help improve the precision of the impact estimate. We will include the specified baseline measures in the analysis model regardless of whether there is high attrition. However, because covariates that are only minimally related to outcomes could potentially reduce precision, we will use a backward selection procedure to eliminate covariates other than the baseline measures and randomization blocks.



EXAMPLE EVALUATION PLAN

Specifically, after adding all covariates to the model, we retain the covariates for gender and ethnicity only if these have a p-value < .20. 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).

Analyses of impacts on students' credit accumulation, academic achievement, and sense of belonging through their 2nd or 4th semester in college will pool both cohorts. Analyses of impacts on students' credit accumulation and academic achievement at the end of their 6th semester in college will include Cohort A (as students in Cohort B will not reach their 6th semester during the grant period). 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.

Impact Analysis Model

We will fit the following linear regression model:

$$Y = \beta_0 + \beta_1 Treatment + \beta_2 CohortA_Group1 + \beta_3 CohortA_Group2 + \beta_4 CohortB_Group1 + \beta_5 SAT + \beta_6 PellEligible + \sum_{j=7}^J \beta_j StudentCharacteristics + \varepsilon$$

Where

- Y is the student outcome;
- β_0 is the covariate-adjusted outcome for students in the control group (i.e., the intercept) and in the reference-category block (Cohort B Group 2);
- β_1 is the average effect of the *Coach4College* intervention;
- Treatment is a dummy variable indicating student treatment status (1 if randomized to treatment and 0 if control);
- CohortA_Group1 is a dummy variable indicating whether the student was randomized to treatment or control status within the block "Cohort A, Group 1", (1=yes, 0=no). The model includes 3 dummies to represent the four blocks, with Cohort B, Group 2 being the omitted (reference) block.
- CohortA_Group2 is a dummy variable indicating whether the student was randomized to treatment or control status within the block "Cohort A, Group 2", (1=yes, 0=no).
- CohortB_Group1 is a dummy variable indicating whether the student was randomized to treatment or control status within the block "Cohort B, Group 1", (1=yes, 0=no).
- $\beta_2 - \beta_4$ is the average deviation in the intercept for each block
- $\beta_5 - \beta_j$ are the effects of baseline measures of student achievement, SES, and other student characteristics;
- SAT is student's baseline SAT combined (critical reading + math) 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.



EXAMPLE EVALUATION PLAN

PellEligible	is student's eligibility for a Pell grant (1=yes, 0=no);
Student Chars	is a vector of other baseline student characteristics, including gender and Hispanic ethnicity;
ε	is the error term (i.e., the deviation between the student's observed and predicted outcome).

Note that for analyses of impacts on student sense of belonging, we will also include in the model a baseline measure of sense of belonging.

The parameter estimate, β_1 , provides a covariate-adjusted estimate of the impact of *Coach4College*.

The hypothesis test for β_1 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 (β_1) by the pooled standard deviation derived from the unadjusted sample standard deviations for the intervention and comparison groups.

We will estimate this linear regression model for all eight 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.

KEEP IN MIND...

Evaluators may choose to analyze impacts on binary outcomes using logistic regression or 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).

Treatment of Missing Data

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.



EXAMPLE EVALUATION PLAN

No imputation of missing outcome (or baseline) data will be conducted. Students missing outcome, baseline, or other covariate data will be dropped from the analysis model using listwise deletion. Thus, analysis samples for each contrast will include only students with non-missing baseline and outcome scores.

Adjusting for Multiple Comparisons

Conducting multiple tests of the intervention impact increases the risk of obtaining a “false positive” finding. In other words, the more tests that are conducted for outcomes in the same domain, the greater the chance of finding a statistically significant effect even when there isn’t one. To protect against false positive findings, we will adjust the threshold for statistical significance using the Benjamini-Hochberg correction when there is more than one impact estimated on outcomes in the same domain. The study will estimate impacts on five outcomes in the credit accumulation/persistence domain (total credits earned at Springdale, completion of the four core courses, declaration of a major, persistence in college through the 4th semester and approval of the IMP proposal by the 6th semester).

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.

TIP!

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

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 each outcome in the same domain: (1) total credits earned

at Springdale at the end of semester six; (2) completion of core courses at the end of semester two; (3) persistence in college at the end of semester two; (4) declaration of major at the end of semester four; and (5) approval of the IMP proposal at the end of semester six. We will not adjust for tests of impacts in earlier semesters.

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.

Cross-Overs/No-Shows

The study will use an intent-to-treat (ITT) analysis, in which students will be analyzed in the group to which they are randomly assigned, even if they fail to participate in the *Coach4College* intervention, or are assigned to the control group and receive the intervention. If a student assigned to the College Squad stops participating in intervention activities, he or she will still be included in the intervention group for analysis purposes. And if a student assigned to the control group winds up participating in College Squad activities, that student will still be treated in analyses as a member of the control group.

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.



3.6 Attrition

Because most of the outcome data will be extracted from Springdale's administrative records, we anticipate little to no missing data (i.e., very low attrition). However, missing data may be possible for some or all of the outcomes for students who choose not to participate in the evaluation (withdraw consent), transfer to another college, or dropout entirely.

For total credits earned, declaration of a major by the end of the fourth semester, and cumulative GPA, data will be considered missing for students who transfer or drop out. For two outcomes – (a) completion of Springdale's four required core courses by the end of second semester and (b) approval of the IMP proposal by the end of the sixth semester, students who transfer or drop out will not have missing data, but will have data from their last semester of enrollment at Springdale.

For persistence in college through four semesters, data will only be considered missing for students who drop out of college (and do not transfer to another institution). National Student Clearinghouse data will be used to supplement administrative data on persistence for transfer students. (We acknowledge that the NSC does not cover all four year institutions in the U.S., nor does it cover international institutions. Nevertheless, prior data on students who transfer from Springdale suggests that the majority do so to an NSC member institution.)

Attrition for the Sense of Belonging Scale may result from students withdrawing from the university prior to collection of the outcome, or declining to complete the scale. The evaluation team will make multiple attempts via email reminders to encourage students enrolled at Springdale to complete the online SOBS scale.

If either overall or differential attrition is high according to the WWC liberal attrition standards for any analytic sample, we will assess the baseline equivalence for that analytic sample (described below). Overall attrition will be calculated as the percentage of students randomly assigned to either condition that have missing outcome data. We will also calculate attrition in the treatment group (the percentage of students randomly assigned to the *Coach4College* condition for whom outcome data are missing) and in the control group (the percentage of students randomly assigned to the control condition for whom outcome data are missing). To assess differential attrition, we will calculate the difference between the attrition rate for the treatment group and the control group. For transparency of attrition calculations, we will report the total number of students randomized to each condition and the total number of students in each condition with non-missing outcome data – for each outcome measure analyzed.

TIP!

- ☑ Include the information necessary for the WWC to calculate overall and differential attrition for each analytic sample, when you report study findings.
- ☑ Report the total number randomized to each condition, and report the total number in each condition for which outcome data are non-missing.
- ☑ Remember that sample sizes can vary from one contrast to another.



3.7 Baseline Equivalence Testing

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 students and the 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 the baseline measures of academic achievement and SES (or without a missing baseline score on the sense of belonging scale, for analyses of impacts on sense of belonging).

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.

Should high attrition occur, 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.

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.

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 (β_1) 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.

$$\text{Baseline} = \beta_0 + \beta_1 \text{Treatment} + \beta_2 \text{CohortA_Group1} + \beta_3 \text{CohortA_Group2} + \beta_4 \text{CohortB_Group1} + \varepsilon$$

TIP!

- ☑ Be prepared to assess baseline equivalence for the analytic sample (or samples) if there is high attrition.
- ☑ Do not include any student who is missing the outcome measure in tests of baseline equivalence. 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.



EXAMPLE EVALUATION PLAN

For the SAT and for the Sense of Belonging Scale baseline measures, we will calculate the effect size of the difference in means using Hedges g . Specifically, we will divide the difference between the mean for the *Coach4College* group and the mean for the control group (i.e., the parameter estimate for the treatment variable, β_1) by the pooled standard deviation for the sample, S :

$$S = \sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}}$$

where n_i and n_c are the sample sizes for the *Coach4College* and control conditions, respectively, and s_i and s_c are the unadjusted student-level SDs derived from the evaluation sample for the treatment and control conditions, respectively.

The treatment and control students will be considered to be equivalent if the baseline difference is ≤ 0.25 , given that we will control for the baseline measure in the impact analysis model.

For the binary measure of Pell grant eligibility, 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 if the baseline difference is ≤ 0.25 , given that we will control for this baseline measure in the impact analysis model.

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.

Hoffman, M., Richmond, J., Morrow, J., & Salomone, K. (2003). Investigating “sense of belonging” in first-year college students. *Journal of College Student Retention*, 4, 227–256.

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.



EXAMPLE EVALUATION PLAN

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).

Tovar, E. and Simon, M.A. (2010). Factorial structure and invariance analysis of the Sense of Belonging Scales. *Measurement and Evaluation in Counseling and Development*. 43(3): 99-217.

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

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

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