Coaching Corps Fitness Evaluation Report
2010-13
Prepared by UC Berkeley August 2013

This report contains an analysis of three years of fitness data, 2010-11 (Year 1), 2011-12 (Year 2), and 2012-13 (Year 3) from Coaching Corps' evaluation of its AmeriCorps program. Coaching Corps coaches collected and aggregated all data. UCSF/ UC Berkeley provided evaluation support and are responsible for all data analysis and reporting. Select results are presented below.

Highlights

- The proportion of students with complete data (data from all 3 yearly data collection times) steadily increased each year (from 31% to 45% to 49%), demonstrating that Coaching Corps improved overall student retention and streamlined and improved their data collection and evaluation procedures across years.
- The proportion of female participants increased from Year 1 (40%) to Year 3 (47%). In addition, the number of participants who are younger than ten steadily increased, from 56% in Year 1, to 67% in Year 2, to 75% in Year 3.
- Average number of days of student attendance increased from Year 1 (39 days) to Year 3 (47 days).
- Across all 3 years, percent increase in students' fitness scores (including all sub-groups of students, broken down by sex and high-need status) ranged from 25% to 600%. These large increases in fitness scores far exceed the 5% change we'd expect to see from maturation alone.
- Across all 3 years, nearly 50% of students who were below the Healthy Fitness Zone at Time 1 moved into the Healthy Fitness Zone by Time 3.
- On average, students with complete data from Years 1 and 2 and Years 2 and 3, increased their average fitness scores across years. This suggests that Coaching Corps may be a sustainable program that has the potential to impact students' fitness over multiple years.
- Whereas we found a positive statistically significant relationship between attendance and change in student fitness scores in Year 1 (on average, for each additional ten days of attendance, the number of laps students could run from the beginning of the year to the end of the year increased by 2 laps), we did not find a statistically significant positive relationship between attendance and change in fitness scores in Years 2 or 3.

Recommendations:
- Shift evaluation focus from time-intensive individual testing to class-level observations, utilizing the System of Observing Fitness Instruction Time (SOFIT) tool. Additionally, incorporate coach interviews/surveys to gather qualitative data on program functioning and coach experiences to use towards program improvement.
- If continuing individual-level testing, in the absence of funder requirements, eliminate the use of the curl-up to assess student fitness levels, as this test is not best suited to measure changes in fitness we'd expect to see as a result of participating in Coaching Corps programming.
- Continue to focus attention on data quality, ensuring data are accurately collected and entered in.
- Coaching Corps may want to consider targeting older high-need students (especially girls) for program participation and work to ensure they remain in the program, as these are groups that diminished program participation across the evaluation period.

The Program

Coaching Corps coaches, trained using Coaching Corps' Building Blocks Curriculum and a comprehensive coach curriculum aimed at increasing fitness levels, work in after-school programs with students from low-income communities across California.

Year 1: Coaching Corps coaches worked with 1,383 students between 12 and 20 hours per week in after-school programs in Los Angeles (26 coaches), the San Francisco Bay Area (24 coaches), Fresno (6 coaches) and San Diego (3 coaches).

Year 2: Coaching Corps coaches worked with 1,571 students between 12 and 20 hours per week in after-school programs in Los Angeles (10 coaches) and the San Francisco Bay Area (20 coaches).

Year 3: Coaching Corps coaches worked with 970 students between 12 and 20 hours per week in after-school programs in Los Angeles (12 coaches) and the San Francisco Bay Area (15 coaches).

Methods

Coaching Corps coaches used the Progressive Aerobic Cardiovascular Endurance Run (PACER) to test the cardio-respiratory health of program participants and the curl-up test to evaluate participants' abdominal strength. Both the PACER test and the curl-up test are subsets of the FitnessGram, which uses "Healthy Fitness Zone" standards that represent levels of fitness that offer some degree of protection against diseases that can result from sedentary living. Tests were performed at up to three points throughout each school year: in the fall (Time 1), in the winter (Time 2), and in the spring (Time 3). Fitness tests (PACER, curl-up, or both) were chosen each year based on Coaching Corps' funders' requirements and "high-need" students were determined as follows:

Year 1: Students ran the PACER test to assess fitness. Students younger than ten were designated as high-need if their score was in the bottom 40% of scores at Time 1. Students ten or older were designated as high-need if their score was less than 75% of the Healthy Fitness Zone standard based on the student's age and sex.

Year 2: Students who were younger than ten did the curl-up test to assess fitness and students who were ten years old or older ran the PACER test. Students younger than ten were designated as high-need if their score was less than the Healthy Fitness Zone standard for curl-ups based on the student's age and sex. As in Year 1, students ten or older were designated as high-need if their score was less than 75% of the Healthy Fitness Zone.

Year 3: All students ran the PACER test and did the curl-up test to assess fitness. As in Year 2, students younger than ten were designated high-need if their score was less than the Healthy Fitness Zone standard for curl-ups based on the student's age and sex. Students ten or older were designated as high-need if their score was less than or equal to 75% of the Healthy Fitness Zone.

Table: PACER and Curl-Up Healthy Fitness Zone and High-Need Cut-points

Age	PACER Healthy Fitness Zone Standards*		Cut-points for high-need students
	Min laps	**Max laps**	**75% of min**
FEMALES			
10	15	41	11
11	15	41	11
12	23	41	17
13	23	51	17
14	23	51	17
15	23	51	17
16	32	61	24
17	41	61	31
17+	41	61	31
MALES			
10	23	61	17
11	23	72	17
12	32	72	24
13	41	72	30
14	41	83	30
15	51	94	38
16	61	94	45
17	61	94	45
17+	61	94	45

* These Healthy Fitness Zone cut-points are based on the FITNESSGRAM Healthy Fitness Zone Standards used from 1999-2005.

Age	Curl-Up Healthy Fitness Zone Standards*		Cut-points for high-need students
	Min curl-ups	**Max curl-ups**	**< Min curl-ups**
FEMALES			
5	2	10	1
6	2	10	1
7	4	14	3
8	6	20	5
9	9	22	8
MALES			
5	2	10	1
6	2	10	1
7	4	14	3
8	6	20	5
9	9	24	8

* These Healthy Fitness Zone cut-points are based on the FITNESSGRAM Healthy Fitness Zone Standards used from 1999-2012

In **Year 1**, 1,383 students were evaluated at Time 1, of whom 429 had complete data (data at all three time points).

In **Year 2**, 1,571 students were evaluated at Time 1; 705 had complete data.

In **Year 3**, 970 students were evaluated at Time 1; 475 had complete PACER data and 472 had complete curl-up data; 471 had complete data for both assessments.

In **Years 1 and 2**, student attrition was not related to student sex (boys were just as likely as girls to have complete data), but students who were older than 10 and students who were high-need were more likely to have complete data than students were younger than ten or who were not high-need. In **Year 3**, student attrition was not related to age, sex, or high-need status, with similar proportions of students younger than ten and students ten or older, girls and boys, and non-high-need/ high-need students having complete data.

Data presented in this report are limited to students with complete data

All students, by age:
　　Year 1:
- 241 of 429 students (56%) were younger than ten.
- 188 of 429 students (44%) were older than 10.

　　Year 2:
- 470 of 705 students (67%) were younger than ten.
- 235 of 705 students (33%) were older than 10.

　　Year 3:
- 352 of 471 students (75%) were younger than ten.
- 119 of 471 students (25%) were older than 10.

High-need students (defined at Time 1):
　　Year 1:
- 217 out of 429 students (51%) were high-need
- 127 of 241 students (53%) were high-need and younger than ten
- 90 of 188 students (48%) were high-need and ten or older.

　　Year 2:
- 460 out of 705 students (65%) were high-need
- 331 of 470 students (70%) were high-need and younger than ten
- 129 of 235 students (55%) were high-need and ten or older

　　Year 3:
- 327 out of 471 students (69%) were high-need
- 283 of 352 students (80%) were high-need and younger than ten
- 44 of 119 students (37%) were high-need and ten or older

Coaching Corps saw changes in the overall number of students who participated in the program over the 3-year evaluation period, with the largest number participating in Year 2 and the smallest number of students participating in Year 3. The decrease in number of participants Year 3 is most likely due to an almost 50% reduction in the number of Coaching Corps sites. Despite the drop in overall students participating between Years 2 and 3, the proportion of high-need students increased, suggesting Coaching Corps continued to do a good job recruiting and retaining this high-need population. Further, the proportion of students with complete data steadily increased across all 3 years (from 31% in Year 1, to 45% in Year 2, and 49% in Year 3), demonstrating that Coaching Corps improved overall student retention and continuously streamlined and improved their data collection and evaluation procedures.

Table 1: Average fitness scores and changes in fitness scores[A,B]

		All students[C]					High-need students[D]				
		Time 1	Time 2	Time 3	Δ 1 to 3	p-value	Time 1	Time 2	Time 3	Δ 1 to 3	p-value
Year 1	PACER (age < 10)	16	18	24	+ 7	<0.001	9	13	16	+ 7	<0.001
	PACER (age ≥ 10)	25	30	32	+ 7	<0.001	11	19	24	+ 14	<0.001
Year 2	Curl-Up (age < 10)	6	7	9	+ 3	<0.001	1	5	7	+ 6	<0.001
	PACER (age ≥ 10)	19	30	33	+ 14	<0.001	9	22	27	+ 18	<0.001
Year 3	Curl-Up (age < 10)	3	5	6	+ 3	<0.001	1	4	5	+ 4	<0.001
	Curl-Up (age ≥ 10)	13	19	20	+ 7	<0.001	8	13	18	+ 10	<0.001
	PACER (age < 10)	16	17	22	+ 6	<0.001	15	17	21	+ 6	<0.009
	PACER (age ≥ 10)	24	25	30	+ 6	<0.001	12	15	22	+ 10	<0.001

[A] Numbers may not add up perfectly, due to rounding. P-values are for the change in laps between Times 1, 2, and 3 and were calculated using linear mixed effects models with random effects for site and student. A p-value less that 0.05 is considered statistically significant and means that observing a change in laps as extreme or more extreme than the change observed by chance alone, given that there really is no change in laps, is very unlikely.

[B] Numbers for the PACER test represent number of laps run.

[C] During Year 1 there were 241 students were younger than ten and 188 students were ten or older. During Year 2 there were 470 students were under ten and 235 students were ten or older. During Year 3 there were 352 students were younger than ten and 119 students were ten or older.

[D] High-need denotes students who were high-need at Time 1. During Year 1 there were 127 students younger than ten and 90 students ten or older were high-need. During Year 2 there were 331 students younger than ten and 129 students ten or older were high-need. During Year 3 there were 283 students younger than ten and 44 students ten or older were high-need.

During all 3 Years, all students and high-need students significantly improved their fitness scores from Time 1 to Time 3.

Year 1:

From Time 1 to Time 3, students improved their PACER scores by:

- **All Students**:
    - *Younger than ten:* 44%
    - *Ten and older:* 28%
- **High-need students**:
    - *Younger than ten:* 78%
    - *Ten and older:* 46%

Year 2:

From Time 1 to Time 3, students improved their scores by:

- **All Students**:
    - *Younger than ten:* 50% (curl-up)
    - *Ten and older:* 74% (PACER)
- **High-need students**:
    - *Younger than ten:* 600% (curl-up)
    - *Ten and older:* 200% (PACER)

Year 3:

From Time 1 to Time 3, students improved their scores by:

- **All Students**:
    - *Younger than ten:* 100% (curl-up)
    - *Ten and older:* 54% (curl-up)
    - *Younger than ten:* 38% (PACER)
    - *Ten and older:* 25% (PACER)

- **High-need students**:
  - *Younger than ten:* 400% (curl-up)
  - *Ten and older:* 125% (curl-up)
  - *Younger than ten:* 40% (PACER)
  - *Ten and older:* 83% (PACER)

These changes, particularly among the high-need students, represent large changes in fitness and far exceed the 5% change we'd expect to see from maturation alone. Some of the improvement likely reflects students' learning to take the tests. The greater improvements seen in Year 2 may be an artifact of test administration or be related to Coaching Corps' increased focus on recruiting and working with high-need students (who started out with lower Time 1 scores and therefore had more room to improve). While the improvements are encouraging, it is difficult to attribute them solely to the presence of Coaching Corps without data from a control group for comparison.

Table 2: Changes in fitness scores for all students, by gender[A]

		All girls[B]					All boys[B]				
		Time 1	Time 2	Time 3	Δ 1 to 3	p-value	Time 1	Time 2	Time 3	Δ 1 to 3	p-value
Year 1	PACER (age < 10)	13	15	19	+ 6	<0.001	18	21	27	+ 9	<0.001
	PACER (age ≥ 10)	19	23	27	+ 8	<0.001	29	35	36	+ 7	0.002
Year 2	Curl-Up (age < 10)	5	6	8	+ 3	<0.001	7	8	9	+ 2	0.011
	PACER (age ≥ 10)	19	28	31	+ 12	<0.001	18	32	35	+ 17	<0.001
Year 3	Curl-Up (age < 10)	3	5	6	+ 3	<0.001	2	4	5	+ 3	<0.001
	Curl-Up (age ≥ 10)	12	19	23	+ 11	<0.001	14	19	18	+ 4	<0.001
	PACER (age < 10)	14	16	20	+ 6	<0.001	17	18	23	+ 6	<0.001
	PACER (age ≥ 10)	23	24	28	+ 5	<0.001	25	27	33	+ 8	<0.001
		High-need girls[C]					High-need boys[C]				
		Time 1	Time 2	Time 3	Δ 1 to 3	p-value	Time 1	Time 2	Time 3	Δ 1 to 3	p-value
Year 1	PACER (age < 10)	9	12	16	+ 7	<0.001	9	14	17	+ 8	<0.001
	PACER (age ≥ 10)	8	12	18	+ 9	<0.001	12	24	29	+ 17	<0.001
Year 2	Curl-Up (age < 10)	1	5	7	+ 6	<0.001	1	6	7	+ 6	<0.001
	PACER (age ≥ 10)	8	17	20	+ 12	<0.001	10	25	32	+ 22	<0.001
Year 3	Curl-Up (age < 10)	2	5	6	+ 4	<0.001	1	4	5	+ 4	<0.001
	Curl-Up (age ≥ 10)	8	13	24	+ 16	<0.008	8	13	16	+ 8	<0.001
	PACER (age < 10)	14	16	20	+ 6	<0.001	16	17	22	+ 6	<0.001
	PACER (age ≥ 10)	9	12	18	+ 9	0.237	13	16	23	+ 10	0.170

[A] Numbers may not add up perfectly, due to rounding. P-values are for the change in laps between Times 1, 2, and 3 and were calculated using linear regression models accounting for clustering by site and student. A p-value less that 0.05 is considered statistically significant and means that observing a change in laps as extreme or more extreme than the change observed by chance alone, given that there really is no change in laps, is very unlikely.

[B] During Year 1 there were 241 students (97 girls and 147 boys) who were younger than ten and 188 students (76 girls and 112 boys) who were ten or older. During Year 2 there were 470 students (237 girls and 233 boys) who were under ten and 235 students (127 boys and 128 girls) who were ten or older. During Year 3 there were 352 students (169 girls and 183 boys) who were younger than ten and 119 students (53 girls and 66 boys) who were ten or older.

[C] High-need denotes students who were high-need at Time 1. During Year 1 there were 127 students (59 girls and 68 boys) who were younger than ten and high-need at Time 1 and there were 90 students (39 girls and 51 boys) who were ten or older and high-need at Time 1. During Year 2 there were 331 students (172 girls and 159 boys) who were younger than ten and high-need at Time 1 and there were 129 students (56 girls and 73 boys) who were ten or older and high-need at Time 1. During Year 3 there were 283 students (130 girls and 153 boys) who were younger than ten and high-need at Time 1 and 44 students (13 girls, 31 boys) who were ten or older and high-need at Time 1.

Across all 3 years, and across all age groups and sexes, both all students and high-need students significantly improved their fitness scores from Time 1 to Time 3, with the exception of boys and girls over the age of 10 who ran the PACER in Year 3. The fact that the p-values for change in PACER laps for high-need boys and girls in Year 3 were not statistically significant is most likely due to the small number of students in these categories (31 boys and 13 girls).

Year 1:

From Time 1 to Time 3, students improved their scores by:

- **All Students, Girls**:
    - *Younger than ten:* 46%
    - *Ten and older:* 42%
- **All Students, Boys**:
    - *Younger than ten:* 50%
    - *Ten and older:* 24%
- **High-need students, Girls**:
    - *Younger than ten:* 78%
    - *Ten and older:* 113%
- **High-need students, Boys**:
    - *Younger than ten:* 89%
    - *Ten and older:* 141%

Year 2:
From Time 1 to Time 3, students improved their scores by:

- **All Students, Girls**:
    - *Younger than ten: 60%* (curl-up)
    - *Ten and older:* 63% (PACER)
- **All Students, Boys**:
    - *Younger than ten: 89%* (curl-up)
    - *Ten and older:* 94% (PACER)
- **High-need students, Girls**:
    - *Younger than ten: 600%* (curl-up)
    - *Ten and older: 150%* (PACER)
- **High-need students, Boys**:
    - *Younger than ten: 600%* (curl-up)
    - *Ten and older:* 220% (PACER)

Year 3:
From Time 1 to Time 3, students improved their scores by:

- **All Students, Girls**:
    - *Younger than ten:* 100% (curl-up)
    - *Ten and older:* 92% (curl-up)
    - *Younger than ten:* 43% (PACER)
    - *Ten and older:* 22% (PACER)
- **All Students, Boys**:
    - *Younger than ten:* 150% (curl-up)
    - *Ten and older:* 29% (curl-up)
    - *Younger than ten:* 35% (PACER)
    - *Ten and older:* 32% (PACER)
- **High-need students, Girls**:
    - *Younger than ten:* 200% (curl-up)
    - *Ten and older:* 200% (curl-up)
    - *Younger than ten:* 43% (PACER)

- o *Ten and older:* 100% (PACER)
- **High-need students, Boys**:
    - o *Younger than ten:* 400% (curl-up)
    - o *Ten and older:* 100% (curl-up)
    - o *Younger than ten:* 38% (PACER)
    - o *Ten and older:* 77% (PACER)

In Year 1, when looking at all students, girls demonstrated greater relative improvement in their fitness scores as compared to boys. This could be because girls started out with lower scores than boys, so therefore had greater room to grow. In Year 2, boys and girls started out with similar fitness test scores. During this year, boys showed a greater relative improvement in their scores as compared to girls. In Year 3, boys and girls started out with similar fitness scores at Time 1 and demonstrated similar relative improvements in scores, with a few exceptions. For example, both high-need boys and girls older than 10 started out at Time 1 able to do 8 curl-ups and by Time 3, high-need girls improved by 16 curl-ups (200% improvement) whereas high-need boys only improved by 8 curl-ups (100% improvement).

While the relative percent increase in curl-ups for the high-need students in Years 2 and 3 is large, students did not demonstrate a substantial absolute increase in curl-ups. The large relative improvements in the curl-up scores during Years 2 and 3 (especially for students younger than ten) could be due, in large part, to test learning, as many students did not know how to do a curl-up when the test was first administered at Time 1. The fact that these absolute and relative changes are not consistent across the 3 evaluation years signifies that these changes in scores by gender may not be attributable solely to the Coaching Corps program.

Table 3: Average fitness scores and changes in fitness scores for students with complete Year 1 and Year 2 data[A]

	All students								
	Year 1			Year 2			Change		
	Time 1	Time 2	Time 3	Time 4 [B]	Time 5 [B]	Time 6 [B]	Δ 1 to 3	Δ 1 to 6	p-value
PACER (age < 10)	13 (n=59)	13 (n=59)	15 (n=59)	5 (n=2)	6 (n=2)	18 (n=59)	+ 3 (n=59)	+ 5 (n=59)	<0.001
PACER (age ≥ 10)	17 (n=13)	28 (n=13)	28 (n=13)	19 (n=13)	31 (n=13)	41 (n=13)	+ 11 (n=13)	+ 24 (n=13)	<0.001
	All girls								
	Time 1	Time 2	Time 3	Time 4 [B]	Time 5 [B]	Time 6 [B]	Δ 1 to 3	Δ 1 to 6	p-value
PACER (age < 10)	12 (n=32)	13 (n=32)	15 (n=32)	5 (n=2)	6 (n=2)	20 (n=32)	+ 3 (n=32)	+ 8 (n=32)	<0.001
PACER (age ≥ 10)	18 (n=8)	29 (n=8)	28 (n=8)	22 (n=8)	33 (n=8)	45 (n=8)	+ 10 (n=8)	+ 27 (n=8)	<0.001
	All boys								
	Time 1	Time 2	Time 3	Time 4 [B]	Time 5 [B]	Time 6 [B]	Δ 1 to 3	Δ 1 to 6	p-value
PACER (age < 10)	13 (n=27)	13 (n=27)	16 (n=27)	- (n=0)	- (n=0)	16 (n=27)	+ 3 (n=27)	+ 3 (n=27)	0.180
PACER (age ≥ 10)	15 (n=5)	26 (n=5)	29 (n=5)	13 (n=5)	29 (n=5)	35 (n=5)	+ 14 (n=5)	+ 20 (n=5)	0.100
	All high-need students								
	Time 1	Time 2	Time 3	Time 4 [B]	Time 5 [B]	Time 6 [B]	Δ 1 to 3	Δ 1 to 6	p-value
PACER (age < 10)	9 (n=34)	12 (n=34)	13 (n=34)	5 (n=2)	6 (n=2)	14 (n=34)	+ 4 (n=34)	+ 5 (n=34)	0.001
PACER (age ≥ 10)	11 (n=8)	18 (n=8)	20 (n=8)	12 (n=8)	23 (n=8)	28 (n=8)	+ 9 (n=8)	+ 17 (n=8)	0.012
	High-need girls								
	Time 1	Time 2	Time 3	Time 4 [B]	Time 5 [B]	Time 6 [B]	Δ 1 to 3	Δ 1 to 6	p-value
PACER (age < 10)	9 (n=20)	11 (n=20)	13 (n=20)	5 (n=2)	6 (n=2)	15 (n=20)	+ 4 (n=20)	+ 6 (n=20)	<0.001
PACER (age ≥ 10)	9 (n=4)	12 (n=4)	12 (n=4)	10 (n=4)	14 (n=4)	22 (n=4)	+ 3 (n=4)	+ 13 (n=4)	0.015
	High-need boys								
	Time 1	Time 2	Time 3	Time 4 [B]	Time 5 [B]	Time 6 [B]	Δ 1 to 3	Δ 1 to 6	p-value
PACER (age < 10)	9 (n=14)	13 (n=14)	13 (n=14)	- (n=0)	- (n=0)	13 (n=14)	+ 4 (n=14)	+ 3 (n=14)	0.335
PACER (age ≥ 10)	14 (n=4)	24 (n=4)	28 (n=4)	13 (n=4)	32 (n=4)	35 (n=4)	+ 15 (n=4)	+ 21 (n=4)	0.097

[A] n=number of students. Numbers may not add up perfectly, due to rounding. P-values are for the change in laps between all 6 times and were calculated using linear regression models accounting for clustering by site and student. A p-value less that 0.05 is considered statistically significant and means that observing a change in laps as extreme or more extreme than the change observed by chance alone, given that there really is no change in laps, is very unlikely.

[B] Time 4 is the same as Year 2, Time 1; Time 5 is the same as Year 2, Time 2; Time 6 is the same as Year 2, Time 3

There were 72 students with data from both Years 1 and 2. Of these students, 59 were younger than ten and 13 were ten or older. From Time 1 to Time 6, all students, all girls, all high-need students, and all high-need girls significantly improved the number of laps they could run. All boys and all high-need boys increased the number of laps they could run from both Time 1 to Time 3 and Time 1 to Time 6, but the increases were not statistically significant. This could be in part because the sample is very small. Because this subsample of 72 students represents only 5% of the students who were tested during Year 1, these numbers should be interpreted with caution. The fact that one-third of the students in this sub-sample are high-need girls may signal that Coaching Corps is doing a good job of connecting with and retaining this traditionally hard-to-reach population.

Table 4: Average fitness scores and changes in fitness scores for students with complete Year 2 and Year 3 data[A B C]

	All students						Change		
	Year 2			Year 3					
	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Δ 1 to 3	Δ 1 to 6	p-value
Curl-Up (age < 10)	4 (n=148)	6 (n=148)	8 (n=148)	5 (n=148)	8 (n=148)	9 (n=148)	+ 4	+ 5	< 0.001
PACER (age ≥ 10)	13 (n=19)	21 (n=19)	39 (n=19)	26 (n=19)	32 (n=19)	39 (n=19)	+ 26	+ 26	<0.001

	All girls								
	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Δ 1 to 3	Δ 1 to 6	p-value
Curl-Up (age < 10)	4 (n = 73)	5 (n = 73)	8 (n = 73)	4 (n = 73)	8 (n = 73)	9 (n = 73)	+ 4	+ 5	<0.001
PACER (age ≥ 10)	14 (n=5)	19 (n=5)	43 (n=5)	29 (n=5)	33 (n=5)	32 (n=5)	+29	+ 18	<0.001

	All boys								
	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Δ 1 to 3	Δ 1 to 6	p-value
Curl-Up (age < 10)	3 (n=75)	7 (n=75)	8 (n=75)	6 (n=75)	9 (n=75)	9 (n=75)	+ 5	+ 6	<0.001
PACER (age ≥ 10)	13 (n=14)	22 (n=14)	37 (n=14)	25 (n=14)	32 (n=14)	41 (n=14)	+ 24	+28	<0.001

	All high-need students[D]								
	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Δ 1 to 3	Δ 1 to 6	p-value
Curl-Up (age < 10)	1 (n=109)	5 (n=109)	7 (n=109)	5 (n=109)	7 (n=109)	8 (n=109)	+ 6	+ 7	<0.001
PACER (age ≥ 10)	9 (n=16)	19 (n=16)	36 (n=16)	26 (n=16)	26 (n=16)	36 (n=16)	+ 27	+ 27	<0.001

	High-need girls								
	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Δ 1 to 3	Δ 1 to 6	p-value
Curl-Up (age < 10)	1 (n=58)	4 (n=58)	7 (n=58)	4 (n=58)	6 (n=58)	8 (n=58)	+ 6	+ 7	.002
PACER (age ≥ 10)	6 (n=3)	14 (n=3)	36 (n=3)	26 (n=3)	20 (n=3)	25 (n=3)	+ 30	+ 19	<0.001

	High-need boys								
	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Δ 1 to 3	Δ 1 to 6	p-value
Curl-Up (age < 10)	1 (n=51)	5 (n=51)	7 (n=51)	5 (n=51)	8 (n=51)	9 (n=51)	+ 6	+ 8	<0.001
PACER (age ≥ 10)	9 (n=13)	20 (n=13)	36 (n=13)	25 (n=13)	27 (n=13)	38 (n=13)	+ 27	+ 29	<0.001

[A] n=number of students. Numbers may not add up perfectly, due to rounding. P-values are for the change in laps between all 6 times and were calculated using linear regression models accounting for clustering by site and student. A p-value less that 0.05 is considered statistically significant and means that observing a change in laps as extreme or more extreme than the change observed by chance alone, given that there really is no change in laps, is very unlikely.
[B] Time 1 is Year 2, Time 1; Time 2 is Year 2, Time 2; Time 3 is Year 2, Time 3; Time 4 is Year 3, Time 1; Time 5 is Year 3, Time 2; Time 6 is Year 3, Time 3.
[C] Age is age at Time 1.
[D] Classified as high-need at Time 1.

Coaching Corps was able to retain more students across Years 2 and 3 than they retained across Years 1 and 2; whereas there were 72 students with complete data from Years 1 and 2 combined, there were 167 students with complete data from Years 2 and 3 combined. Of these 167 students, all groups increased their fitness scores on average between both Times 1 and 3 and Times 1 and 6, with the largest improvements happening between Times 1 and 3 (during Year 2). All groups decreased their fitness scores between Times 3 and 4 (when change in school year occurred). A decrease in fitness and physical activity levels during the summer break has been demonstrated in other studies, so this finding is not surprising. Positively, while students' scores decreased between Times 3 and 4, the scores did not regress to the Time 1 scores (with the exception of all girls under 10 years-old, who had a curl-up score of 4 at both Times 1 and Time 4). This positive finding could be due partially to an improvement in scores due to maturation, as well as due to real changes in fitness maintained from participation in Coaching Corps the previous year. Without a control group, it is difficult to

identify the exact cause or to attribute these changes in fitness scores to Coaching Corps alone. Still, these changes do point in a positive direction, suggesting that Coaching Corps may be a sustainable program that has the potential to impact students' fitness over multiple years.

There were only 19 students older than 10 who had complete data for both Years 2 and 3; of these, 3 were high-need girls and 13 were high-need boys. While Coaching Corps was not able to retain a large number of students older than 10, of those who remained in the program across both years, 84% were high-need. With such a small number of students, it is difficult to draw real conclusions from this data, but it suggests that this could be a hard-to-reach (and retain) population that may require greater efforts to enroll and maintain as program participants (this could also be a product of the ages of students who were at participating sites). Of the high-need students older than 10 that Coaching Corps was able to retain across Years 2 and 3, both boys and girls demonstrated substantial relative increases (> 300%) in the number of laps they could run between Time 1 and Time 6.

* *Note*: There were 34 students with data from Years 1, 2, and 3.   Of these students, 32 were under 10 years old. We did not analyze the data from these students, as the sample it too small and prevents us from drawing meaningful conclusions.

Table 5: Number of Students who were below the Healthy Fitness Zone (HFZ) at Time 1 who moved into the HFZ at Time 3, by gender[A]

		All students			All Girls			All Boys		
		# Below HFZ at Time 1	# Who Moved into HFZ at Time 3	% Who Moved into HFZ at Time 3	# Below HFZ at Time 1	# Who Moved into HFZ at Time 3	% Who Moved into HFZ at Time 3	# Below HFZ at Time 1	# Who Moved into HFZ at Time 3	% Who Moved into HFZ at Time 3
Year 1	PACER (age < 10)	45	25	56%	22	13	59%	23	12	52%
	PACER (age ≥ 10)	115	46	40%	45	17	38%	70	29	41%
Year 2	Curl-Up (age < 10)	331	169	51%	172	87	51%	159	82	52%
	PACER (age ≥ 10)	154	69	45%	69	32	47%	85	37	44%
Year 3	Curl-Up (age < 10)	283	132	47%	130	61	47%	153	71	46%
	Curl-Up (age ≥ 10)	83	39	47%	37	19	51%	46	20	43%
	PACER (age < 10)	86	44	51%	46	23	50%	40	21	53%
	PACER (age ≥ 10)	61	24	39%	22	11	50%	39	13	33%

[A] During Year 1 there were 241 students (97 girls and 147 boys) who were younger than ten and 188 students (76 girls and 112 boys) who were ten or older. During Year 2 there were 470 students (237 girls and 233 boys) who were under ten and 235 students (127 boys and 128 girls) who were ten or older. During Year 3 there were 352 students (169 girls and 183 boys) who were younger than ten and 119 students (53 girls and 66 boys) who were ten or older.

Across all 3 years, an average of 47% of students who were below the HFZ at Time 1 moved into the HFZ by Time 3. It is important to note that a far greater number of participants who were younger than ten in Year 3 were below the HFZ for the curl-up (81% of students) as compared to for the PACER (29% of students); 198 students who were classified as below the HFZ based on their curl-up score were classified as above the HFZ for their PACER score. This substantial difference could be related to the tests themselves, and not due to true fitness differences (i.e. typically younger students are more familiar with running and have less experience with the curl-up, which could lead them to do better on their initial PACER test than on their initial curl-up test). The curl-up may not be an ideal stand-alone test of younger student's fitness, as it appears to have the potential to misclassify students as unhealthy.

The relationship between changes in fitness scores and attendance:

Average number of days of student attendance increased from Year 1 to Year 3, moving from 39 days in Year 1, to 31 days in Year 2, to 47 days in Year 3. There is some question about the quality of the attendance data as several students with complete data had unrealistically low attendance data. This could impact these analyses.

Year 1: We used data from the 429 students who ran the PACER test at Time 1, Time 2, and Time 3. Of these students, 241 were younger than ten and 188 students were ten or older. Taking into account student's fitness levels at Time 1, whether they were a girl or a boy, their age, and their program site, we found that attendance is positively related to improvements in student fitness; students with higher attendance showed greater improvement in fitness.

Students of all ages analyzed together
For each additional ten days of attendance, the number of laps students could run from Time 1 to Time 3 increased by:
- **All Students**: 2 laps
- **All Students, Girls**: 1.7 laps
- **All Students, Boys**: 2.2 laps
- **High-need students**: 3 laps
- **High-need students, Girls:** 2.3 laps
- **High-need students, Boys**: 2.5 laps

Students younger than ten and ten and older analyzed separately
For each additional ten days of attendance, the number of laps students could run from Time 1 to Time 3 increased by:
- **All Students**:
  - *Younger than ten:* 2.5 laps
  - *Ten and older:* 1.6 laps
- **All Students, Girls**:
  - *Younger than ten:* 3.3 laps
  - *Ten and older:* 0.8 laps (but this was not statistically significant).
- **All Students, Boys**:
  - *Younger than ten:* 2.8 laps
  - *Ten and older:* 1.9 laps (but this was not statistically significant).
- **High-need students**:
  - *Younger than ten:* 3.7 laps
  - *Ten and older:* 2.6 laps
- **High-need students, Girls**:
  - *Younger than ten:* 5 laps.
  - *Ten and older:* 1.7 laps (but this was not statistically significant)
- **High-need students, Boys**:
  - *Younger than ten:* 3.4 laps
  - *Ten and older:* 2.8 laps

Year 2: We used data from the 470 students who did the curl-up and the 235 students who ran the PACER test at Time 1, Time 2, and Time 3. Taking into account student's fitness levels at Time 1, whether they were a girl or a boy, their age, and their program site, we found that, overall, attendance was not positively associated with improvements in student fitness. When we looked at the data by gender, we similarly found that there was no statistically significant relationship between attendance and student fitness. When we looked at high-need students, specifically, we found that for every additional ten days of program attendance, high-need boys significantly increased the number of curl-ups they could do by 1 curl-up.

Year 3:

We used data from the 471 students who did the curl-up and ran the PACER test at Time 1, Time 2, and Time 3. Taking into account student's fitness levels at Time 1, whether they were a girl or a boy, their age, and their program site, we found that, overall, attendance was not significantly associated with improvements in student fitness. When we looked at the data by gender, as well as by high-need status, we similarly found no statistically significant relationship between attendance and student fitness.

Year 1 and 2 combined: When looking at the subset of 72 students who ran the PACER test during both Year 1 and Year 2, we did not separate students out by younger and older than ten, because the groups would become too small.

>**For students of all ages analyzed together:**
>For each additional ten days of attendance the number of laps students could run from Time 1 to Time 3 increased by:
>- **All Students, Boys**: 1.7 laps
>- **High-need students**: 1.8 laps
>- **High-need students, Boys**: 2.8 laps
>- **All students; all students, girls**; and **high-need students, girls** did not significantly increase the number of laps they could run from Time 1 to Time 6.

While we did see a positive relationship between increased attendance and fitness scores in Year 1, we did not see the same relationship in Year 2. When we looked at the subset of students who had data from both Year 1 and Year 2 we found that for each additional ten days of attendance, high-need boys significantly increase the number of laps they could run from Time 1 to Time 6 by almost 3 laps.

Year 2 and 3 combined:

There were only 2 students that had PACER data for both Years 2 and 3 (the rest were younger than 10 and therefore had curl-up data only for Year 2). We cannot meaningfully assess the relationship between fitness score and attendance in this small sample.

Summary:

On average, students demonstrated improved fitness from the beginning to the end of the program each year that far exceeds the improvement we would expect to see from maturation alone. These increases persisted when looking at changes in fitness among high-need students, boys and girls, and among high-need boys and girls. Additionally, nearly 50% of students who started out each year classified as below the Healthy Fitness Zone for curl-up and/or PACER moved into or above the Healthy Fitness Zone by Time 3. It is particularly heartening to see the increase in girls' fitness levels, as national research has shown that girls' physical activity and fitness levels decline as they get older. The fact that the proportion of students with complete data steadily increased across all 3 years (from 31% in Year 1 to 49% in Year 3), coupled with the increase in the number of students who participated in the program in both Years 2 and 3 combined, demonstrates that Coaching Corps improved overall student retention and streamlined and improved their data collection and evaluation procedures across the evaluation period. This further suggests that Coaching Corps may be a sustainable program that has the potential to impact students' fitness over multiple years. The decrease in student participants and number of sites between Years 2 and 3 may have been due to funding, but could afford Coaching Corps the opportunity to intensify efforts to recruit, retain, and successfully coach high-need students and females (particularly older females) in the remaining sites.

Average number of days of student attendance increased from Year 1 to Year 3, moving from 39 days in Year 1 to 47 days in Year 3. The attendance data from Year 1 were promising; taking into account student's fitness levels at Time 1, their sex, program site, and age, we found students with higher attendance showed greater improvement in fitness. However, the attendance data from Years 2 and 3 did not show a significant association between student attendance and fitness scores. The lack of an association in Years 2 and 3 sheds some doubt on findings in Year 1. It is important to recognize that without data from control students, we cannot attribute these findings to Coaching Corps alone. It is also noteworthy that the quality of the attendance data may not be high, as several students with complete data had unrealistically low attendance data.

Overall, these data suggest that participation in an organized after-school sports program with a trained coach may help improve fitness outcomes for students. Further research is still needed to better understand the impact of this program on student fitness and other youth development outcomes.

Recommendations:

- Shift focus from time-intensive individual testing to class-level observations.
  - o Continue to use the System for Observing Fitness Instruction Time (SOFIT) at sites in order to assess how much moderate-to-vigorous physical activity students are getting when they are working with a coach. This will continue to help Coaching Corps quantify how physical activity session time is being spent, as well as better understand the day-to-day operation of the program on-site.
- If continuing to collect individual-level data, in the absence of funder requirements, eliminate the use of the curl-up to assess student fitness levels, as this test is not best suited to measure changes in fitness we'd expect to see as a result of participating in Coaching Corps programming.
- Maintain continued attention to data quality, including:
  - o Ensuring SOFIT data collectors are rigorously trained and tested for reliability so that observations are conducted consistently across sites and across years
  - o Ensuring SOFIT data is filled in completely (with no portions missing) and that data entry is checked for errors
  - o Ensuring individual-level test administration is consistent across sites and across years.
  - o Getting solid attendance and program dosage data
  - o Ensuring student's age and sex are correct
- Collect process evaluation data, such as surveys/interviews with coaches, to better understand how the program is working at certain sites. These data can provide context for quantitative results and may help with quality assurance and program improvement.