2015 College Possible: Closing the Achievement Gap for Low-Income Students

i3 National Development Study

November 2019

Submitted to:
Lara Dreier
Director of Federal Partnerships
College Possible™ National

Submitted by:
Samantha Spinney, PhD
Kazuaki Uekawa, PhD
Johnavae Campbell, PhD

Grant Number: U411C140049

Table of Contents

I. Introduction

The College Possible program provides intensive coaching by near-peer mentors, peer group support, and opportunities for academic and non-cognitive skill building to low-income students through an intensive two-year curriculum in junior and senior years of high school. Coaches, serving as AmeriCorps members, guide students through all the key aspects of preparing for college during after-school sessions for two hours twice a week. Over the course of their junior and senior years, students complete 320 hours of curriculum in a supportive group of college-bound peers. The junior year curriculum orients students to the college application process, provides extensive preparation for the ACT/SAT exams, introduces students to college life through campus tours and allows time for students to apply for summer enrichment opportunities. The senior year curriculum leads students through the college application process, assists students in applying for financial aid and scholarships, and guides students through the transition to college. The goal of the program is to close the achievement gap between low-income students and their more affluent peers.

College Possible was founded in St. Paul, MN in 2000 and expanded to Milwaukee, WI in 2008; Omaha, NE in 2011; Portland, OR in 2012; and Philadelphia, PA in 2013.

College Possible's positive results have been confirmed by five independent evaluations. A Harvard University external evaluation of College Possible utilized a randomized controlled trial (RCT) design and found that the intervention has a significant positive effect on four-year college enrollment outcomes for low-income students (Avery, 2013). Furthermore, a 2013 evaluation conducted by ICF found that the College Possible coaching model has a significant positive influence on college success, helping to reduce historical achievement gaps in persistence.

1. Description of the Intervention

The College Possible program model employs two key intervention components with the goal of improving participating low-income students' non-cognitive skills and academic achievement outcomes:

(1) **Training and Support for College Possible High School Coaches:** AmeriCorps coaches who are recent college graduates are recruited and assigned to participating schools to serve a caseload of no more than 40 low-income students per year during their junior and senior years of high school. Coaches receive orientation training and attend 30 weekly meetings with other coaches during each school year. Training and support provided during these sessions prepares coaches to build effective relationships with students and to effectively and efficiently use College Possible's comprehensive curriculum to identify and support low-income student's academic and non-cognitive needs. In addition, coaches use the weekly meetings to share resources with each other and develop effective strategies to help address student needs. Throughout the year, coaches use a variety of real-time data on each individual student using the Naviance data system and later Salesforce. Coaches use these data and their personal interactions with students to provide intensive, targeted supports that help students navigate high school completion and the college preparation and enrollment process.

(2) **College Focused Sessions for College Possible High School Students:** College Possible coaches arrange their caseloads into small peer groups consisting of 10–15 students each. These small peer groups meet after school and in the evenings for two hours, twice each week over the course of their junior and senior years of high school. The small cohort size and consistent meeting schedule allow students to build a peer group of support, increasing their odds for successfully completing high school and enrolling in college. Coaches support students in their caseload as they practice goal setting, utilizing academic discipline, and building confidence and engagement as they realize their goals can be met through hard work and persistence. Participating students also receive a variety of additional academic supports as College Possible participants. One-on-one sessions are provided in addition to peer group sessions.

These key components of the College Possible program, when enacted in combination, are hypothesized to have a positive impact on participating students' non-cognitive skill development, especially the areas of academic engagement, commitment to learning, and sense of belonging. First, the intervention provides training and support for peer coaches. This is posited as necessary for coaches to build appropriate relationships with and properly identify the needs of participating students. Next, coaches act as facilitators providing college-focused sessions where College Possible students are thought to develop ongoing peer support networks and receive support to address their academic needs. Results of successful intervention are believed to build essential non-cognitive skills, leading to increases in successful graduation and postsecondary enrollment outcomes. While not measured in this evaluation, the College Possible model also posits that the long-term impact of the intervention on students' development is an increase in college persistence rates, financial security, and the development of a scalable program model.

For more details about the intervention and intended impacts, see the College Possible logic model in Appendix A.

2. Evaluation Overview

Under this Investing in Innovation (i3) development grant, College Possible served approximately 1,300 students in two cohorts (the class of 2018 and the class of 2019) during their junior and senior years of high school in 18 high schools across seven school districts and five states (i.e., Minnesota, Nebraska, Oregon, Pennsylvania, and Wisconsin).

The i3-funded evaluation of College Possible represents the first rigorous evaluation of the impact of the College Possible model upon participating low-income students' non-cognitive skills and is also the first comprehensive multi-site evaluation of the College Possible program. Despite these differences in outcomes and scale, the intervention to be evaluated in this i3-funded development study is identical to the intervention that has been previously evaluated.

College Possible contracted with ICF to conduct a federally mandated third-party implementation and impact evaluation of the 2015 i3 development grant. Throughout the evaluation, the National Evaluation of i3 (NEi3) technical assistance and support team provided feedback to the evaluation team to ensure an approved What Works Clearinghouse (WWC) impact design and high-quality fidelity of implementation (FOI) study.

3. Purpose of this Report

The purpose of this report is to provide an overview of evaluation findings at the culmination of the College Possible grant, including impact and implementation results. Impact and implementation study findings are presented separately in the following sections.

II. Impact Study

The independent evaluation of the College Possible program, led by ICF, explored program impact on students' non-cognitive skills and high school graduation status. The program expectation was that students' participation in the College Possible program should bring about positive changes in students' non-cognitive skills and increase the likelihood of their graduating from high school on time. To determine program impact, the ICF evaluation team conducted a quasi-experimental design (QED) study using two adjacent student cohorts (class of 2018 as Cohort 1 and class of 2019 as Cohort 2) that included treatment and comparison students from each cohort. The ICF evaluation team applied propensity score matching (PSM) to identify matched comparison students to those students participating in the College Possible program. Using the matched data that met WWC standards with reservations, the ICF team conducted statistical analyses to estimate the program impact on student outcomes. Non-cognitive skills were measured in both groups using the REACH Survey at the beginning of students' junior year and again at the conclusion of students' senior year. The REACH Survey consisted of items related to students' levels of academic effort, motivation, aspiration, cognitive focus, and understanding of their own interest. Graduation outcomes were measured by high school completion data provided by participating school districts. The following sections will detail the research questions, data sources, methods, analytical models, and results.

1. Research Questions

Implementation of the College Possible model was expected to ultimately have a positive impact on participating students' non-cognitive skill development. In turn, the increased non-cognitive skills were expected to contribute to increases in successful graduation and postsecondary enrollment outcomes. While postsecondary enrollment data were not available for all students included in the evaluation, the impact study sought to understand the impact of the College Possible model on both non-cognitive skills and graduation outcomes with the following two research questions:

Confirmatory Q1: What is the two-year impact of the College Possible program upon high-needs high school students' non-cognitive skills, as measured by the REACH Survey, as compared to similar high school students in the business-as-usual condition by the conclusion of the senior year?

Exploratory Q2: What is the two-year impact of the College Possible program upon high-needs high school students' high school completion rates as compared to similar high school students in the business-as-usual condition by the conclusion of the senior year?

2. Impact Study Methodology

To assemble the comparison group for the analysis of student non-cognitive skills, the ICF team worked with College Possible to administer the REACH Survey to all students in the same grade as those receiving the treatment. For two cohorts of students, the REACH Survey pre-test and post-test were administered, respectively, at the beginning of junior year and at the end of senior year. In addition, ICF requested demographic and academic data from the participating school districts for all students in the same grade level as those receiving the treatment. From the available pool of non-treatment students who took the REACH Survey pre-test and post-test, the ICF team used PSM to select the comparison students who resembled the treatment students on demographic and academic characteristics. The statistical analysis compared the two groups in terms of post-test REACH Survey scores while statistically adjusting various factors, such as prior-to-the intervention REACH Survey score, grade-point average (GPA), demographic characteristics, and school districts.

2.1 Impact Study Data Sources and Variables

2.1.1 REACH Survey Scores as Students' Non-Cognitive Skill Measure

ICF and College Possible selected the Search Institute's REACH Survey to measure students' non-cognitive skills (the confirmatory research question). Specifically, the evaluation team used an abbreviated version of the survey developed by the Search Institute with 28 items. The survey assessed the degree to which students have developmental relationships with teachers and the strengths necessary to achieve academic success in school and in life. The survey's 28 items covered five domains, including relationships in school, effort, academic aspirations, cognition (or the ability to manage thinking and be positive in the face of challenges), and understanding about one's own interests and talents. The internal consistency of the REACH Survey, as measured by Cronbach's Alpha, is 0.92. The reliability estimates for the five subscales range from 0.73 to 0.89. See Table B4, Appendix B for the wording and response values of the 28 items in the REACH Survey.

2.1.2 High School Graduation Outcome

ICF collected high school completion data from seven school districts to measure the impact of College Possible on high school graduation (the exploratory research question). The completion data included student status at the end of the student's senior year—whether the student completed high school on time, dropped out of high school, did not complete high school on time and were still enrolled in school, or transferred out of district and had stopped being tracked. ICF created a binary outcome indicator of high school completion either as "completed" if students completed high school on time and as "not completed" if students dropped out or did not finish on time. Students who moved out of district and students whose status was unknown to the district were treated as "missing" and did not become part of the analysis modeling process.

The graduation percentages collected mostly fell in the 90–100% range and lacked between-school and between-district variations necessary for the multilevel modeling analysis technique used for the REACH score analysis. The high graduation percentages were specific to the analysis sample and only reflected data from junior and senior years (i.e., they were not the

same as the four-year graduation rate). The graduation percentages used for the analysis sample should be interpreted only in the context of research questions posed. The public data shows that the four-year public high school graduation percentages (2016–17) were 83% (Minnesota), 89% (Wisconsin), 89% (Nebraska), 87% Pennsylvania, and 77% (Oregon).[1]

2.1.3 Data Structure and Merges

Table 1 summarizes the four data sources used for the data analysis. The original list of treatment students included 725 treatment students from Cohort 1 and 734 from Cohort 2. ICF obtained demographic data of all 11th graders from the seven participating school districts for each cohort (Cohort 1: 4,952 students; Cohort 2: 6,942 students). The REACH Survey pre-test dataset and post-test dataset for Cohort 1 included records from 2,946 students (pre-test) and 2,124 students (post-test). For Cohort 2, the REACH Survey dataset included 2,927 students (pre-test) and 1,822 students (post-test). Since the treatment students were known to the College Possible program, they were more consistently and systematically followed to be part of the data collection than the comparison students. The effort to include comparison students relied on the availability of students at the time of test administration.

Table 1. Raw Datasets and the Number of Student Records

	Cohort 1			Cohort 2		
	Total	Treatment group	Comparison group	Total	Treatment group	Comparison group
Total treatment students	**N/A**	725	N/A	**N/A**	734	N/A
Demographic data for all Grade 11 students	**4,952**	725	4,227	**6,942**	734	6,208
Pre-test REACH Survey data	**2,946**	663	2,283	**2,927**	582	2,345
Post-test REACH Survey data	**2,124**	518	1,606	**1,822**	389	1,433

Note: Students included in the study came from 18 schools in 7 different school districts.

The ICF team combined these multiple data sources to create two datasets to address the confirmatory REACH question and the exploratory high school graduation question. The high school graduation database combined demographic data, pre-test REACH data, and graduation information data collected from seven districts. As detailed in the next subsection, the database included pre-test REACH scores, together with baseline GPA, to improve the quality of the PSM. The post-test REACH scores, on the other hand, were not needed for the graduation analysis.

The REACH score analysis was based on a database that combined demographic data, pre-test REACH data, and post-test REACH data. Both post-test REACH scores and pre-test REACH scores were important parts of the statistical analysis.

[1] National Center for Education Statistics, Public high school 4-year adjusted cohort graduation rate (ACGR) (https://nces.ed.gov/ccd/tables/ACGR_RE_and_characteristics_2016-17.asp).

Table 2 summarizes how the original data were reduced in size in the treatment and comparison groups due to data attrition. Data loss may have occurred for different reasons: some students may have moved out of the school district, some students may have been retained in the prior grade, and some students may not have been available at the time of the pre-test or post-test administration.

For the REACH score analysis (confirmatory question), the number of observations available for the PSM was 2,312 (see Table 2, Row 3; Cohort 1: 1,286; Cohort 2: 1,036). The data available for the high school graduation outcome (exploratory analysis) included 4,507 cases (see Table 2, Row 4, Cohort 1: 2,197; Cohort 2: 2,310).

The important set of numbers to monitor are the number of treatment students that the analysis sample retained to keep the treatment sample as close as possible to the original list of treatment students. The original roster of treatment students included 725 Cohort 1 students and 734 Cohort 2 students. These numbers reduced to 447 (62% of the Cohort 1 treatment list) and 328 for the REACH analysis (45% of the Cohort 2 treatment list), respectively. For the graduation outcome analysis, the numbers reduced to 570 (79%) and 552 (75%), respectively.

Table 2. Results of Data Merges (Before Matching)

	Cohort 1			Cohort 2		
	Total	Treatment group	Comparison group	Total	Treatment group	Comparison group
1. Total treatment students	N/A	725	N/A	N/A	734	N/A
2. Demographic and pre-test REACH dataset	2,524	619	1,905	2,643	582	2,061
3. Demographic, pre-test REACH dataset, post-test REACH dataset	1,286	447	839	1,036	328	708
4. Demographic, pre-test REACH dataset, graduation outcome data	2197	570	1627	2310	552	1758

Note: The number of school districts (7) and schools (18) did not decrease after student-level attrition was taken into consideration. Students who did not assent to the testing were excluded.

2.2 PSM and Baseline Equivalence Test

ICF conducted PSM using one-to-one matching (i.e., one comparison student was matched to one treatment student) based on the datasets that combined multiple data sources. The predictors included in the PSM model were REACH Survey pre-test scores, baseline GPA (standardized with a mean of 0 and standard deviation of 1 within cohort and district combined), race and ethnicity groups (Asian, Black, Hispanic, White, other race groups), disadvantaged status, gender, and school district. Cohort and school district were used as the exact matching criteria (i.e., treatment and comparison students were matched within the same school district and the same cohort). Disadvantaged status may have been defined differently by states and/or

school district (e.g., free or reduced lunch status, household income or other poverty data, socioeconomic characteristics of residential areas); however, it did not affect the quality of matching result since students were matched within districts.

For the REACH score analysis (confirmatory question), baseline equivalence was established under the WWC. Table B1, Appendix B details the baseline equivalence analysis results. Post-test GPA and the proportion of Asian students had a standardized difference greater than 0.05 but smaller than 0.25 and, per WWC guidelines, such predictors need to be included in the statistical model. All predictors used in the PSM model were used in the final statistical model. "Other race category" did not meet the WWC threshold; however, it is a variable involving a very small number of students (i.e., 2 students in the treatment group and 15 students in the comparison group out of the whole sample of 1,250 students). Table 3 summarizes the number of students included in the analysis sample for meeting the matching criteria imposed by the PSM analysis.

Table 3. PSM Results for the REACH Score Analysis

	Total	Treatment Group Total	Comparison Group Total	Baseline Equivalence
Cohort 1	716	358	358	Established. Baseline GPA and proportion of Asian students need statistical adjustment. See Table B1, Appendix B for details.
Cohort 2	534	267	267	
Total	1,250	625	625	

Notes: the number of cases before the matching were: total 2,322; Cohort 1: 1,286 (T: 447, C: 839), Cohort 2: 1,036 (T: 328, C: 708).

The ICF team used the same matching method for the construction of the graduation outcome analysis. As mentioned earlier, post-test REACH score was not required for this analysis and thus the database used for matching and the resulting analysis sample were larger than the one used for the REACH score analysis. Pre-test REACH score, however, was an important part of the PSM matching process as it provided baseline information about students' non-cognitive skills. Since non-cognitive skill development is posited in the College Possible logic model (see Appendix A) to contribute to higher graduation outcomes, incorporating pre-intervention non-cognitive skills into the PSM model helped ensure that treatment students and comparison students were matched accordingly.

Baseline equivalence was also established by WWC standards (see Table B3, Appendix B for descriptive statistics and baseline equivalence test details). Post-test GPA and the proportions of disadvantaged students, Asian students, and White students had a standardized difference greater than 0.05 but smaller than 0.25 and, per WWC guidelines, such predictors need to be included in the statistical model. All predictors used in the PSM model were used in the final statistical model. "Other race category" did not meet the WWC threshold; however, it is a variable involving a very small number of students (i.e., 5 students in the treatment group and 18 students in the comparison group out of the whole sample of 2,048 students). Table 4 summarizes the number of students included in the analysis sample for meeting the matching criteria imposed by the PSM analysis.

Table 4. PSM Results for the Graduation Outcome Analysis

	Total	Treatment Group Total	Comparison Group Total	Baseline Equivalence
Cohort 1	1,026	513	513	Established. Baseline GPA, proportions of disadvantage status, Asian students, and White students need statistical adjustment. See Appendix B Table B3.
Cohort 2	1,022	511	511	
Total	2,048	1,024	1,024	

Notes: the number of cases before the matching were: total 4,507; Cohort 1: 2,197 (T: 570, C: 1,627), Cohort 2: 2,310 (T: 552, C: 1,758).

2.3 Analytical Models

2.3.1 Analytical Models for REACH SCORE Analysis

Using the analysis sample consisting of treatment students and matched comparison students, the analysis team used hierarchical linear modeling (HLM) to estimate program impact on students' non-cognitive skills. To address the clustering issue, the model estimated the intercepts (i.e., school effects) as random effects. The model used all covariates included in the PSM: pre-test REACH score, baseline GPA, gender, race groups (Black, Hispanic, Asian, other race groups, and White), and disadvantaged status. The model also included districts as a series of dummy variables and student cohort indicator (Cohort 1 vs. Cohort 2). As mentioned, school differences were estimated as random effects in the HLM framework. The program effect was estimated as the coefficient of the treatment status (1 if treatment, 0 if control) and the standardized effect size was presented to facilitate interpretation. The standardized program effect was derived by rerunning the same statistical model using the z-score version of post-test REACH scores (z-score used the sample mean and sample SD). The following equation summarizes the model described above.

$$Posttest_{ij} = \beta_{00} + \beta_{10} * pretest_{ij}$$
$$+ \beta_{20} * treatment_j + ... + r_{ij} + u_j$$

where

- Posttest represents post-test REACH scores
- Pretest represents the pre-test REACH scores
- Postscripts *i* and *j*, respectively represent student and school
- βs are parameters to be estimated and r and u are error terms
- The three ellipses (i.e., "…") indicate that the model will include multiple predictors (discussed earlier) and corresponding parameters
- Treatment represents the treatment status (1 if treatment group; 0 if control group)

2.3.2 Analytical Models for Graduation Outcome Analysis

To analyze program impact on graduation outcomes, the ICF evaluation team used comparison of simple descriptive statistics (percentages) instead of multivariate modeling. We compared the graduation percentages between treatment and comparison students separately by district and cohort. The complex multivariate Hierarchical Linear Modeling (HLM) was not used for this analysis, as the policy of student graduation and precise definition of student graduation seemed to vary by district. Furthermore, data quality of graduation information seemed to differ

by school district. To account for local variance associated with graduation outcomes, ICF decided to descriptively analyze the variance of graduation rates by school district.

3. Impact Study Analysis Results

3.1 Results for Confirmatory Analysis of Program Impact on Non-Cognitive Skills

In the HLM analysis, the program impact is the average score difference between the treatment and matched comparison groups. The program impact estimate was adjusted for all predictors included in the model, as well as between-school outcome differences (treated as random effects in the HLM framework). Table 5 shows the regression model results and Figure 1 summarizes the estimated program impact graphically. Table B1, Appendix B shows the descriptive statistics of the analysis sample.

The estimate for treatment students, or the program impact, was 0.06 and the result was statistically significant (at p=0.01). When standardized, the program effect was 0.12, which indicates that the program impact was positive, and the impact size was smaller than 0.20, which Cohen (1988) considered a small effect. WWC considers the effect size of 0.25 substantively important (2017).[2]

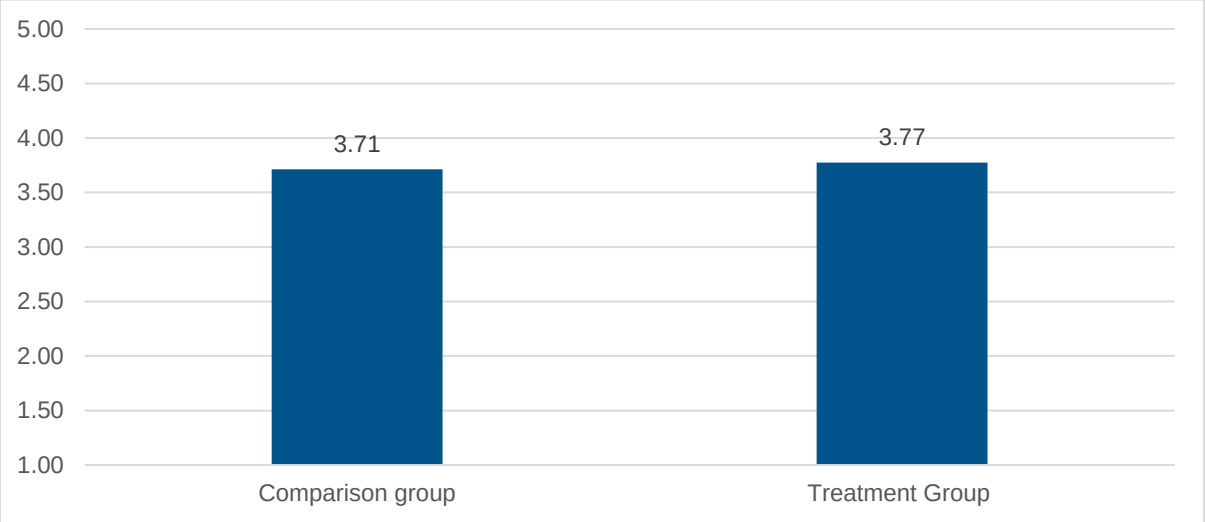
[2] What Works Clearinghouse Standards Handbook Version 4.0, page 77.

Table 5. Results of HLM Analysis for Cohort 1 and Cohort 2 Combined Sample (*n*=1,250)

	Estimates	Std. Error	*p*-value	Sig.	Standardized Effects
Intercept	3.65	0.06	0.00	***	-0.18
Treatment	0.06	0.02	0.01	**	0.12
Cohort 2	-0.04	0.02	0.07		-0.08
Pre-test REACH (centered)	0.56	0.02	0.00	***	1.08
Baseline GPA (z-score)	0.06	0.02	0.00	***	0.11
Disadvantage status	-0.03	0.03	0.32		-0.06
Male	0.02	0.02	0.32		0.05
Black	0.17	0.06	0.00	**	0.33
Hispanic	0.08	0.06	0.15		0.16
Asian	0.04	0.06	0.45		0.08
Other race group	0.08	0.11	0.49		0.15
Columbia Heights	0.11	0.07	0.11		0.21
Milwaukee	-0.05	0.04	0.27		-0.09
Minneapolis	-0.01	0.06	0.92		-0.01
Omaha	-0.02	0.04	0.66		-0.04
Park Rose	0.13	0.08	0.11		0.26
Philadelphia	0.13	0.15	0.39		0.25

Note: The omitted categories were White students and SPPS (St. Paul Public Schools); their estimates are represented by the intercept value (3.65). Significance test (2-tailed): ~ if $p<0.10$, * $p< 0.05$; ** $p< 0.01$, *** $p< 0.001$. HLM analysis estimated level-1 (within-individual) variance, level-2 (between-school) variance, and intra-c)lass correlation (level-2 variance / level1+level2 variance): Intercept-only model: 0.27, 0.01, 2%; The final model (presented above): 0.17, 0.00, 0%. The total variances (sum of level1 and level 2 variance) for the two models are, respectively, 0.27 and 0.17 (variance explained 37%). See Appendix B Table B1 for descriptive statistics of the variables used in the analysis.

Figure 1. Comparison of Post-Test Adjusted Average REACH Scores by Treatment Status



Notes: For ease of interpretation, the comparison group's average was fixed at the comparison group's unadjusted average score from the descriptive statistics (3.71; see Table B1, Appendix B) and the treatment group's average estimate (3.77) was the sum of comparison group average (again 3.71) and the coefficient from the multivariate model presented in Table 5 (0.06).

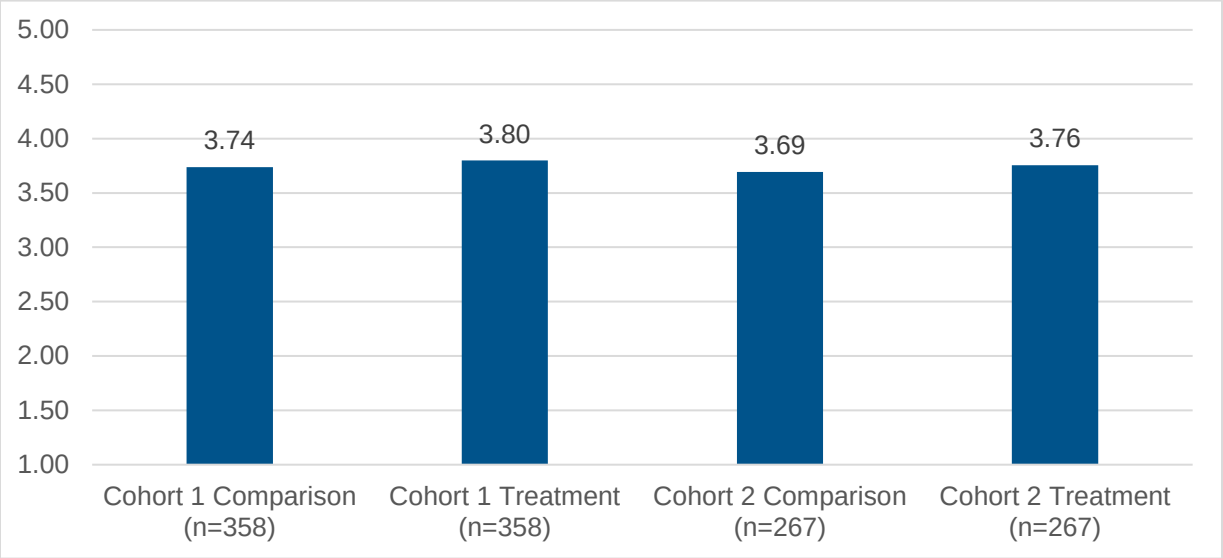
Additional analysis was conducted to see if the program impact varied by cohort. The model included four subcategories defined by cohort and treatment status and the adjusted post-test REACH average score was derived for each group (full results of the model are presented in Table B2, Appendix B). Table 6 summarizes the results from the final model. Figure 2 graphically represents the same information. The results suggest that the program impact did not vary by cohort. The standardized effect sizes were 0.12 for the whole sample (as already discussed), 0.12 for Cohort 1 (statistically not significant at p=0.05; significant at 0.10), and 0.12 for Cohort 2 (statistically not significant at p=0.05; significant at 0.10).

Table 6. Comparison of Program Impact Estimates by Final Model and Additional Model

	Adjusted Average Post-test REACH score	Program Impact	Standardized effect	Sig.
Final Model as Reference (Already presented in Table 4)				
Comparison group	3.71			
Treatment Group	3.77	0.06	0.12	**
Additional Analysis with Subgroups Defined by Cohort and Treatment Status Modeled				
Cohort 1 Comparison	3.74			
Cohort 1 Treatment	3.80	0.06	0.12	~
Cohort 2 Comparison	3.69			
Cohort 2 Treatment	3.76	0.06	0.12	~

Note: Significance test (2-tailed): ~ if p<0.10, * if p< 0.05; ** if p< 0.01, *** p< 0.001. For ease of interpretation, Cohort 1 comparison group's average scores were fixed at the unadjusted averages (3.74) and other subgroups' averages were based on the multivariate model estimates (see Appendix B Table B2). See Table B1, Appendix B for descriptive statistics of the sample; Table B2 for the full results of the ad-hoc analysis.

Figure 2. Comparison of Post-test Adjusted Average REACH Scores by Treatment Status and Cohort



3.2 Subgroup Impact Analysis

To further explore the data, we estimated program impacts specific to subgroups defined by gender, race and ethnicity groups, and achievement level. We created multiple subgroup datasets: female student sample, male student sample, high-achiever sample (baseline GPA equal to or above the median value), low achiever sample (below median), and subsamples defined by race and ethnicity groups (White, Black, Hispanic, and Asian). To estimate subgroup-specific program effects, we used the same HLM regression models used in the main analysis.

Table 7 shows the results of eight subgroup analyses alongside the result of the main analysis (as a reference). Figures 3 and 4 graphically represent the subgroup program impact effect sizes using the same information.

The impact effects derived from the whole sample, the female sample, and the high GPA sample were statistically significant at p=0.01. As these analyses were exploratory based on subsets of the main analysis sample, we should focus on the standardized effect sizes. The rightmost column shows the subgroup impact sizes' deviation from the main sample estimate.

The findings suggest the following:

a) The program effect is greater for female students than for male students (standardized program impacts are 0.16 vs. 0.06; the difference is 0.10).
b) The program effect is greater for students with a high baseline GPA than for those with a low baseline (standardized program impacts are 0.17 vs. 0.06).
c) The program impact for White students (0.31) appears greater than other race groups (e.g., the impact for Hispanic students is 0.08). The analysis is based on a small sample size (n=75), though, and one-third of White students are concentrated in one school in Omaha (n=24).

We conducted a formal testing of whether the program impacts were greater for female students, White students, or students with a high baseline GPA, using the whole sample and the statistical interaction test. None of the interaction effects were statistically significant.

Table 7. Comparison of Program Impact Estimates by Subgroup Analysis Models

Subgroup	Total cases	Impact Estimate	Sig.	Standardized Effect	Standardized Effect Size Deviation from the Whole Sample Estimate
Whole sample	1,250	0.06	**	0.12	N/A
Male students	507	0.03		0.06	-0.06
Female students	743	0.08	**	0.16	0.04
High pre-test GPA	625	0.09	**	0.17	0.05
Low pre-test GPA	625	0.03		0.06	-0.06
White students	75	0.16		0.31	0.19
Black students	336	0.07		0.13	0.01
Hispanic students	334	0.04		0.08	-0.04
Asian students	488	0.07	~	0.14	0.02

Note: Significance test (2-tailed): ~ if p<0.10, * if p< 0.05; ** if p< 0.01, *** p< 0.001.

Figure 3. Subgroup-Specific Standardized Program Effects by Gender and Pre-Test GPA Levels

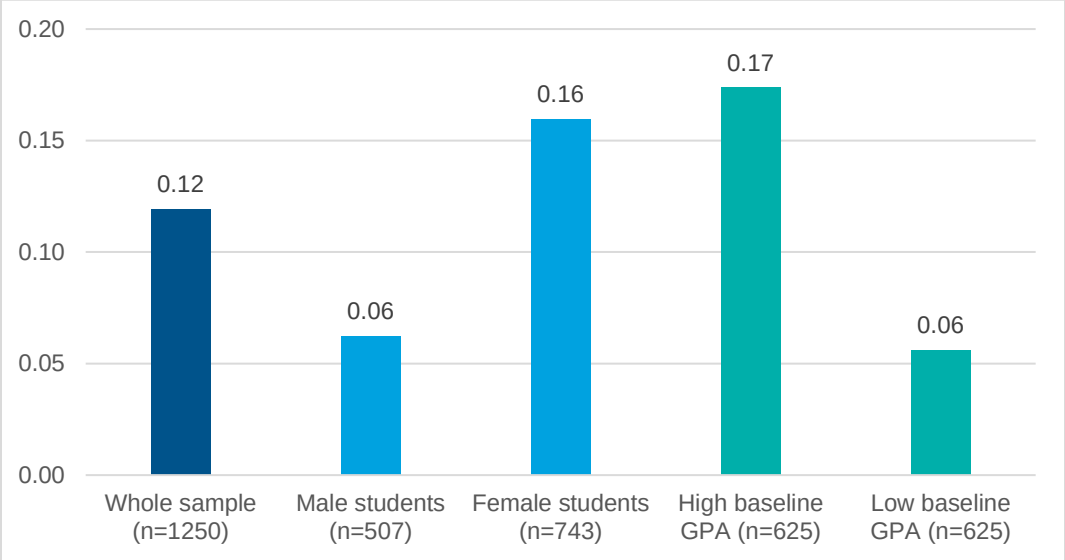
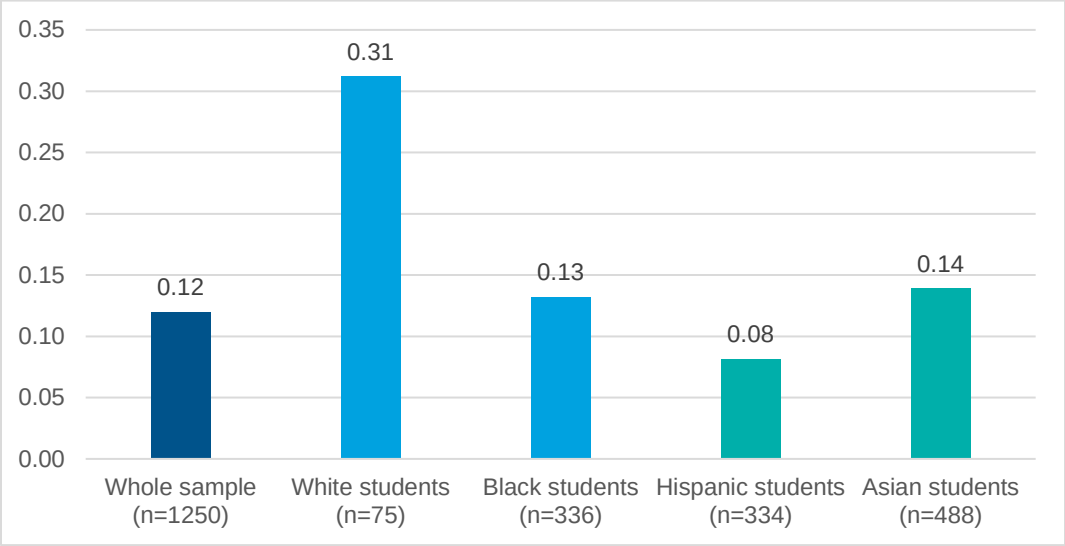


Figure 4. Subgroup Specific Standardized Program Effects by Race Categories



One possible explanation for the not statistically significant but suggestive subgroup findings is that subgroups that are academically advantaged may be more prone to the intervention as they may be more accepting of what is provided to them as mentoring. This is consistent with the finding that the program impact was greater for the high achiever sample (baseline GPA higher than the median value) than the low achiever sample. As shown in Table 8, we compared the baseline GPA averages by subgroups. The baseline GPA scores were higher for White students and female students than other groups.

Table 8. Comparison of Pre-test GPA Averages by Student Demographic Characteristics

		Pre-test GPA (z-score)	
	Total	Mean	Std. Dev.
Race			
Black	336	-0.46	1.02
Hispanic	334	0.03	0.96
Asian	488	0.25	0.90
White	75	0.39	0.98
Other race	17	-0.48	0.88
Gender			
Male	507	-0.15	1.04
Female	743	0.10	0.96

Notes: For ease of interpretation, individual-level pre-test GPA scores were standardized with the analysis sample mean of zero and a standard deviation of 1 and the subgroup averages were derived. The differences among the subgroup means can be interpreted as z-score differences (e.g., the average value 0.39 from the White student group means that the value is 0.39 higher than the sample average in standard deviation unit).

3.3 Results of Graduation Outcome Analysis

This section compares the study participant districts by the proportion of students who graduated high school on time and examines how the proportion was different by treatment status (treatment vs. matched comparison). As discussed earlier, the result is based on the

matched analysis sample and two years of data; thus, the graduation rates derived from this analysis should not be misunderstood to be the four-year graduation rates of these districts.

We begin with three districts whose overall number of students in the matched analysis sample was small. Accordingly, the results may be unreliable. As shown in Table 9, the data were defined by two cohorts and three districts and thus the table includes six contrasts. Per each contrast, the table compares the number of students and graduation percentages. The rightmost column summarizes the percentage difference between the treatment and comparison groups. The difference was 0% in three contrasts as a result of both treatment and comparison group showing the perfect graduation rate (Philadelphia Cohort 2, Park Rose Cohort 2, Columbia Heights Cohort 1). When graduation rates varied by group in other contrasts, the treatment group had a higher graduation rate than comparison group (Philadelphia Cohort 1, 8.3%; Park Rose Cohort 1, 40%; Columbia Heights Cohort 2, 5.7%).

Table 9. Three Districts with Small Number of Cases: Comparison of Graduation Rates by Districts and Cohort

	Cohort	Group	*n*	% Graduated	Group difference in %
Philadelphia (*n*=34)	1	Comparison	12	92%	8%
		Treatment	12	100%	
	2	Comparison	5	100%	0%
		Treatment	5	100%	
Park Rose (*n*=40)	1	Comparison	10	60%	40%
		Treatment	10	100%	
	2	Comparison	10	100%	0%
		Treatment	10	100%	
Columbia Heights (*n*=124)	1	Comparison	27	100%	0%
		Treatment	27	100%	
	2	Comparison	35	94%	6%
		Treatment	35	100%	

Notes: The graduation percentages reported in this table are 2-year graduation percentages of the analysis sample and not of the real district populations.

Table 10 reports on the four other school districts whose number of cases in the matched analysis sample was relatively large. Figures 5 and 6 summarize the same information separately for two cohorts. The table summarizes the proportions of student graduation for four districts, two cohort of students, and two groups (treatment vs. control), generating eight contrasts. The rightmost column summarizes the result of the treatment vs. comparison group contrast by showing the difference of graduation percentages. One contrast group, Milwaukee Cohort 1, had the identical percentage of graduation (100% and 100%). All other seven group contrasts showed percentages in positive numbers (ranges: 1–10%), meaning that the treatment students had a higher rate of graduation than comparison students.

Table 10. Four Districts with Large Number of Cases: Comparison of Graduation Rates by Districts and Cohort

	Cohort	Group	*n*	% Graduated	Group difference in %
Minneapolis (*n*=172)	1	Control	48	90%	6%
		Treatment	48	96%	
	2	Control	38	89%	8%
		Treatment	38	97%	
Milwaukee (*n*=506)	1	Control	127	100%	0%
		Treatment	127	100%	
	2	Control	126	98%	1%
		Treatment	126	98%	
Omaha (*n*=480)	1	Control	124	90%	10%
		Treatment	124	100%	
	2	Control	116	90%	10%
		Treatment	116	100%	
St. Paul (*n*=692)	1	Control	165	89%	8%
		Treatment	165	98%	
	2	Control	181	92%	3%
		Treatment	181	95%	

Notes: The graduation percentages reported in this table are 2-year graduation percentages of the analysis sample and not of the real district populations.

Figure 5. Cohort 1 Student Graduation Percentage per Group in Four Districts

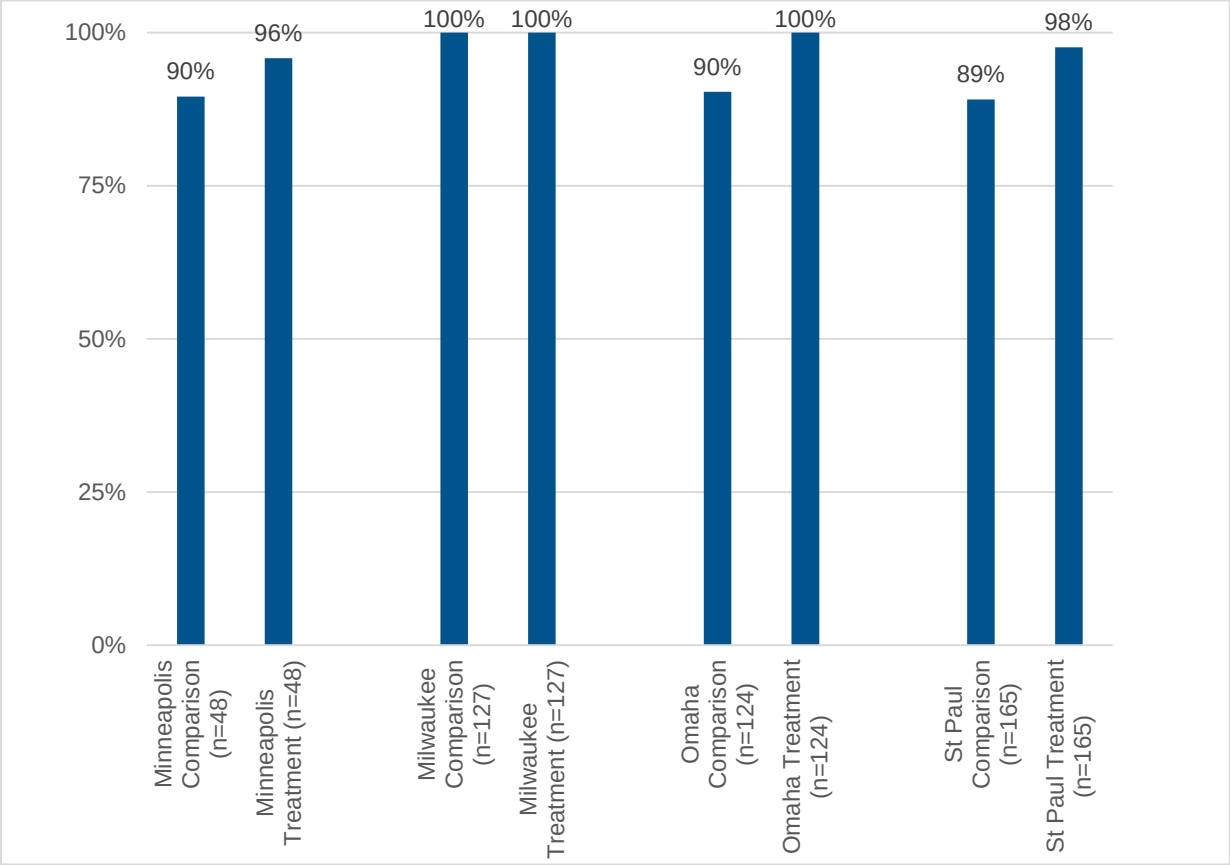
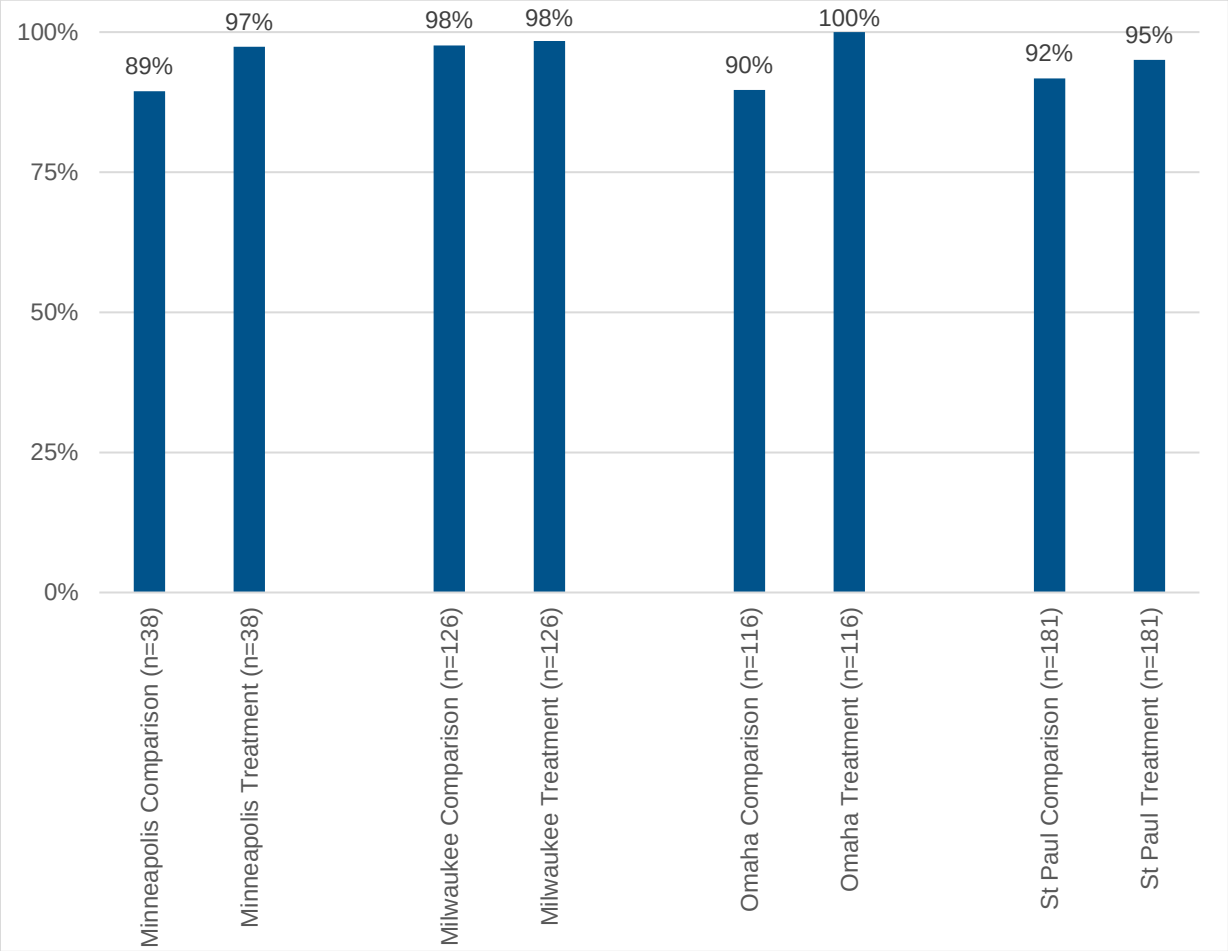


Figure 6. Cohort 2 Student Graduation Percentage per Group in Four Districts



3.4 Summary of Program Impact Analysis

The impact analysis section addressed two research questions regarding students' non-cognitive skills (REACH scores) and students' on-time high school graduation. The findings are specific to the analysis samples used for each analysis (analysis samples were defined by students' participation in the College Possible program and by comparison students whose characteristics were similar to the treatment students). This means that the graduation rates presented should not be interpreted as those of all students in the seven districts.

The statistical analysis suggested that the treatment students (in Cohorts 1 and 2 combined) had a higher level of non-cognitive skills (measured by REACH) than matched comparison students. The program impact estimated was statistically significant (at p=0.01) and the standardized effect size was 0.12. This effect size is considered rather small by WWC guidelines that considers the effect size of 0.25 as substantively meaningful. The effect size around 0.10, however, is not negligible in education research (Lipsey & Wilson, 2001). As two years of intervention returned an effect size of 0.12, more years of intervention throughout high school years could add to the effect that may accumulate to be of a meaningful size.

The subgroup analysis suggested that the program impacts for female students, White students, and high achieving students seemed greater than other subgroups. These findings are only

suggestive as none of the subgroup impact differences were statistically significant when evaluated by the statistical interaction models; however, it may point to the possibility that subgroups of students who are academically oriented may be better recipients of the program benefit.

The findings from the second analysis are based on basic descriptive comparison of high school graduation rates by school district. The analysis sample lacked between-district outcome variance necessary for meaningful analysis. However, when there were group differences, the treatment students' graduation rate was higher than that of the comparison students. This is consistent with the program expectation that the intervention is positively correlated with students' graduation outcome.

III. Implementation Study

ICF conducted an FOI study, required by the NEi3 project, to measure the extent to which the College Possible program was implemented as intended in participating schools during the 2016–17 school year (SY16–17), 2017–18 school year (SY17–18), and 2018–19 (SY18–19). The College Possible logic model specifies two key components (KCs) to the intervention: training and support for College Possible high school coaches (KC 1) and college-focused sessions for College Possible high school students (KC 2). The program used both KCs to support a college readiness intervention model that was posited to build participating low-income high school students' non-cognitive skills. The timeframe and sample of students in the implementation study aligned with those previously described for the impact study.

1. Research Questions

The implementation study documented the extent to which the intervention's two KCs were implemented with fidelity. The following two research questions align with the KCs found in the logic model (Appendix A) and were studied using the Fidelity of Implementation framework developed and submitted to NEi3 in the study's design summary template:[3]

1. To what extent do College Possible peer coaches participate in training and support provided through College Possible as intended?
2. To what extent do College Possible high school students engage in college-focused sessions as intended?

2. Implementation Study Methodology

FOI was calculated and coded to represent the extent to which each participant met the associated indicator's implementation threshold (typically measured as low, medium, or high). Once indicator implementation scores were derived for each individual participant—coach and student—they were summed across indicators to arrive at a KC Implementation Score for each participant (typically measured as low, medium, or high). We then calculated the percentage of

[3] The full name of the design summary template submitted to NEi3 is the DEV90 Design Summary Template.

study participants meeting the criteria for "high" implementation for each KC and compared this to an established threshold for "high" fidelity (e.g., greater than 80%). If the percentage of study participants in the entire sample who met the criteria for "high" implementation met or exceeded this threshold, fidelity of implementation was considered to be met with "high" fidelity for the DC (see Table C1, Appendix C for details of the reporting system).

2.1 Implementation Fidelity Measurement System

The ICF evaluation team measured FOI during SY16–17, SY17–18 and again in SY18–19 using a collaboratively developed implementation fidelity measurement system that included 10 indicators aligned to the two KCs of the College Possible program logic model. ICF and College Possible identified each indicator and set implementation thresholds at the outset of SY16–17. KCs, associated fidelity indicators, and data sources appear in Table 11. In 2015, ICF and College Possible identified each initial indicator and set implementation thresholds for the study of College Possible (see Appendix C, Table C2 and C3 for details of the implementation fidelity system).

In 2017, indicators were revised to reflect additional data sources for Cohort 2. KCs, associated fidelity indicators, and data sources appear in Table 11.

Table 11. KC, Indicator, and Data Sources for the FOI Study

Measuring Implementation Fidelity		
Key Component	**Indicator**	**Data Source**
KC 1. Training and support of College Possible coaches	1A. Coaches are assigned a caseload of Grade 11 students	Administrative roster records
	1B. Near-peer coaches receive annual two-part training orientation	Training attendance records
	1C. Near-peer coaches attend weekly meetings with their colleagues	
	1D. Near-peer coaches receive adequate support to serve students	Survey Data Annual coach survey and focus group transcripts
	1E. Near-peer coaches share resources and develop effective strategies	
	1F. Near-peer coaches use data to guide student intervention	
KC 2. College focused sessions for College Possible high school students	2A. College Possible students placed in peer groups	Attendance records
	2B. College Possible students attend weekly sessions	
	2C. College Possible students participate in one-on-one sessions	

Measuring Implementation Fidelity		
Key Component	**Indicator**	**Data Source**
	2D. Students receive opportunities to practice non-cognitive skill building	Annual student survey (SY16–17) and (SY17–18) Attendance records and program documents

2.2 Data Sources

The implementation study drew data about the 10 indicators in Table 11 from the following sources: (1) administrative roster records, (2) student surveys, (3) coach evaluation surveys, (4) attendance records, (5) coach focus groups transcripts, and (6) student focus groups transcripts. Details about each data source are as follows:

Administrative roster records: College Possible staff submitted roster records of student group assignment to ICF annually in January. These records were used to track student group assignment. Each student was given a session code by site meeting times, and meeting days to establish a final peer group placement count.

Coach evaluation survey: ICF worked with the College Possible staff to develop survey questions to be included in the administration of an existing survey (SY16–17) and (SY17–18) to all coaches in participating schools. The survey asked about: (1) the extent to which coaches received adequate and appropriate support and professional development related to College Possible sessions; (2) the quality, relevance, and usefulness of this support; (3) perceptions about educators' needs for future support; and (4) the extent to which educators engaged in meaningful interactions with peer coaches.

Student evaluation survey: In the first two school years, ICF created a survey question to be included in the existing student survey. Students were asked to indicate whether they practice the use of non-cognitive skills and in SY17–18 students were additionally asked about certain types of non-cognitive skills. This question was replaced with attendance records and program documents that align with non-cognitive skill practice sessions for the final year.

Attendance records: College Possible attendance records were tracked and submitted to ICF with final session attendance count, total number of sessions offered and overall percentage. Final count also included total attendance by full or partial attendance.

Coach focus groups: In addition to the above data sources, virtual focus groups with the coaches provided additional sources of data. These sessions took place annually in 2017, 2018, and 2019 with each of the participating coaches. The purpose of these focus groups was to better understand the details of College Possible implementation, training supports, and document best practices and challenges to implementation and sustainability.

Student focus groups: ICF conducted virtual focus groups with students from six school sites. These students were asked questions about their perception of College Possible services and impact on their academic and non-academic skill development. Focus groups provided additional sources of data. These sessions took place annually in 2017, 2018, and 2019.

2.3 Data Collection

College Possible was responsible for data collection and implementation of survey items for each school year. Data were transmitted to ICF in December and July of each calendar year. After Year 2 of the study, the evaluation team consulted with the program developer and revised and consolidated data collection efforts to reflect different expectations for schools to better align with the new data sources available and program theory. Some changes were made to KCs and their associated indicators before the last year of the study. These changes are summarized below by KC.

- KC1: Six indicators were associated with this KC in each year of the study. The threshold for the indicators was not changed, however the measurement tool for this component shifted away from survey data and focused instead on focus group interview data, which allowed College Possible to better understand context around usage and participation.
- KC2: Four indicators were associated with this KC in each year of the study. The threshold for the indicators was changed from 70% of available sessions to a threshold of 15 sessions for seniors and 25 sessions for juniors per year. Additionally, for the final year of the grant, the measurement tool was updated from a self-reported survey to the attendance records system, which allowed ICF to better measure participation in lessons focused on non-cognitive skills.

All data sources were developed and maintained by College Possible, with consultation from ICF. College Possible was responsible for data collection, and implementation data were transmitted to ICF in summer/fall 2017 and summer/fall 2018, and again in summer 2019, respectively.

3. Analysis

3.1 Implementation Study Analysis

ICF evaluation staff first calculated individual indicator implementation scores for each of the 42 treatment coaches at schools and 721 students remaining in the study at the conclusion of SY18–19. All 10 fidelity indicators were scored for each individual site. The resulting scores were then coded to represent the extent to which each site met the associated indicator's implementation threshold (typically measured as low, medium, or high). Once indicator implementation scores were derived, they were summed within each KC to arrive at a single KC implementation score for each treatment school (typically measured as low, medium, or high).

The ICF evaluation team then calculated the percentage of treatment coaches and students meeting the criteria for "high" implementation for each KC and compared this to an established threshold for "high" fidelity (e.g., greater than 80%). If the percentage of schools in the entire sample who met the criteria for "high" implementation met or exceeded this threshold, fidelity of implementation was considered to be met for the KC at the sample level. Fidelity was calculated and reported in this manner for two years as part of the study investigating implementation for Cohort 1 (SY16–17 and SY17–18). Fidelity was also calculated and reported the same way for Cohort 2 (SY17–18 and SY18–19).

It is important to note that the denominator for FOI calculations included only those coaches/students that were remaining in the treatment group at the end of each school year. All 10 fidelity indicators were scored for each school. The resulting scores were then coded to represent the extent to which each school met the associated indicator's implementation threshold (typically measured as low, medium, or high). Once indicator implementation scores were derived, they were summed within each KC to arrive at a single KC implementation score for each treatment school (typically measured as low, medium, or high).

The ICF evaluation team then calculated the percentage of treatment schools meeting the criteria for "high" implementation for each KC and compared this to an established threshold for "high" fidelity at the sample level (e.g., greater than 80%). If the percentage of schools in the entire sample who met the criteria for "high" implementation met or exceeded this threshold, FOI was considered to have been met for the KC at the sample level.

4. Implementation Study Results

As previously described, the FOI study was supplemented by focus groups conducted in both cohort phases of the study. Data collected during these groups indicated mostly positive findings about the College Possible experience, with many successes stemming from the College Possible trainings, regular College Possible student sessions and one-on-one support, and opportunities that come from practicing non-cognitive skills in informal as well as during formal session instruction. Overall, students and coaches participating in focus groups reported College Possible implementation within their respective schools to be successful as it related to their expectations and impact on college readiness, including increased ACT scores, access to scholarships, and better college fit selection. Findings from the focus groups suggested progress in students' non-cognitive skills, communication and confidence. The following subsections discuss implementation findings related to each of the KCs.

4.1 Implementation Fidelity by Key Component

In this section, we provide a summary of KC-level fidelity outcomes for both cohorts. Fidelity outcomes are based on the following numbers of coaches and students for each school year of the study:

- SY16–17 (Junior Year Cohort 1): 23 coaches and 721 students;
- SY17–18 (Senior Year Cohort 1): 20 coaches and 672 students;
- SY17–18 (Junior Year Cohort 2): 22 coaches and 723 students; and
- SY18–19 (Senior Year Cohort 2): 20 coaches and 678 students.

The following findings are organized by KC and list the status as reported to the NEi3 in the study's design summary template.[4] Table 12 summarizes these findings. The results are highlighted for both cohorts in the study. Overall, College Possible reached adequate fidelity on one out of two components for first cohort of the study and reached the threshold for adequate FOI for the first year on one out two components for Cohort 2 during the study.

[4] The full name of the design summary template submitted to NEi3 is the DEV90 Design Summary Template.

KC1: High implementation fidelity to KC1, training and support of College Possible coaches, was not met with high fidelity for the final year of the study.

This KC consists of six indicators, five of which focus on the participation and quality of training and support offered through the College Possible program: caseload, orientation and weekly session attendance, and adequate support and meaningful interaction throughout the school year. The remaining indicator focused specifically on the use of data to guide student intervention. To achieve high fidelity for this component, coaches had to attend at least 75% of each part of a two-part orientation training held during the months of August and September and participate in at least 25 weekly sessions during the year. Additionally, coaches surveyed had to report high levels of agreement with the quality, support, and usefulness of their preparation experience and trainings for their College Possible implementation efforts, as well as agree that the technology and materials were useful.

Calculated annually, but reported once over each two-year period, a total of 18 out of 20 Cohort 1 coaches (90%) met the threshold in SY17–18 and 17 out of 20 (85%) Cohort 1 coaches met the threshold in SY16–17. For Cohort 2, 82% of the 22 coaches met the threshold in SY17–18 and 13 out of 20 coaches (65%) met the threshold in SY18–19 with high fidelity. Overall, most indicators relevant for KC1 were met with high fidelity for all four years and at the conclusion of the two-year period for both cohorts. However, for the final year, the indicators that seemed to be particularly challenging were 1D, near-peer coaches receive adequate support to serve students, and 1E, near-peer coaches share resources and develop effective strategies.

KC2: High implementation fidelity to KC2, college focused sessions for College Possible high school students, was not met for the final year study.

KC2 refers to student participation in college-focused College Possible sessions. Initially, the 2015 grant fidelity was based on attendance and survey results as data sources, but the final year was measured solely by attendance records. This KC was calculated annually and reported once for each student cohort across two years of the study.

For Year 1, 56% of the 731 Cohort 2 students met the threshold with high fidelity. For Year 2, while KC2 was again not met with fidelity at the established threshold of 81%, 534 of the 678 Cohort 2 students (79%) participated at medium fidelity (40%) and high fidelity (39%). The remaining 21% of 678 students were measured at low fidelity in SY18–19. For the final year, the indicators that seemed to be particularly challenging in this KC were 2B and 2C attendance at one-on-one and weekly sessions.

Table 12: FOI Status by KC and Cohort

KCs in the Logic Model	Definition of High Implementation	Definition of "implementation with fidelity" at program level	Cohort 1		Cohort 2	
			SY16–17	SY17–18	SY17–18	SY18–19
KC 1: Training and support of College Possible coaches	Calculation based on 6 indicators (#1A through #1F) at the end of each year of implementation	81–100% of coaches have "high" implementation at the end of each year	Met "High" Fidelity	Met "High" Fidelity	Met "High" Fidelity	Low Fidelity
KC 2: College focused sessions for College Possible high school students	Calculation based on 4 indicators (#2A through #2D) at the end of each year of implementation	81–100% of students received college-focused sessions with "high" implementation at the end of each year	Low Fidelity	Low Fidelity	Low Fidelity	Low Fidelity

4.2 Implementation Fidelity by Indicator

Table 13 presents a breakdown of fidelity performance data comparing local and national study results by year. The table is organized by KC and lists the corresponding indicator(s) and scoring details and thresholds (e.g., low, medium, and high) as reported to the NEi3.

Fidelity to each indicator was assessed using the same scoring criteria established for each indicator's respective KC. For example, the threshold for high fidelity to KC1 is that 81% of the sample will achieve high implementation fidelity when data are aggregated across indicators 1A – 1F. To make a fidelity determination separately for each individual indicator (i.e., 1A, 1B, 1C, and 1D), we first assessed what percentage of the sample met the criteria for "high" fidelity on each indicator. If at least 81% of the sample met the criteria for "high" fidelity at the indicator level, we determined fidelity was "met" for the indicator. When survey items are used for fidelity indicators, denominators are determined by the actual number of respondents for that item rather than the number in the study.

Of the two KCs and corresponding indicators highlighted in Table 13, KC 1 targeted at coach training and support, included six indicators 1A through 1F. Indicators 1A, 1B, and 1C were implemented with high fidelity consistently across both cohorts of the study. Indicator 1A measured whether each coach was assigned a caseload. Indicators 1B and 1C measure coach attendance at orientation and trainings throughout the school year. For indicator 1B, coaches were required to attend at least 81% of weekly sessions offered. For indicator 1C, 75% of attendance across both orientation segments was required for coaches to meet the threshold; and in the final year 90% of coaches met this indicator with fidelity.

Over the course of the four-year study, College Possible made some positive changes to the program and subsequently ICF modified data collection sources to include the addition of focus group interview data. Indicator 1D measured the adequacy of support received by coaches from

the College Possible staff and supervisors as measured by survey items and supported by focus group data. The goal of this support was to build coach confidence and capacity to meet students' needs and deliver quality services in schools. An average rating of agree or strongly agree was required across survey items to meet fidelity, and for the first three years, fidelity was met. However, the final year survey data coupled with focus group interviews revealed that while some coaches expressed adequate support by College Possible staff, 75% of 20 reported they did not receive adequate support, resulting in not meeting the "high" fidelity threshold for indicator 1D in the final year. Coaches shared in focus group interviews that while College Possible was successful in effectively providing college entrance examination training and data resources, there was inconsistent communication and limited staff support to meet high expectations. Additionally, indicator 1E was not met with fidelity. Survey findings and focus group interviews indicated that approximately 60% of 20 coaches met fidelity for indicator 1E. However, others reported that they did not receive enough time to share resources and develop effective strategies, with one another.

Indicator 1F described the level of data use by the coaches for student intervention. This indicator was met with "high" fidelity. Approximately, 17 of 20 (85%) of coaches reported using some type of data to guide their student intervention. It appears that during the senior year cohort experience, coaches utilized data more frequently than during the students' junior year. While coaches varied in their reported amount of data use and support, most coaches commented on the effort by their supervisors and peers to provide ongoing mentorship and peer support. This effort translated to strong outcomes for students.

Four indicators supported KC2, students' receipt of adequate support and training. For the final year, all indicators were measured by attendance records provided by College Possible staff. Indicator 2A measured whether each student was assigned a peer group between 5 and 25 students in attendance, a number optimizing the small group peer interaction. Indicator 2A was implemented with high fidelity consistently across both cohorts of the study. For indicators 2B and 2C, students had to attend at least 15 weekly sessions as seniors and participate in at least two one-on-one sessions; the 81% threshold for fidelity was not met. The final indicator, 2D measured whether students were offered the opportunity to practice non-cognitive skills with peers. While this indicator was previously evidenced by a survey item, the final year data was measured by student attendance. To capture this data in another meaningful way, College Possible supplied lesson plans and curriculum by session data to establish that non-cognitive skills content was promoted and offered. ICF used this data as evidence to capture whether the students were offered non-cognitive skills participation in these sessions, replacing the survey question asking about whether the non-cognitive skills were practiced. Seventy-five percent of senior students attended these sessions, just under the required 81% to meet fidelity. Thus, indicator 2D did not meet the threshold of "high" fidelity.

Table 13. FOI for SY 2016–17 through SY 2018–19 by KC and Indicator

			Cohort 1		Cohort 2	
KC	**Indicator**	**Scoring Details**	**SY16–17 (Year 1) Performance Data (*n*=20)**	**SY17–18 (Year 2) Performance Data (*n*=20)**	**SY17–18 (Year 1) Performance Data (*n*=22)**	**SY18–19 (Year 2) Performance Data (*n*=20)**
KC1. Training and support of College Possible coaches	1A. Coaches are assigned a caseload of Grade 11 students.	<9 or >45 = Low (0) 10–19 = Med (1) 20–45 = High (2) # Assigned to Coach	*95% of 20 Met "high" fidelity*	*90% of 20 Met "high" fidelity*	**95%** of 21 Met "high" fidelity	**100%** of 20 Met "high" fidelity
	1B. Near-peer coaches receive annual two-part training orientation	<75% Part I & II = Low (0) >/=75% Part I or II = Med (1) >/=75% Part I & II = High (2) % Participation	*95% of 20 Met "high" fidelity*	*85% of 20 Met "high" fidelity*	**90%** of 21 Met "high" fidelity	**90%** of 20 Met "high" fidelity
	1C. Near-peer coaches attend weekly meetings with their colleagues	0–18 = Low (0) 19–24 = Med (1) 25 or more = High (2) # Meetings Attended	*83% of 20 Met "high" fidelity*	*100% of 20 Met "high" fidelity*	**100%**of 21 Met "high" fidelity	**95%** of 20 Met "high" fidelity
	1D. Near-peer coaches receive adequate support to serve students.	1–2.99 = Low (0) 3–3.49 = Med (1) 3.5–5.0 = High (2) Survey Ratings	*83% of 18 Met "high" fidelity*	*94% of 18 Met "high" fidelity*	**82%** of 22 Met "high" fidelity	**25%** of 20 Did not meet "high" fidelity
	1E. Near-peer coaches share resources and develop effective strategies	disagree = Low (0) neutral = Med (1) agree = High (2) Agreement in *focus group or survey*	*55% of 20 Did not meet "high" fidelity*	*70% of 20 Did not meet "high" fidelity*	**59%** of 22 Did not meet "high" fidelity	**60%** of 20 Did not meet "high" fidelity

	1F. Near-peer coaches use data to guide student intervention	<1 = Low (0) 1 = Med (1) 2> = High (2) Reported monthly frequency	*70% of 20 reported bi-weekly use* *Did not meet "high" fidelity* *Measured by survey item*	*90% of 20 reported bi-weekly use* *Met "high" fidelity* *Measured by survey item*	**77%** of 22 Did not meet "high" fidelity *Measured by focus group or survey*	**85%** of 20 Met "high" fidelity *Measured by focus group or survey*
2. College focused sessions for College Possible high school students	2A. College Possible students placed in peer groups	0–4 or > 30 = Low (0) 25–30 = Med (1) 5–25 = High (2) # Students in peer group	*97% of 721* *Met "high" fidelity*	*93% of 672* *Met "high" fidelity*	96% of 732 Met "high" fidelity	95% of 678 Met "high" fidelity
	2B. College Possible students attend weekly sessions	*At least 25 sessions junior year and 15 sessions senior year = High (1)*	*72% of 721* *Did not meet "high" fidelity*	*75% of 672* *Did not meet "high" fidelity*	68% of 730 *Did not meet "high" fidelity*	66% of 678 Did not meet "high" fidelity
	2C. College Possible students participate in at least 2 one-on-one follow-up check-in sessions per curriculum guide	*None = Low (0)* *1 = Med (1)* *2 = High (2)* *# One-on-one sessions students attended*	*80% of 721* *Did not meet "high" fidelity*	*76% of 672* *Did not meet "high" fidelity*	*78% of 732* *Did not meet "high" fidelity*	69% of 678 Did not meet "high" fidelity
	2D. College Possible students report that near-peer mentoring services provided opportunities to practice and develop non-cognitive skills.	*None = Low (0)* *1 = Med (1)* *>2 = High (2)* *# non-cognitive skills indicated*	*89% of 466* *Met "high" fidelity* *Measured as an agreement scale in SY16–17*	*88% of 310* *Met "high" fidelity* Measured as select all non-cognitive skills indicated in SY17–18	87% of 350 Met "high" fidelity Measured as select all non-cognitive skills indicated in SY17–18	75% of 662 Did not meet "high" fidelity met Measured as attendance at select non-cognitive skills sessions

4.3 Summary of Program Implementation Analysis

The FOI study revealed that while the majority of the ten indicators were delivered with high fidelity during the final year, neither of the two KCs met the threshold for adequate implementation. Challenges with fidelity were more apparent during the first year and were later revised to consider number of sessions rather than percentages, reflecting more of site program scheduling. While these revisions were more representative of program realities, attendance was still not met with fidelity and participation during specific non-cognitive skills targeted sessions was not met with high fidelity, as measured during the final year. Another factor contributing to the low FOI in attendance may include data system changes that occurred in the final year, making it challenging for both coaches and staff to interpret attendance accurately. However, while the two indicators, 2C and 2D, relying on attendance data were not met with high fidelity, students reported positive achievement outcomes overall. Lastly, coaches' attendance at the necessary training events and provision of data as resource 1A, 1B, 1C, 1D were met with high fidelity, but the remaining indicators 1D and 1E related to the quality of support were not met with fidelity. Thus, while the training provided an overall framework and organization for coaches to provide the intervention, the specific needs of coach peer interaction and leadership support were not consistently delivered to coaches during this study.

IV. Discussion and Conclusions

The findings from the REACH score analysis based on the confirmatory question showed the two-year program impact was relatively small (standardized effect of 0.12; statistically significant at p=0.05). While there was some suggestive subgroup variation, the overall program effect size may be related to the issue that FOI that was found lower than intended by the analysis of implementation data. Based on the findings of the implementation data analysis, lack of consistency in student attendance may have contributed to the result.

The subgroup analysis suggested that some subgroups had a slightly larger size of program impact. It is possible that the program impact may be stronger for subgroups that are high achieving. The ideal program effect would be the one that is achieved regardless of student characteristics or their level of achievement. Lack of consistency again may have been related as students more willing to participate in mentoring assistance may be the ones who were already motivated, compared to those who were unwilling.

There are a couple factors that may have contributed to the limited level of student level impact: (1) student peer group participation and (2) relevancy of coach training to local context.

The College Possible model was designed to guide students through all the key aspects of preparing for college during after-school sessions for two hours, twice per week. Over the course of their junior and senior years, students were to complete 320 hours of curriculum in a supportive group of college-bound peers. The junior year high school curriculum was designed to orient students to the college application process, provide extensive preparation for the ACT/SAT exam, introduce students to college life through campus tours, and allow time for students to apply for summer enrichment opportunities. The senior year high school curriculum

led students through the college application process, assisted students in applying for financial aid and scholarships, and guided students through the transition to college.

Since fidelity relies on student engagement in the peer group sessions, it may be challenging for the program to fully implement the program model if there is a lack of consistency in attendance. Coaches recognized the scheduling conflicts that cohort juniors, and particularly cohort senior students, faced and replaced group sessions with one-on-one meet ups. However, this lack of attendance in the peer group sessions could be detrimental for overall program success. Based on participant interviews, students across both cohorts were enthusiastic about their accomplishments as part of the project, but many shared concerns over peer commitment to the program and how attendance was low. Although KC2 was not met with adequate fidelity due to overall student attendance, several focus groups conducted in the first and second year of the study confirmed the active student participants reported that the program exceeded their expectations.

Another factor that may have played a role in program impact was that coaches' preparation and training focused too much on national metrics related to college readiness rather than localized context and individual student planning. College Possible training components—orientation and weekly sessions—were sometimes provided to coaches with limited regard to coach best practices in leveraging school resources and local community efforts to maximize reach. According to coaches, their preparation could be more effective if it capitalized on school and community relationships developed through College Possible presence in the school and implemented collaboratively. For example, while students were challenged by unique considerations of immigration laws, school violence, or parental absence, coaches could work more closely with school and community personnel to support exposure in targeted peer settings that may bring about change in non-cognitive skills—and higher program impact.

In considering areas for program improvement and further study, it is important to note that the small program impact on REACH scores was based on the two-year study intervention during junior and senior school years. If the mentoring was extended to earlier years, it is possible that the program effect, if cumulative, may be larger.

College Possible program leaders may also wish to review program implementation to see how the program effect can be more universal on students of different demographics. The inconsistency of program participation may be an area of improvement. A future study may examine students' program attendance itself as an intermediary outcome to achieve a better predictive understanding of who participates in the program and who does not. Students who are relatively high achieving in school seem to benefit more from the program intervention; however, a future analysis can address the issue of how to approach and involve low achieving students.

Based on the exploratory and descriptive analysis findings, the graduation rates were greater for the College Possible students. This finding is specific to the matched analysis sample and not generalizable to the wider student population. Still, this positive finding is consistent with College Possible's program goal. The study deserves replication based on a more rigorous analysis with refined data collection methods. The area of methodological improvement for a more reliable analysis would involve the data collection approach that reduces the inconsistency between school districts as to the definition of high school graduation. By tracking students from

freshman year and focusing on four-year graduation rates, the analysis will be able to achieve wider variation of graduation outcomes by schools and districts and the findings will be more reliable. Another approach would to study the proxy measures of high school graduation and college aspiration. The survey should include questions related to students' educational aspiration and plans.

V. References

Avery, C. (2013). *Evaluation of the College Possible program: Results from a randomized controlled trial* (Working Paper 19562). Retrieved from National Bureau of Economic Research website at https://www.nber.org/papers/w19562.pdf

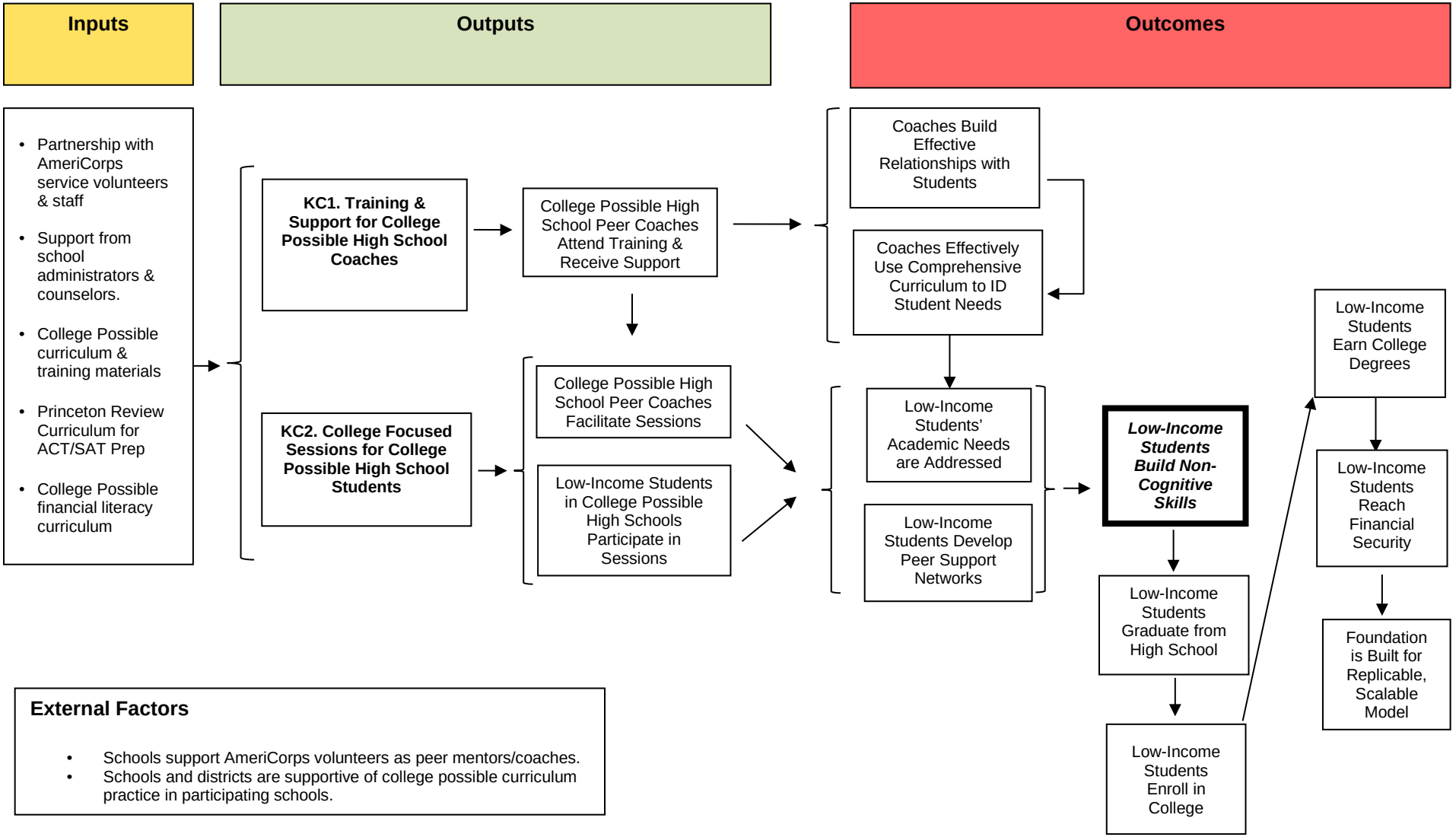
Cohen, Jacob. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

ICF International. (2013). *Evaluation of College Possible postsecondary outcomes, 2007-2012.* Retrieved from College Possible website at http://www.collegepossible.org/wp-content/uploads/2017/01/3-evaluation_of_college_possible_postsecondary_outcomes_2007-2012_final_summary_report.pdf

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

What Works Clearinghouse. (2017). *What Works Clearinghouse Standards Handbook Version 4.0.* Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Appendix A: Logic Model



Appendix B: Impact Study Tables

Table B1. Descriptive Statistics and Baseline Equivalence Analysis Results for the REACH Analysis Sample (Confirmatory Analysis)

	Total sample (n=1250)		Treatment group (n=625)		Comparison group (n=625)		Group Mean Difference		
	Mean	SD	Mean	SD	Mean	SD	Difference	Standardized Difference	WWC Baseline Equivalence Test
Post-test REACH scores	3.75	0.52	3.78	0.53	3.71	0.51	0.07	0.13	N/A
Treatment	0.50	0.50	1.00	0.00	0.00	0.00	1.00	N/A	N/A
Pre-test REACH scores (centered)	0.00	0.53	0.00	0.51	0.00	0.55	0.00	0.01	A
Pre-test REACH scores (original)	3.73	0.53	3.73	0.51	3.73	0.55	0.00	0.01	A
Cohort 2 student	0.43	0.49	0.43	0.50	0.43	0.50	0.00	0.00	A
Baseline GPA (z-score)	0.35	0.76	0.40	0.72	0.30	0.80	0.10	0.13	B
Disadvantage status	0.84	0.37	0.84	0.37	0.84	0.37	0.00	0.00	A
Male	0.41	0.49	0.40	0.49	0.41	0.49	-0.01	-0.01	A
Black	0.27	0.44	0.27	0.44	0.28	0.50	-0.01	-0.05	A
Hispanic	0.27	0.44	0.27	0.44	0.28	0.50	-0.01	-0.05	A
Asian	0.39	0.49	0.39	0.49	0.38	0.50	0.01	0.07	B
Other race group	0.01	0.12	0.01	0.12	0.00	0.10	0.01	1.23	C
White	0.06	0.24	0.06	0.24	0.07	0.30	-0.01	-0.19	A
Columbia Heights	0.06	0.23	0.06	0.23	0.06	0.23	0.00	0.00	A
Milwaukee	0.24	0.43	0.24	0.43	0.24	0.43	0.00	0.00	A
Minneapolis	0.07	0.26	0.07	0.26	0.07	0.26	0.00	0.00	A
Omaha	0.24	0.43	0.24	0.43	0.24	0.43	0.00	0.00	A
Park Rose	0.03	0.16	0.03	0.16	0.03	0.16	0.00	0.00	A
Philadelphia	0.01	0.08	0.01	0.08	0.01	0.08	0.00	0.00	A
St. Paul	0.36	0.48	0.36	0.48	0.36	0.48	0.00	0.00	A

Note: Standardized difference was derived as Hedge's *d* for continuous variables and as Cox's index for binary variables (the same algorithms described in the WWC documentation were used; see https://ies.ed.gov/ncee/wwc/Docs/OnlineTraining/wwc_training_m3.pdf). Final column indicates A if the absolute value of standardized difference was smaller than 0.05, B if smaller than 0.25. WWC states that if B, the final statistical model should include the predictors to adjust for the pre-test differences. If greater than 0.25, the dataset fails a WWC baseline equivalence test. The proportion of other race group received C; however, the difference is trivial in real numbers (2 students in the treatment group; 15 students in the comparison group out of the sample of 1,250) and this is not a crucial variable in the predicative model.

Table B2. Ad-Hoc Analysis: Results of the Hierarchical Linear Modeling (HLM) Analysis for Cohort 1 and Cohort 2 Combined Sample – with Subgroups Defined by Cohort and Treatment Status Explicitly Modeled (n=1250)

Predictors	Estimate	Std. Error	*p*-value	Sig.	Standardized Estimates
Intercept	3.66	0.06	<.0001	***	-0.17
Cohort 1 Comparison	0	.	.	*	0.00
Cohort 1 Treatment	0.06	0.03	0.05	~	0.12
Cohort 2 Comparison	-0.04	0.03	0.19		-0.09
Cohort 2 Treatment	0.02	0.03	0.57		0.04
Pre-test REACH (centered)	0.56	0.02	<.0001	***	1.08
Baseline GPA (z-score)	0.06	0.02	0.00	***	0.11
Disadvantage status	-0.03	0.03	0.33		-0.06
Male	0.02	0.02	0.32		0.05
Black	0.17	0.06	0.00	**	0.33
Hispanic	0.08	0.06	0.15		0.16
Asian	0.04	0.06	0.45		0.08
Other race group	0.08	0.11	0.49		0.15
Columbia Heights	0.11	0.07	0.11		0.21
Milwaukee	-0.05	0.04	0.27		-0.09
Minneapolis	-0.01	0.06	0.92		-0.01
Omaha	-0.02	0.04	0.66		-0.04
Park Rose	0.13	0.08	0.11		0.26
Philadelphia	0.13	0.15	0.39		0.25

Notes: Statistical significance (2-tailed):~ if $p<.10$, *$p<.05$, ** $p<.01$, *** $p<.001$.

Table B3. Descriptive Statistics and Baseline Equivalence Analysis Results for the High School Outcome Graduation Sample (Exploratory Analysis)

	Total Sample (n=2048)		Treatment Group (n=1024)		Comparison Group (n=1024)		Group Mean Difference		
Variables	Mean	SD	Mean	SD	Mean	SD		Standardized Difference	WWC Baseline Equivalence test
Graduated	0.95	0.21	0.98	0.13	0.93	0.26	0.05	0.91	N/A
Treatment	0.50	0.50	1.00	0.00	0.00	0.00	0.50	N/A	N/A
Pre-test REACH scores (original)	3.73	0.53	3.74	0.52	3.73	0.55	0.01	0.01	A
Cohort 2 student	0.50	0.50	0.50	0.50	0.50	0.50	0.00	0.00	A
Baseline GPA (z-score)	0.00	1.00	0.03	0.98	-0.03	1.02	0.06	0.06	B
Disadvantage status	0.83	0.37	0.84	0.36	0.83	0.38	0.01	0.07	B
Male	0.41	0.49	0.4	0.49	0.41	0.49	-0.01	-0.03	A
Black	0.31	0.46	0.31	0.46	0.32	0.46	-0.01	-0.02	A
Hispanic	0.27	0.45	0.27	0.44	0.28	0.45	-0.01	-0.05	A
Asian	0.35	0.48	0.36	0.48	0.34	0.47	0.02	0.06	B
Other race group	0.01	0.11	0.02	0.13	0.00	0.07	0.02	0.78	C
White	0.05	0.23	0.05	0.22	0.06	0.24	-0.01	-0.14	B
Columbia Heights	0.06	0.24	0.06	0.24	0.06	0.24	0.00	0.00	A
Milwaukee	0.25	0.43	0.25	0.43	0.25	0.43	0.00	0.00	A
Minneapolis	0.08	0.28	0.08	0.28	0.08	0.28	0.00	0.00	A
Omaha	0.23	0.42	0.23	0.42	0.23	0.42	0.00	0.00	A
Park Rose	0.02	0.14	0.02	0.14	0.02	0.14	0.00	0.00	A
Philadelphia	0.02	0.13	0.02	0.13	0.02	0.13	0.00	0.00	A
St. Paul	0.34	0.47	0.34	0.47	0.34	0.47	0.00	0.00	A

Note: Standardized difference was derived as Hedge's d for continuous variables and as Cox's index for binary variables (the same algorithms described in the WWC documentation were used; see https://ies.ed.gov/ncee/wwc/Docs/OnlineTraining/wwc_training_m3.pdf). Final column indicates A if the absolute value of standardized difference was smaller than 0.05, B if smaller than 0.25. WWC states that if B, the final statistical model should include the predictors to adjust for the pre-test differences. If greater than 0.25, the dataset fails a WWC baseline equivalence test. The proportion of other race group received C; however, the difference is trivial in real numbers (5 students in the treatment group; 18 students in the comparison group out of the sample of 2,048) and this is not a crucial variable in the predicative model.

Table B4. REACH Survey Items for College Possible Study (item n.=28)

Question	Response Options
How much do you agree or disagree with the following?	Strongly disagree Disagree Somewhat Disagree Somewhat Agree Agree Strongly Agree
I can get smarter by working hard.	
I am inspired by the example the adults in my school set for me.	
I work hard on all assignments even if they won't affect my grade.	
I am certain I can master the skills taught in school this year.	
How much are the following like you?	Not at all like me A little like me Somewhat like me Mostly like me Very much like me
I can focus my attention on the things that matter.	
I organize my time so I can focus on school.	
I stay positive even when I'm facing challenges.	
I try to develop my interests and talents by practicing and working on them.	
I have interests and talents that I really enjoy spending my time on	
How much do you agree or disagree with the following?	Strongly disagree Disagree Somewhat Disagree Somewhat Agree Agree Strongly Agree
If I make a plan, I can usually make it work out.	
My main reason for working hard in school is to learn new knowledge and skills.	
How well I do in school depends more on how hard I work than on how naturally smart I am.	
The adults in my school and I respect one another's point of view	
How often do you talk about your interests and talents with … --	Never About once a month About 2-3 times a month Almost once a week More than once a week
… other students in your classes?	
… your parents or guardians?	
… a teacher or other adult at your school (such as a coach, counselor)?	
How well do each of the following describe the adults in your school?	Not at all like the adults in my school A little like … Somewhat like … Mostly like … Very much like …
The adults in my school treat me like I am important and valued.	
The adults in my school have high expectations for me.	
The adults in my school help me learn from my mistakes.	
The adults in my school value my ideas and opinions.	
The adults in my school give me useful advice.	
The adults in my school try to find out what I am interested in.	
Think about a time when you did something really challenging when working on something that interest you. With that experience in mind, how true is the following statement?	Not at all true A little true Somewhat true Mostly true Very true
It gave me confidence that I could tackle other hard challenges	
How much are the following like you?	Not at all like me A little like me
I set goals that are actually possible for me to reach.	

I have plans for my future.	Somewhat like me Mostly like me Very much like me
I don't give up when things get tough.	
I can list specific goals I want to achieve in the next year.	
I try to imagine the person I want to be in the future.	

Appendix C: Implementation Study Tables

Table C1. Reporting Sample-Level Component Fidelity Scores by Key Component (KC)

Key Components in the Logic Model	Definition of High Implementation	Definition of "implementation with fidelity" at program level	Cohort 1 Reporting Year 1 (2017-18) On Data Collected 2016-17 and 2017-18		Cohort 2 Reporting Year 2 (2018-19) On Data Collected 2017-18 and 2018-19	
			No. of Years with "high" implementation	Implementation with fidelity for Reporting year	No. of Years with "high" implementation	Implementation with fidelity for Reporting year
KC1 – Training and support of CP coaches	Calculation based on 6 indicators (#1A through #1F) at the end of each year of implementation	81 – 100% of coaches have "high" implementation at the end of each year	2 = high	N/A	1 = low	No = low
KC2 – College focused sessions for CP high school students	Calculation based on 4 indicators (#2A through #2D) at the end of each year of implementation	81 – 100% of students received college-focused sessions with "high" implementation at the end of each year	0 = low	N/A	0 = low	No = low

Table C2. Fidelity of Implementation Indicators and Scores for Key Component 1 (KC1)

Indicators	Definition	Unit of implem-entation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
Key Component 1 (KC1) : Training and support for CP High School Coaches										
#1A. Coaches are assigned a caseload of Grade 11 students.	Each of the 19 near peer coaches will be assigned between 20 and 40 students	Coach	Administrative records maintained by CP	CP staff will submit to ICF on December 1 of each year	# of Students assigned to Coaches 0 (low) = 0-9 or >40 1 (moderate) = 10-19 students 2 (high) = 20-40 students	High implementation at coach level = score of "2"				
#1B. Near peer coaches receive annual two-part training orientation	Coaches will receive required two-part training orientation	Coach	Attendance records collected at training by CP staff	CP staff will submit to ICF on December 1 of each year	# of Trainings received by Coaches 0 (low) = receive no training 1 (med) – receive 1 part training 2 (high)= receive 2 part training	High implementation at coach level = score of "2"				
#1C. Near peer coaches attend weekly meetings with their colleagues	Coaches attend 30 weekly meetings with other coaches during school-year	Coach	CP staff will collect weekly sign in sheets	CP staff will submit to ICF on July 30 of each year	# Sessions attended for coach 0 (low) = 0 - 18 meetings 1 (moderate) = 19 – 24 meetings 2 (high) = 25 or more meetings	High implementation at coach level = score of "2"				
#1D. Near peer coaches receive adequate	Coaches report training adequately prepared	Coach	CP staff will use a coach feedback survey	CP staff will submit to ICF on July 30 of each year	Coach either agrees or strongly agrees (4 or 5) that training provided adequate support 0 (low) = 1-2	High implementation at coach level = score of "2"				

Indicators	Definition	Unit of implem-entation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measure ment
support to serve students.	them to serve students.		collected annually		1 (moderate) = 3 2 (high) = 4-5					
#1E. Near peer coaches share resources and develop effective strategies	Coaches report meaningful interaction with their colleagues during weekly meetings	Coach	CP staff will use a coach feedback survey collected annually	CP staff will submit to ICF on July 30 of each year	Coach either agrees or strongly agrees (4 or 5) that meeting provided opportunities for meaningful interaction 0 (low) = 1-2 1 (moderate) = 3 2 (high) = 4-5	High implementation at coach level = score of "2"				
#1F. Near peer coaches use data to guide student intervention	Coaches report using data bi-weekly to guide student interventions	Coach	CP staff will use a coach feedback survey collected annually	CP staff will submit to ICF on July 30 of each year	# times coaches reported data use 0 (low) = Less than once per month 1 (moderate) = Once per month 2 (high) = Twice or more per month	High implementation at coach level = score of "2"				
All indicators Component 1 Scoring (Year 1)					Year 1: Component Indicator & Implementation Score <u>Indicator Implementation</u> < 8 0 = Low 8 - 10 1 = Med 11-12 2 = High	Coach-level: "High" implementation score = 12		Program-level: At least 81% of near-peer coaches in sample have "High" implementation	All coaches employed to implement intervention	2016-17 (Cohort 1) 2017-18 (Cohort 2)
All indicators Component 1 Scoring					Year 2: Component Indicator & Implementation Score	Coach-level: "High" implementation score = 12		Program-level: At least 81% of near-peer coaches in	All coaches employed to	2017-18 (Cohort 1), 2018-19 (Cohort 2)

Indicators	Definition	Unit of implem-entation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measure ment
(Year 2)					Indicator Implementation <8 0 = Low 8 - 10 1 = Med 11 - 12 2 = High			sample have "High" implementation	implement intervention	

Table C3. Fidelity of Implementation Indicators and Scores for Key Component 2 (KC2)

Indicators	Definition	Unit of implem-entation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measure ment
Key Component 2 (KC2): College Focused Sessions For CP High School Students										
#2A. CP students placed in peer groups	CP students placed in college-focused peer groups with between 10-15 fellow mentees and one coach.	Student	CP will query data for roster by school	CP staff will submit to ICF on December 1 of each year	Students placed in peer group of: 0 (low) = 0-4 or >20 peers 1 (high) = 5-20 peers	High implementation at student level = score of "1"				
# 2B. CP students attend weekly sessions	CP students attend 60, two-hour weekly mentoring sessions with coach/mentor and fellow mentees per year	Student	CP will provide an attendance tracking spreadsheet	CP staff will submit to ICF on July 30 of each year	Students attend sessions 0 (low) = <25 sessions 1 (moderate) = 25-50 sessions 2 (high) = >50 sessions	High implementation at student level = score of "2"				
#2C. CP students participate in one-on-one sessions	CP students participate in at least 2 one-on-one follow-up check-in	Student	CP will provide an attendance tracking spreadsheet	CP staff will submit to ICF on July 30 of each year	Students participate in one-on-one follow-up sessions 0 (low) – No sessions 1 (moderate) – 1 session 2 (high) – 2 sessions	High implementation at student level = score of "2"				

Indicators	Definition	Unit of implem-entation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measure ment
	sessions per curriculum guide									
#2D. Students receive opportunities to practice non-cognitive skill building.	CP students report that near-peer mentoring services provided opportunities to practice and develop non-cognitive skills.	Student	CP conducts an annual survey, ICF will add items to it annually	CP staff will submit to ICF on July 30 of each year	Student either agrees or strongly agrees (4 or 5) that opportunities to practice and develop non-cognitive skills are provided 0 (low) = 1-2 1 (moderate) = 3 2 (high) = 4-5	High implementation at student level = score of "2"				
All indicators Component 2 Scoring (Year 1)					Year 1: Component Indicator & Implementation Score <u>Indicator</u> <u>Implementation</u> < 4 0 = Low 5 1 = Med 6 - 7 2 = High	Student-level: "High" implementation score = 7		Program-level: At least 81% of students in sample receive college-focused sessions with "High" implementation	All students in which intervention is being implemented	2016-17 (Cohort 1) 2017-18 (Cohort 2)

Indicators	Definition	Unit of implem-entation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for adequate implementation at unit level	Roll-up to next higher level if needed (score and threshold): Indicate level	Roll-up to program level (score and threshold for adequate implementation at sample level)	Expected sample for fidelity measure	Expected years of fidelity measure ment
All indicators Component 2 Scoring (Year 2)					Year 1: Component Indicator & Implementation Score Indicator Implementation < 4 0 = Low 5 1 = Med 6 - 7 2 = High	Student-level: "High" implementation score = 7		Program-level: At least 81% of students in sample receive college-focused sessions with "High" implementation	All students in which interventio n is being implement ed	2017-18 (Cohort 1), 2018-19 (Cohort 2)