Peer Influences on High School FAFSA Completion: A Quasi-Experimental Impact Analysis of College Summit's PeerForward Program.

Paul Scott

University of Pittsburgh

Executive Summary

College Summit is an education nonprofit that over the last 20 years has placed more than 350,000 students from 500 high schools in low-income communities on the path to college and career success. PeerForward is an exciting innovation that builds on College Summit's decades of experience, as well as third-party research about what really works in schools.

PeerForward Program and Background

Launched in 2016, College Summit's PeerForward program is designed to leverage the influence and power of peer leadership by training teams of eight high school juniors and seniors (who are referred to as Peer Leaders) and their PeerForward Advisor (a trained high school staff member) to guide their classmates to college. The PeerForward model is composed of three campaigns, each tied to an outcome that has been proven to boost college enrollment: applying to three or more colleges (Smith, 2011), filing early for financial aid (Roderick et al., 2008), and connecting academics to college and career (Bedsworth & Doctor, 2006). Through PeerForward, College Summit partners with high schools to identify, train, and support these Peer Leaders and an Advisor to plan and execute the model. There is an emphasis on high schools in low-income communities, where many participating College Summit students would be the first generation of college graduates in their families, and on schools in which the counselor–student ratios exceed 1:500. A copy of the PeerForward Logic Model can be found in Appendix A. The PeerForward program is unique in unleashing the power of student-driven change to address college preparation and enrollment in high schools serving low-income communities.

This study tests the PeerForward Theory of Action, depicted in the PeerForward Logic Model, that a team of influential and trained Peer Leaders with the support of a PeerForward Advisor and College Summit coaching staff can influence the number of students that file early for financial aid when compared to similar schools not implementing PeerForward, as measured by Free Application for Federal Student Aid (FAFSA) completion data as of March, 3rd 2017.

Study Design

This study employs propensity score methods to evaluate the effectiveness of the PeerForward program in influencing students to complete the FAFSA by March 3rd; a key campaign metric of the PeerForward program. An overview of the theory underlying the usage of propensity scores in causal inference studies is given below. The general idea is to create treatment and control groups that are balanced across a set of characteristics that have been shown to have an impact on high school FAFSA completion rates. In effect this helps to eliminate confounding of a treatment effect that might be due to one, or a set, of these characteristics.

State-by-state datasets were pulled from state department of education websites and included various school characteristics such as prior achievement and demographics. For the most part, states keep the same basic information. However, the specific measures often vary (e.g., variation in state testing systems) as does the richness of the dataset (i.e., some states collect more information than others). To empirically determine important covariates to balance between the groups, the state datasets were mined to find influential covariates on FAFSA completion rates. More technically speaking, Boosted Poisson Regression Trees were applied to the FAFSA returns using the full set of covariates. This resulted in the creation of a dataset of the most influential school characteristics, from the available state data, which were then used to predict group membership and subsequently generate propensity scores for each school. With these propensity scores, we were able to eliminate or trim schools from the sample that were not similar across the treatment and control groups. After trimming the sample, a sampling weight was created using the propensity score to further reduce bias in estimating the effect of PeerForward on early FAFSA returns due to pre-existing group differences. Additionally, when possible, schools were stratified on their propensity scores within states such that the closest matching schools across groups are compared in a random effects model. Because our outcome of interest is a count of completed FAFSA applications we utilize a Poisson Regression. To account for variation in school size we set the senior enrollment count as an exposure such that FAFSA returns are evaluated relative to the size of the senior class. Treatment effect results are reported as incidence rate ratios (IRR) for their interpretability and value as effect size measures.

Results

The overall results indicate a positive influence of the PeerForward program on the rates of early FAFSA returns. The state-by-state reporting indicates greater efficacy in some states than in others, but consistency in a positive direction. The individual states were pooled into a random effect model for meta-analysis. On the whole, across states, it appears that PeerForward schools had about 22 percent higher FAFSA returns than their counterparts. When we apply a standardized weight at the within strata level, such that the propensity score scales are the same relative scale in all states, we find the rate estimate increases slightly, showing a 26 percent higher FAFSA completion rate for PeerForward schools when compared to similar schools not receiving the PeerForward program. However, the significance level is not altered.

General Overview of the Theory underlying the Methods used for Evaluating PeerForward

Random assignment has long been considered the gold standard in deriving causal inference from a study on the effects of an intervention. In essence, the value of randomization is that by helping to ensure each individual has an equal probability of being in any of the groups being evaluated we should end up with groups which are similar in character. Having such assurance is desirable as it mitigates concerns that group-specific characteristics are overly influential on measured outcomes. When estimating the effects of an intervention, it's important that we can trust that observed differences in an outcome amongst groups is due to the intervention instead of group-specific differences in characteristics (i.e., confounding influences). In the real world, true random assignment of interventions is often not an option, thus we must retroactively create a pseudo-randomization in order to evaluate intervention effects. Such pseudo-randomization can be accomplished by balancing covariates amongst groups. The aim is to help ensure that groups are equal in expectation. The key to such a process is removing bias introduced in the outcome by pre-existing differences amongst intact groups. The ideal is to control out the effects of influential covariates on the outcome, thus allowing for an unbiased estimate of the effect of an intervention. One approach to achieving this aim is to use propensity scores as a means of equalizing groups in their expected outcomes (Abadie & Imbens, 2009; Rosenbaum & Rubin, 1983).

Following the aforementioned points, analysis based on propensity score matching, similar to experimental research, proceeds in two basic stages: (1) setting up the study design, then (2) running analysis on the outcome (Stuart, 2010). When the design stage is not characterized by actual random assignment to treatment groups, it becomes necessary to establish a pseudo-randomized design. Propensity Matching belongs to a broader class of methods that pseudo-randomize treatment groups by balancing covariates to reduce any systematic bias in group performance on outcomes aside from the treatment condition. The essence of this approach is to select important covariates that are apt to influence performance on the outcome, and equalize the groups on these covariates to control out their influence on estimating a treatment effect. The specific way in which propensity scores do this is by taking the conditional probability of belonging to a group given some set of covariates and balancing groups on this probability. In this way, the propensity score serves as a summary score for a larger set of covariates specifically in terms of how those covariates bear upon the treatment. The design aspect of the analysis begins by finding covariates that differentially influence performance on an outcome across groups, next we estimate a probability of being in a specific group conditioned on these covariates, then we find a range within this conditional probability distribution which contains individuals in both groups. Following the design stage, we incorporate the probability value for an individual (i.e. propensity score) into the analysis of group differences on an outcome in such a way as to reduce bias induced by these probabilities of group membership.

General Overview of the Methodological Approach to this Evaluation

Data

PeerForward has been implemented in multiple states, which allows us to evaluate the program within states and across states. However, states tend to vary both in the data collected and their educational policies and practices. Thus, we acquire better datasets by compiling state specific information provided at the state level instead of using a federal dataset such as the Common Core of Data. The type of information at the state level tends to be consistent and comparable, e.g. enrollment counts at schools, academic achievement, demographics, etc., but can vary in the exact ways in which it is reported. Our outcome of interest, school-level FAFSA completion counts, were collected using the Federal Student Aid FAFSA Completion by High School dataset released by the Department of Education (DOE). These DOE state datasets were matched with 2017 state reported senior class enrollment numbers by high school. College Summit compiled these state specific datasets and provided this data to an independent consultant for analysis. In this way, each analysis begins with a state specific dataset.

Establishing Covariates Influential on Return Rates for FAFSA

Variable selection involves a combination of empirical and theoretical considerations as to what covariates will be most important to balance between the treatment and control groups to reduce bias in estimating a treatment effect. The main issue is to consider important predictors of the outcome that we want to be equal between the groups prior to evaluating the treatment effect. The selection of such covariates will be guided by the researchers conceptual understanding of an outcome, knowledge of important covariates from prior research, as well as empirical analysis. In this study, theoretical judgement guided the compilation of the initial datasets. The initial datasets contained a large and rich set of covariates, which is important since the performance of a propensity score analysis can be heavily degraded by omitted variable bias (Heckman, 2005; Guo & Fraser, 2010). When one has a large set of covariates, particularly helpful techniques for empirically determining influential predictors of an outcome can be found in the data mining literature (Hastie, Tibshirani, & Friedman,2009; James et al., 2013). One of the most valuable techniques for this purpose is the usage of Generalized Boosted Regression which can be implemented through the R package 'gbm' (Ridgeway, 2017;2007). Because boosting is known to generate low error in prediction and classification along with its ability to sort out the relative influence of various predictors on an outcome, it is an important technique that can be incorporated into propensity score analysis (McCaffrey et al., 2004; Guo & Fraser, 2010).

Setting up the Propensity Score Analysis

After establishing a robust set of covariates that prove influential on the outcome, we can predict group membership from these covariates to estimate a conditional probability that will serve as the basis for the propensity score. The next step is to eliminate propensity score ranges where only one group is represented, this is what we call the region of common support. The propensity score is then used in two different ways: (1) we create strata in order to match individuals across groups that can be assumed to be most similar given the closeness of their propensity scores; and (2) we can create an inverse probability weight (Lunceford & Davidian, 2004) to representatively adjust cases in the same way as sample weights (Horvitz & Thompson, 1952). The weight can be defined as:

$$\frac{Treatment_i}{\widehat{Pr}_i} + \frac{1-Treatment_i}{1-\widehat{Pr}_i}$$

where $\widehat{Pr}_i$ represents the estimated propensity score for an individual. This weight allows for the estimation of the average treatment effect in the sample by appropriately weighting both the treatment and controls up to the full sample.

Analyzing the Outcome

Given that our outcome concerns the count of FAFSA returns by early March, 2017, we model this as Poisson distributed and estimate the treatment effect in terms of an incidence rate ration (IRR) using Stata 14. Further, since the count of FAFSA is likely to be dependent on the number of students enrolled, we use the enrollment numbers as an exposure variable so that we get a more correct representation for the rates of return. In some states, multiple propensity strata are created; whereas within other states only one strata was feasible. For the states with multiple strata, we use the strata as clusters in a weighted, mixed-effect Poisson Regression such that the analysis is focused at the within strata level with the propensity weights applied at the within strata level. For states where only one strata are feasible, we simply do the state-specific analysis as a weighted Poisson Regression. Each state specific sample is appended to a larger dataset containing all state data. We use a similar approach in estimating the treatment effect across states. We employ a weighted mixed-effect Poisson Regression with the strata specified at the clustering level and propensity weights applied within strata. As a further adjustment, we also employed clustered robust standard errors, with the state specified at the clustering level, and allowed the treatment effect to vary across strata. As a further examination of the between state effects we can apply a propensity weight at the between state level. However, because the propensity scales might vary from state to state (e.g., in some states schools may, in general, have larger probabilities than in other states) it is of value to create a standardized propensity score. To do this we estimate a z score for propensities within each state, then place this z score onto a probability scale by using an inverse logit function $\left[\frac{\exp(z)}{\exp(z)+1}\right]$. After doing this we can then estimate an inverse probability weight as was done before. We can then create a mean standardized weight for each state and enter it into the mixed-effect model as a sampling weight applied to the between strata level.

Results

Across States

Analysis was done at the state-by-state level, then each state was pooled into a larger dataset where each state had 1) a variable indicating whether the school participated in PeerForward or not, 2) propensity scores and weights, 3) the enrollment count of their 12th grade class, 4) the state and strata to which they belonged, and 5) the count of FAFSA returns as of early March, 2017. Thus, this analysis can be considered as a kind of meta-analysis. The analysis, based on the collection of states, indicated significantly positive effects of PeerForward on the rates of FAFSA returns. This result is demonstrated across various weighting specifications explained below:

Employing the raw propensity weight within the state strata level the incidence rate ratio implies that PeerForward schools had a rate of about 22 percent higher returns for PeerForward schools, IRR=1.218, SE=0.102, p=0.019, 95% CI= (1.033, 1.436). Adding the mean standardized weight to the between strata level doesn't alter the results by much, leading to the same conclusion as before, IRR=1.220, SE=0.118, p=0.040, 95% CI=(1.009, 1.474).

When we apply a standardized weight at the within strata level we find the rate estimate increases slightly to about 26 percent; however, the significance level is not altered, IRR=1.256, SE=0.123, p=0.019, 95% CI=(1.038, 1.521). The increase in the rate estimate may be due to the fact that the standardized weight values are larger than the raw weight values. As before, the results are not altered much by entering the mean standardized weight at the between strata level, IRR=1.261, SE=0.136, p=0.031, 95% CI=(1.021, 1.558). As we will demonstrate in the state-by-state analysis, in some states a significant treatment effect was detected, while in other states it was not. However, all in-state analyses indicated the effect was at least moving in a positive direction.

We conducted a sensitivity analysis to evaluate the influence of each individual state on the estimation of the treatment effect. The sensitivity analysis involved estimating the treatment effect while excluding each state in turn. For example, the first entry in Table 1, labeled CA in the first column, displays the estimated treatment effect when we base the treatment effect estimation on all states except California. In other words, it represents the estimated treatment effect if California were not included in our sample. The reported results have the unstandardized propensity weight specified at the within state level only. Table 1 presents these results, the consistency in the estimated effect implies that no one state is overly influential on the final results. Appendix B demonstrates that the results of the sensitivity analysis are similar across the different weighting specifications.

Table 1. *Analyses Excluding Each State in Turn.*

Excluded State	*N*	*IRR*	*SE*	*p(Sig.)*	*95% CI*	
CA	170	1.208	0.086	0.008	1.05	1.39
FL	186	1.302	0.147	0.019	1.044	1.624
MD	192	1.228	0.12	0.036	1.014	1.488
MI	180	1.19	0.106	0.05	1	1.417
MO	192	1.259	0.114	0.011	1.054	1.504
NY	107	1.269	0.13	0.021	1.037	1.551
OH	185	1.211	0.111	0.037	1.012	1.45
SC	174	1.136	0.059	0.014	1.026	1.258

State-by-State[1]

In the following we will go over the analysis as it was implemented in each state. As mentioned before, we observed differences in the estimated treatment effect and the sample sizes tend to be smaller. Table 2 gives an overview of the results

Table 2. *State-by-State Results*

State	*Treatment*	*Control*	*IRR*	*SE*	*p(Sig.)*	*95% CI*		*Number of Covariates*
CA	7	21	1.07	0.126	0.569	0.849	1.348	37
FL	5	7	1.076	0.166	0.636	0.796	1.454	34
MD	3	3	1.172	0.062	0.002	1.058	1.3	34
MI	6	12	1.633	0.376	0.033	1.039	2.565	52
MO	2	4	1.007	0.002	0.763	0.964	1.051	13
NY	12	79	1.045	0.157	0.772	0.778	1.403	83
OH	3	10	1.319	0.284	0.198	0.866	2.01	28
SC	4	20	1.774	0.346	0.003	1.211	2.599	79

[1] Detailed information concerning the specific methods and results within each state can be directly requested via email: pws5@pitt.edu

California

In the state of California, we began by dropping three schools that received a hybrid program. This gave us 10 PeerForward schools and 2,463 potential controls. The State of California provides information on school type. As all PeerForward schools are considered traditional public high schools any school not considered as such was dropped from the analysis. This reduced the potential controls to 1,250. Schools with missing information on the outcome were dropped, as were schools representing grades outside of the 9th through 12th grade range. This left us with 1,093 potential controls. When treatment was boosted on the covariates, perfect separation was achieved, and due to perfect prediction propensity scores the logistic regression could not be used. Propensities were generated from a linear regression using maximum likelihood estimation. When we attempted to generate propensity scores from the full set of influential predictors we ran into collinearity issues. Hence, treatment was regressed on each covariate in turn to generate propensity scores from each individual predictor. The distribution of each of these propensities was examined across treatment groups and cases whose propensities were in a non-overlapping region were dropped. The resulting trimmed sample had 21 control and 10 PeerForward schools. We then averaged across these propensities to create a mean propensity score. Additional cases outside the overlapping region of the mean propensity score were dropped, leaving us with 21 control and seven PeerForward schools. Schools in California could not be stratified on the propensity score due to the restricted range. Attempts to create such strata resulted in a failure to estimate the treatment effect. Hence, a propensity weighted Poisson Regression with only fixed effect was applied to estimate the treatment effect. The results of this analysis showed a positive, but non-significant rate of return. The results are displayed in Table 2.

Florida

For Florida, cases missing values on the outcome were dropped. Some variables did not vary within the treatment groups and thus were dropped from the dataset. We began with a sample of 10 PeerForward schools and 598 potential controls. Accordingly, with the standard procedure this dataset was taken into R and a Boosted Poisson Regression was run with FAFSA count as the outcome and with all covariates in the dataset, except the treatment indicator, serving as predicting variables. We then examined the distributions of the influential covariates across treatment groups. The propensity from boosting resulted in near perfect separation, thus preventing matches from being made. Because there were too many perfect predictions from the logistic regression, propensity scores generated from the logit were not feasible. A linear regression with maximum likelihood estimation, to handle missing values, was used to generate a propensity from each covariate. We then trimmed cases which were outside of the overlap of propensity scores between treatment and control. This left us with 14 control and nine PeerForward schools. These propensity scores were then averaged to create a composite propensity score. Additional cases were dropped that were outside of the overlap for the mean propensity score. The final trimmed sample had seven control and five treatment schools. The creation of strata within the state was not feasible, hence only one strata is represented in Florida. As seen in Table 2, the estimated effect of PeerForward was positive but non-significant.

Maryland

The Maryland state analysis began by removing cases with missing values on the outcome then moving the dataset into R to run a boosted regression to identify variables most influential for predicting FAFSA returns. When the boosted regression was run to predict PeerForward schools from the influential predictors, the resulting propensity scores created perfect separation of PeerForward and non-PeerForward schools, thus no matches could be made. However, we were able to use a logit to generate propensity scores from each influential covariate. We then trimmed the cases which had propensities in the non-overlap region and further trimmed cases in the non-overlap on the mean propensity. For Maryland, we were able to create two propensity strata. There were three individuals in each strata, and we had three PeerForward and three control schools. As shown in Table 2 the results were significantly positive with about 17 percent higher FAFSA returns by early March.

Michigan

In the Michigan dataset, there were many variables with a large amount of missing values, thus those with less than 20 percent of values represented were dropped. Cases with missing values on the outcome were also dropped. A boosted regression was run on FAFSA returns, and the most influential variables were retained and used to generate propensity scores. When boosting the treatment on the influential predictors, we ended up with perfect separation of PeerForward and non-PeerForward schools. We then attempted a logit, but there was a problem with perfect prediction due to a lack of variation within treatment groups for some predictors. Propensities were then generated via linear regression with maximum likelihood estimation (i.e., regressions were run in a structural equation modeling framework). Cases outside of the overlap region of the propensity scores for each individual predictor were dropped. Then a mean propensity score that pooled across the propensities from each individual predictor was created and further trimming was done to cases outside of this common support region. Within Michigan strata were not feasible, hence only one strata was represented by Michigan. The sample is comprised of six PeerForward and 12 control schools. Table 2 shows that Michigan had a positive and significant rate of return with about 63 percent higher rates of returns.

Missouri

Missouri was evaluated in a similar fashion as the other states. We removed cases missing values on the outcome and then ran a boosted regression on FAFSA counts to acquire a set of variables influential on the outcome. Predicting PeerForward schools from the set of influential predictors using boosted regression resulted in propensity scores that perfectly separated PeerForward from non-PeerForward schools, hence there was no common support to work with. Logistic regression couldn't be used due to predictors that did not vary amongst the treatment groups. Thus, we calculated propensity scores via linear regression with maximum likelihood estimation. In this case the propensity scores are based on predicting the treatment from the full set of influential predictors. Cases outside this region of common support were dropped. This resulted in four control and two PeerForward schools. Missouri was represented by two strata. The results for Missouri implied no treatment effect for PeerForward (i.e., the rate of return ratio between PeerForward and non-PeerForward was essentially one, see Table 2).

New York

Using the full set of covariates deemed most influential on FAFSA returns via a boosted regression, we attempted to first establish a common support based on a boosted propensity. However, as occurred in many other states, no common support region existed. We then attempted to utilize logistic regression to produce a common support for the propensity scores; however, some predictors did not vary amongst the treatment groups (i.e., perfect prediction). Hence, we generated a propensity score using a linear regression with maximum likelihood estimation. In this case, we were able to generate this score using the full set of influential predictors. We removed cases with non-overlapping propensity scores, then created three strata for New York. This sample contained a total of 79 control and 12 PeerForward schools. More specifically, in Strata 1 there were three PeerForward and 28 control schools, in Strata 2 there were five PeerForward and 45 control schools, and in Strata 3 we had four PeerForward and six control schools. Table 2 shows that results were positive but non-significant for New York.

Ohio

Ohio was analyzed in a similar fashion to the other states. First, by boosting FAFSA returns on the full set of variables, then retaining the covariates with the highest influence. Next, we attempted to find a common support region on the propensity score created by boosting treatment on the influential predictors. When no such region could be identified, we attempted logistic regression. Due to some predictors not varying amongst the treatment groups, propensity scores could not be calculated using the logistic approach. We then attempted to estimate propensities from a linear regression using maximum likelihood estimation with the full set of influential covariates, but due to collinearity issues we could not do this either. Hence, we generated propensity scores with a linear regression using maximum likelihood estimation for each individual covariate in turn. We then trimmed cases outside of the common support for each of these respective covariates, and then pooled across these propensities to create a mean propensity score which we additionally trimmed on. The creation of strata was not feasible for the state of Ohio. The resulting sample had 10 control and three PeerForward schools. Table 2 shows that results were positive, but didn't reach conventional significance levels. The rate estimate in Ohio implied nearly 32 percent higher rates of return; however, the large standard error led to a non-significant finding.

South Carolina

Analysis in South Carolina followed the same basic procedure as with the other states. We identified the most influential predictors of FAFSA returns in the state's dataset using a boosted regression. Then, treatment was boosted on the set of influential predictors. In South Carolina, we were able to establish a common support region on the propensity scores from the boosted regression. Thus, we were able to bring the propensity score from R into Stata. We then dropped treatment cases with propensity scores above the maximum control propensity score and control cases below the minimum propensity for treatment. This resulted in a sample of 20 control and four PeerForward schools, three strata were created. The resulting analysis indicated that the rate of returns for PeerForward schools was significant and positive, with nearly 77 percent higher FAFSA returns by early March for PeerForward schools.

Limitations

A major limitation of the study is that, for most states, the resulting analysis samples tended to be small. The PeerForward treatment schools included in the analysis may not be representative of the entire PeerForward population of schools, and this raises the need for further inquiry into the characteristics of the schools on which the treatment effects were estimated.

As can be seen in the above analysis, boosting often failed to create regions of common support on the propensity score. Boosting is known to be a notably stringent classifier, thus the fact th at the number of PeerForward schools in each state is small relative to the control schools is likely driving increased precision in classification. This relates to the notion that the algorithm for boosting is better able to hone in on the individual characteristics of each school within the treatment allowing PeerForward schools to be easily distinguished from the control.

Another limitation is that the information available from state to state was not entirely consistent. Accordingly, the necessary approach to conduct the propensity analysis isn't entirely consistent from state to state. In a more ideal situation we would be able to create propensities from the same set of information across states and in the exact fashion. Appendix C gives the set of covariates utilized for each state. This resulted in propensity scales that were not on the same scale (e.g., some states may have propensity score values that are greater or less than the propensity score values in other states). An attempt to address this issue was made by creating standardized propensity scores. The state-by-state results using this approach are given in Appendix D. As can be seen, the findings are relatively robust to the different propensity specifications. Unfortunately, meta-analytic approaches to propensity score analyses, as was done in this study, are (to this researcher's knowledge) underrepresented. Thus, further methodological studies concerning this are warranted.

As with any propensity analysis, our results are dependent on the information we have observed. Thus, one must always be cautious that if any unobserved features have a strong influence on the treatment groups, then the propensity analyses will be biased. Moreover, with any program evaluation, it is important that we take into account variation in implementation which in this case is not observed. Similarly, there is no information available on whether the non-PeerForward schools were receiving services from another program that may be influencing FAFSA returns. This study would greatly benefit from having such information.

Next steps

As noted above, in order to improve the internal validity (i.e., causal inference) of this study we were required to work with greatly reduced samples. The constriction of the sample obviously raises questions concerning the generalizability (i.e., external validity) of the results found in this study. This is the classic example of the trade-off between increasing internal validity at the expense of external validity that researchers are often forced to choose between. It will be valuable to follow-up on this analysis to assess the generalizability of these findings. It is recommended that College Summit closely examine the analytic sample of this study relative to the larger sample from which they were drawn. For example, we may want to know if the schools in the analytic sample significantly differ from schools excluded from the analysis on key demographic features.

One of the most important aspects of conducting causal inference studies is to verify the robustness of the findings. We will address any additional concerns and considerations that may come up to the best of our ability with the appropriate robustness checks. In this study we estimated an average treatment effect (ATE), which aims to capture the mean difference between treatment and control groups in the population. Another common treatment effect of interest pertains to the effect as specific to those who actually received treatment, this effect is known as an average treatment effect on the treated (ATET). For more details on the different treatment effects one can consult the literature on the Potential

Outcomes Modeling framework (e.g.,Rubin, 2005; Heckman, 2005). Follow-up analyses assessing various treatment effects could be of interest.

Given that PeerForward seeks to influence not only the number of students completing the FAFSA early but the number of students who enroll, persist, and ultimately complete higher education, additional analyses will be applied to other outcomes that are important to the greater goal of increasing college access and completion in underserved student populations. Additional data collection concerning the presence of PeerForward in a school over time will allow for more sophisticated longitudinal analyses that will serve to enhance the effectiveness of subsequent program evaluations of PeerForward.

References

Abadie, A., & Imbens, G. W. (2009). *Matching on the estimated propensity score* (No. w15301). National Bureau of Economic Research.

Bedsworth, W., Colby, S., & Doctor, J. (2006). Reclaiming the American Dream. *Bridgespan Group*.

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications* (Vol. 11). Sage Publications.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data mining, Inference, and Prediction, 2nd edition.* New York: Springer Series in Statistics.

Heckman, J. J. (2005). 1. The Scientific Model of Causality. *Sociological methodology*, *35*(1), 1-97.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, *47*(260), 663-685.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with applications in R*. New York: Springer Series in Statistics.

Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine*, *23*(19), 2937-2960.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, *9*(4), 403.

Ridgeway, G. (2007). Generalized Boosted Models: A guide to the gbm package. *Update*, *1*(1), 2007.

Roderick, M., Nagaoka, J., Coca, V., Moeller, E., Roddie, K., Gilliam, J., & Patton, D. (2008). From high school to the future: Potholes on the road to college. Chicago, IL: Consortium on Chicago School Research at the University of Chicago.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 41-55.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322-331.

Smith, J. (2011). Can Applying to More Colleges Increase Enrollment Rates? Research Brief. *College Board*.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Appendix A: *PeerForward Logic Model*

Inputs

Program Model
- Belief in the power of positive peer influence.
- Selection process for identifying influential high school students to serve as Peer Leaders.
- PeerForward brand: evidence-based student-driven college access program.
- PeerForward materials: school signage, branded gear, student and Advisor playbooks, and online LinkForward resources.
- Technology leveraged to implement effective and scalable remote support model.

Human Capital
- Trained and effective College Summit program staff, workshop volunteers, and alumni.
- Strong Executive Team leadership and Board members.
- Effective and efficient organizational support staff.

Financial Model
- Financial model that allows for economies of scale.
- School pricing structure that ensures the program is affordable.

Strategic Partnerships
- Sustainable relationships with funders and school partners.
- Recognized as an innovative and forward thinking organization in the college access field.

Activities

Support
- College Summit staff offers ongoing support in the form of Advisor calls and monthly Peer Leader Team meetings.
- PeerForward Advisors attend a summer professional development training on college access, Peer Leadership, campaign planning, and program implementation in order to support Peer Leaders.
- Access to a nationwide network of PeerFoward Advisors and Peer Leaders to serve as thought partners in program implementation.

Peer Influence Cultivation
- Rising 12th grade Peer Leaders attend summer workshops and are trained in college knowledge, leadership, teamwork, and campaign organizing.
- Ongoing 11th and 12th grade Peer Leader coaching through fall and spring training camps focused on leadership development and campaign implementation.

Signaling
- PeerForward campaign signaling materials placed strategically in partner school buildings.

Outputs

- Number of students in grades 9-12 in partner high schools that are exposed to PeerForward campaign programming.
- Number of Advisors that are trained at summer workshops.
- PeerForward Teams implements three research-based college and career campaigns during the school year: 1) applying to three or more colleges, 2) early filing for financial aid, and 3) connecting academics to college and career.
- Number of Peer Leaders trained as change agents.
- Number of students that are exposed to college signaling.

Outcomes

Short-Term

- ★ Increase in schoolwide college-going culture.
- ✦ Increase in the number of students who complete FAFSA by March 1st.
- ✦ Increase in the number of students who submit three or more college applications.
- ✦ Increase in the number of 9th- 11th graders who complete a Postsecondary Plan.
- ★ In addition to increases in college-going predictive outcomes, Peer Leaders demonstrate increases in Power Skills, including: communication, leadership, teamwork, problem-solving, and grit.

Medium-Term

- ★ Increase in the number of students enrolling in college due to successful campaign implementation.
- ★ Increase in the number of students persisting in college due to better fit college decisions.

Long-Term

- ★ Increase in the number of students from low-income backgrounds with improved life outcomes due to college degree attainment.

Mission: College Summit transforms the lives of youth from low-income communities by connecting them to college and career.

★ = Outcome Measures

✦ = Milestone Measures

Appendix B: *Results of the sensitivity analysis across different weighting specifications.*

Results based on the standardized weight at the within-state level only.

Excluded State	*N*	*IRR*	*SE*	*p(Sig.)*	*95% CI*	
CA	170	1.26	0.134	0.030	1.02	1.55
FL	186	1.26	0.134	0.028	1.03	1.56
MD	192	1.27	0.145	0.034	1.02	1.59
MI	180	1.24	0.130	0.043	1.01	1.52
MO	192	1.30	0.141	0.017	1.05	1.60
NY	107	1.33	0.133	0.004	1.10	1.62
OH	185	1.25	0.132	0.036	1.01	1.53
SC	174	1.15	0.062	0.010	1.03	1.28

Results based on the unstandardized weight at the within-state level and the standardized mean weight at the between-state level

Excluded State	*N*	*IRR*	*SE*	*p(Sig.)*	*95% CI*	
CA	170	1.21	0.092	0.014	1.04	1.40
FL	186	1.35	0.185	0.030	1.03	1.76
MD	192	1.23	0.129	0.050	1.00	1.51
MI	180	1.20	0.127	0.087	0.97	1.47
MO	192	1.26	0.132	0.025	1.03	1.55
NY	107	1.28	0.135	0.021	1.04	1.57
OH	185	1.21	0.121	0.051	1.00	1.48
SC	174	1.12	0.052	0.014	1.02	1.23

Results based on the standardized weight at the within-state level and the standardized mean weight at the between-state level

Excluded State	*N*	*IRR*	*SE*	*p(Sig.)*	*95% CI*	
CA	170	1.27	0.149	0.046	1.00	1.59
FL	186	1.27	0.148	0.041	1.01	1.59
MD	192	1.28	0.156	0.047	1.00	1.62
MI	180	1.25	0.144	0.056	0.99	1.56
MO	192	1.30	0.157	0.028	1.03	1.65
NY	107	1.35	0.149	0.007	1.09	1.67
OH	185	1.25	0.145	0.050	1.00	1.57
SC	174	1.14	0.058	0.014	1.03	1.26

Appendix C: *Listing of Influential Variables for Each State*

California Dataset	
Type	Variable
ACT Performance	NumGE21_ACT_16
	NumTstTakr_ACT_16
	PctGE21_ACT_16
Cohort Graduation Rates	grad_cohort_prct_2012_2013
	grad_cohort_prct_2013_2014
	grad_cohort_prct_2014_2015
English Language Learner Status	TotalELLStudents
Enrollment Counts	Enroll12_ACT_16
	EnrollmentK12_2017
	SumofGR_12_2016
	SumofGR_12_2017
FAFSA Completion	FAFSA_12_30_16
	FAFSA_6_30_16
	FAFSA3_3_16
Income Status	FreeMealCountK12_2017
	FRPMCountK12_2017
Prior Academic Achievement on State Standardized Tests	ELA_Y1_2015
	ELA_Y2_2015
	MATH_Y1_2015
	MATH_Y2_2016
	SCI_Y1_2014
	SCI_Y3_2016
	SSCI_Y2_2015
Race/Ethnicity	AfricanAmericanNotHispanic
	AmericanIndianorAlaskaNative
	AsianNotHispanic
	FilipinoNotHispanic
	HispanicorLatino
	PacificIslanderNotHispanic_2
	NotReported_race_2017
	WhiteNotHispanic
	TwoorMoreRacesNotHispanic
SAT Performance	AvgScoreMath_SAT_2016
	AvgScrERW_SAT_2016
	PctCCR_Benchmark_SAT_2016
	NumCCR_Benchmark_SAT_2016
	NumTstTakr_SAT_2016

Florida Dataset	
Type	Covariate
Cohort Graduataion Rates	GraduationRate201415
	GraduationRate201314
Enrollment Counts	Grade_12_2017
	Grade_12_2016
	Grade2016
	# of Students_2017
FAFSA Completion	FAFSA_6_30_16
	FAFSA_12_30_16
	FAFSA_3_3_16
Income Status	# of Free Lunch Students_2017
	# of Reduced price Lunch Students_2017
	PercentofEconomicallyDisadvan
Prior Academic Achievement on State Standardized Tests/Sate Grading Systems	CollegeandCareerAcceleration
	EnglishLanguageArtsAchievemen
	EnglishLanguageArtsLearningG
	MathematicsAchievement_2016
	MathematicsLearningGains_2016
	InfoBaseGrade2016
	ScienceAchievement_2016
	SocialStudiesAchievement_2016
	PercentTested_2016
	PercentofTotalPossiblePoints
	TotalPointsEarned_2016
Race/Ethnicity	Asian_2017
	BlackorAfricanAmerican_2017
	HispanicLatino_2017
	White_2017
	PercentofMinorityStudents_201
	TwoorMoreRaces_2017
School Type	CharterSchool
	TitleI

Maryland Dataset	
Type	Variable
Graduation Rates	EarnDiplom164
	EarnDiplom154
	AdjCohortCount153
Enrollment Counts	Seniors16
	Seniors17
	StudentCount_2015
FAFSA Completion	FAFSA_12_30_16
	FAFSA_6_30_16
	FAFSA3_3_16
Prior Academic Achievement on State Standardized Tests	Level2pctELA10
	Level4pctELA10
	Level4ctELA10
	Level3pctELA10
	Level3ctELA10
	Level2pctAlg1
	Level3ctAlg1
	Level1ctAlg1
	Level5ctELA10
	Level3pctAlg1
	Level4pctAlg1
	Level1pctAlg1
	Level2ctELA10
	Level1pctELA10
	Level5pctELA10
	Level1ctELA10
	Level2ctAlg1
	Level1ctAlg2
	Level3ctAlg2
	Level2ctAlg2
	TestedCountELA10
	TestedCountAlg1
Race/Ethnicity	Black
	Hispanic
College Enrollment	Number of 2015 senior class enrolled in college

Michigan Dataset	
Type	Variable
Cohort Graduation Rates	GRADUATION_RATE_5_YEAR_2015
	DROPOUT_RATE_5_YEAR_2015
	GRADUATES_5_YEAR_2015
	DROPOUTS_5_YEAR_2015
	CONTINUING_5_YEAR_2015
	OTHER_COMPLETER_5_YEAR_2015
	GRADUATION_RATE_6_YEAR_2014
	DROPOUT_RATE_6_YEAR_2014
	GRADUATES_6_YEAR_2014
	DROPOUTS_6_YEAR_2014
	OTHER_COMPLETER_6_YEAR_2014
	GRADUATION_RATE_4_YEAR_2016
	DROPOUT_RATE_4_YEAR_2016
	GRADUATES_4_YEAR_2016
	DROPOUTS_4_YEAR_2016
	CONTINUING_4_YEAR_2016
	OTHER_COMPLETER_4_YEAR_2016
English Language Learner Status	ENGLISH_LANGUAGE_LEARNERS_ENROLL_2017
Enrollment Counts	Grade_12_ENROLLMENT_2016
	GRADE_12_ENROLLMENT_2017
	TOTAL_ENROLLMENT_2017
	MALE_ENROLLMENT_2017
	FEMALE_ENROLLMENT_2017
FAFSA Completion	FAFSA_12_30_16
	FAFSA_6_30_16
	FAFSA3_3_16
Income Status	ECONOMIC_DISADVANTAGED_ENROLLMENT_2017
Race/Ethnicity	AMERICAN_INDIAN_ENROLLMENT_2017
	ASIAN_ENROLLMENT_2017
	AFRICAN_AMERICAN_ENROLLMENT_2017
	HISPANIC_ENROLLMENT_2017
	WHITE_ENROLLMENT_2017
	TWO_OR_MORE_RACES_ENROLLMENT_2017
Special Education Status	SPECIAL_EDUCATION_ENROLLMENT_2017

Missouri Dataset		
Type	Variables	
ACT Performance	ACT_TESTS_ADMINISTERED_2016	ACT_COMPOSITE_SCORE_2016
	GRADUATES_WITH_ACT_SCORE_ABOVE_N	ACT_SCIENCE_SCORE_2016
	ACT_ENGLISH_SCORE_2016	ACT_READING_SCORE_2016
	GRADUATES_2016_ACT	ACT_MATH_SCORE_2016
Cohort Graduation Rates	GRADUATION_RATE_5YR_COHORT_2014	GRADUATION_RATE_5YR_COHORT_2015
	ADJUSTED_5YR_COHORT_2014	ADJUSTED_5YR_COHORT_2015
	GRADUATES_2016	GRADUATES_4YR_COHORT_2016
	ADJUSTED_5YR_COHORT_2016	GRADUATES_5YR_COHORT_2014
	GRADUATES_5YR_COHORT_2015	GRADUATION_RATE_4YR_COHORT_2016
	GRADUATES_4YR_COHORT_2014	GRADUATES_4YR_COHORT_2015
	ADJUSTED_4YR_COHORT_2014	GRADUATES_5YR_COHORT_2016
	GRADUATION_RATE_4YR_COHORT_2015	GRADUATION_RATE_4YR_COHORT_2014
	ADJUSTED_4YR_COHORT_2016	GRADUATION_RATE_5YR_COHORT_2016
	ADJUSTED_4YR_COHORT_2015	
College Enrollment Rates	GRADUATE_FOLLOWUP_COLLEGE_CER_20	GRADUATE_FOLLOWUP_2YR_PCT_CER_20
	GRADUATE_FOLLOWUP_COLLEGE_PCT_20	GRADUATE_FOLLOWUP_4YR_PCT_CER_20
	GRADUATE_FOLLOWUP_4YR_CER_2015	GRADUATE_FOLLOWUP_4YR_PCT_2016
	GRADUATE_FOLLOWUP_4YR_CER_2014	GRADUATES_PREVIOUS_YEAR_CER_2016
	GRADUATE_FOLLOWUP_4YR_2016	GRADUATE_FOLLOWUP_2YR_CER_2014
	GRADUATE_FOLLOWUP_4YR_PCT_CER_2	GRADUATE_FOLLOWUP_2YR_CER_2015
	GRADUATES_PREVIOUS_YEAR_CER_2015	GRADUATE_FOLLOWUP_COLLEGE_PCT_CE
	GRADUATES_PREVIOUS_YEAR_CER_2014	GRADUATE_FOLLOWUP_2YR_2016
	GRADUATE_FOLLOWUP_2YR_PCT_2016	
English Language Learner Status	ENROLLMENT_ELL_LEP	ENROLLMENT_ELL_LEP_PCT
Enrollment Counts	ENROLLMENT_GRADES_12_2016	GRADUATES_2014
	ENROLLMENT_GRADES_K_12	GRADUATES_2015
	ENROLLMENT_GRADES_12_2017	ENROLLMENT_GRADES_9_12_2017
	JANUARY_MEMBERSHIP	
FAFSA Completion	FAFSA_6_30_16	FAFSA3_3_16
	FAFSA_12_30_16	
Income Status	LUNCH_COUNT_FREE_REDUCTED_PCT	LUNCH_COUNT_FREE_REDUCED
Race/Ethnicity	ENROLLMENT_ASIAN	ENROLLMENT_HISPANIC
	ENROLLMENT_ASIAN_PCT	ENROLLMENT_HISPANIC_PCT
	ENROLLMENT_BLACK	ENROLLMENT_MULTIRACIAL
	ENROLLMENT_BLACK_PCT	ENROLLMENT_MULTIRACIAL_PCT
	ENROLLMENT_INDIAN_AMERICAN_PCT	ENROLLMENT_WHITE
	ENROLLMENT_INDIAN_AMERICAN	ENROLLMENT_WHITE_PCT
Special Education Status	IEP_SCHOOLAGE_CHILDCOUNT	

New York Dataset		
Type	Variable	
Graduation Rates	REG_ADV_CNT_Grad_2012_4_yr	STILL_ENR_CNT_Grad_2010_6_yr
	GRAD_CNT_Grad_2010_4_yr	DROPOUT_CNT_Grad_2010_4_yr
	REG_CNT_Grad_2012_4_yr	REG_CNT_Grad_2011_5_yr
	GRAD_CNT_Grad_2011_5_yr	LOCAL_CNT_Grad_2012_4_yr
	ENROLL_CNT_Grad_2012_4_yr	STILL_ENR_CNT_Grad_2011_5_yr
	REG_ADV_CNT_Grad_2009_6_yr	LOCAL_CNT_Grad_2009_6_yr
	ENROLL_CNT_Grad_2010_6_yr	LOCAL_CNT_Grad_2011_5_yr
	ENROLL_CNT_Grad_2011_5_yr	GRAD_CNT_Grad_2010_5_yr
	REG_ADV_CNT_Grad_2010_5_yr	STILL_ENR_CNT_Grad_2009_6_yr
	REG_ADV_CNT_Grad_2010_4_yr	GED_CNT_Grad_2009_6_yr
	REG_ADV_CNT_Grad_2011_5_yr	DROPOUT_CNT_Grad_2010_6_yr
	REG_CNT_Grad_2010_4_yr	GED_CNT_Grad_2010_6_yr
	LOCAL_CNT_Grad_2010_6_yr	GED_CNT_Grad_2010_4_yr
	REG_CNT_Grad_2010_5_yr	GED_CNT_Grad_2012_4_yr
	STILL_ENR_CNT_Grad_2012_4_yr	LOCAL_CNT_Grad_2010_4_yr
	ENROLL_CNT_Grad_2010_5_yr	GRAD_CNT_Grad_2012_4_yr
	LOCAL_CNT_Grad_2010_5_yr	REG_CNT_Grad_2009_6_yr
	DROPOUT_CNT_Grad_2012_4_yr	GRAD_CNT_Grad_2010_6_yr
	NON_DIPLOMA_CREDENTIAL_CNT_Grad_	ENROLL_CNT_2009_6_yr
	GRAD_CNT_Grad_2009_6_yr	STILL_ENR_CNT_Grad_2010_4_yr
	STILL_ENR_CNT_Grad_2010_5_yr	DROPOUT_CNT_Grad_2009_6_yr
English Language Learner Status	EnglishLanguageLearner_2017	Not English Language Learner
Enrollment Counts	GRADE12_2017	PK12TOTAL_2017
	GRADE12_2016	
FAFSA Completion	FAFSA_12_30_16	FAFSA 3_3_16
	FAFSA_6_30_16	
Income Status	Not Economically Disadvantaged	EconomicallyDisadvantaged_2017
Prior Academic Achievement on State Standardized Tests	PI_G_RATE_ELA_2015	NUM_PERF_Math_2014
	PI_G_RATE_ELA_2014	NUM_PERF_ELA_2015
	NUM_PERF_Math_2015	NUM_ENROLL_ELA_2015
	NUM_ENROLL_ELA_2014	NUM_PARTIC_ELA_2015
	NUM_PARTIC_Math_2015	CURRENT_SH_TARGET_ELA_2014
	NUM_PARTIC_ELA_2014	PI_G_RATE_Math_2015
	NUM_ENROLL_Math_2014	PER_PARTIC_ELA_2015
	CURRENT_SH_TARGET_ELA_2015	CURRENT_SH_TARGET_Math_2015
	CURRENT_SH_TARGET_Math_2014	NUM_ENROLL_Math_2015
	AMO_STAND_ELA_2014	PI_G_RATE_Math_2014
Race/Ethnicity	AmericanIndianAlaskaNative_Ra	Hispanic_Race_2017
	AsianPacificIslander_Race_2017	Multiracial_Race_2017
	Black_Race_2017	White_Race_2017
Special Education Status	GeneralEducationStudents_2017	StudentswithDisabilities_2017

Appendix D: *Standardized Results from the State-by-State analyses*

State	*N*	*IRR*	*SE*	*p(Sig.)*	*95% CI*	
CA	28	1.21	0.121	0.057	0.99	1.47
FL	12	1.17	0.191	0.350	0.85	1.61
MD	6	1.14	0.057	0.007	1.04	1.26
MI	18	1.54	0.338	0.050	1.00	2.37
MO	6	1.05	0.059	0.392	0.94	1.17
NY	91	1.02	0.155	0.891	0.76	1.37
OH	13	1.38	0.268	0.094	0.95	2.02
SC	24	1.85	0.503	0.024	1.09	3.15