Evaluation Studies

A GFSF Matched Schools Study

January 2021

Russell W. Rumberger, UCSB, Jay Plasman, *Ohio State University*

Highlights

This study used a quasi-experimental design to estimate whether schools that adopted the GFSF program had a stronger impact on graduation rates, A-G course completion rates, and suspension rates than similarly matched schools that did not adopt the program.

Comparing four different groups of GFSF schools and five, highly-regarded individual GFSF high schools yielded mixed results:

- Only one comparison produced a positive and meaningful program effect in graduation rates: the graduation rate of the 2009-10 freshman cohort at Carpinteria High School was 6.7 percentage points higher than would have been expected had the school not offered the GFSF program.
- Five of nine comparisons produced positive program effects in A-G completion rates; for example, students who attended schools that implemented the GFSF program with high fidelity had higher improvements in A-G completion rates than students who attended similar schools that did not provide the program.
- All five comparisons produced positive program effects in suspension rates, although only two of the effects were robust.

Introduction

Get Focused…Stay Focused!™ (GFSF) is a high school program designed to develop the skills and knowledge that lead to high school graduation, college readiness and completion, and successful entry into the workforce. Despite strong theoretical underpinnings, some supporting empirical research, and extensive testimony from school officials and students about the value of the program, the GFSF program has never been subjected to a rigorous evaluation. To address this need, two ongoing evaluation studies have been undertaken to assess the implementation of the GFSF program and its impact on a series of student outcomes related to performance and success in high school. The first study employed an experimental design in which 20 high schools from throughout California were randomly assigned to either a treatment group or a control group, comparing 9th grade students in the treatment (cohort 1) schools that received the program in the 2016-17 school year with 9th grade students in the control (cohort 2) schools who did not receive the program in the 2016-17 school year. This brief summarizes the activities and findings from the second study, which relies on school-level administrative data collected by the California Department of Education.

Research Design

The second study utilizes a quasi-experimental design called a Comparative Interrupted Time Series (CITS). These designs have been used extensively to evaluate interventions in such areas as epidemiology, employment programs, and education interventions.[1] In a CITS design, program impacts are evaluated by looking at whether schools that implemented an intervention (in this case, the GFSF program) "deviated" from their baseline trends by a greater amount than a group of similar comparison schools. This provides an indication of whether schools that adopt GFSF have a stronger impact on students' outcomes than schools that do not adopt the program. Studies have shown that CITS designs can, in some cases, produce results similar to those produced through randomized experiments.[2]

The first step in conducting a CITS design is to create two samples of matched schools: (1) treatment schools that adopted the GFSF program and (2) comparison schools that did not adopt

[1] Kemple, J. J., Herlihy, C. M., & Smith, T. J. (2005). *Making progress toward graduation: Evidence from the Talent Development High School Model*. New York: MDRC; Dee, T. S., & Jacob, B. (2011). The Impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management, 30*(3), 418-U460. doi:10.1002/pam.20586

[2] St.Clair, T., Cook, T. D., & Hallberg, K. (2014). Examining the Internal Validity and Statistical Precision of the Comparative Interrupted Time Series Design by Comparison With a Randomized Experiment. *American Journal of Evaluation, 35*(3), 311-327. doi:10.1177/1098214014527337

the GFSF program. We identified 33 high schools in California that adopted the GFSF program in the five-year period from 2009-10 through 2014-15. In order to identify comparable matched schools, we collected school-level data from the California Department of Education website on a series of student outcome measures:

1. High school graduation rates
2. A-G completion rates for high school graduates
3. Suspension rates

The data were collected over a number of years before and after the first year of program implementation, as shown in Table 1. For example, four schools first implemented the GFSF program in 2009-10. For those schools, we collected student outcome measures for the two years before the program was first implemented—2007-08 and 2008-9—and the four years after the program was first implemented—2009-10, 2010-11, 2011-12, and 20012-13. Since the main component of the program is taught in the ninth grade, the first year to judge its impact on graduation and A-G completion rates is four years later. For the students who took the freshman course in 2009-10, the first year to judge its impact on graduation and A-G completion rates is 2012-13.

Table 1
Year of Implementation for GFSF Program and Observation Years for Collecting Data

Implementation year	**Observation years**					
2009-10 (4 schools)	2007-08	2008-09	**2009-10**	2010-11	2011-12	2012-13
2010-11 (1 school)	2008-09	2009-10	**2010-11**	2011-12	2012-13	2013-14
2011-12 (2 schools)	2009-10	2010-11	**2011-12**	2012-13	2013-14	2014-15
2013-14 (10 schools)	2011-12	2012-13	**2013-14**	2014-15	2015-16	2016-17
2014-15 (16 schools)	2012-13	2013-14	**2014-15**	2015-16	2016-17	2017-18
	1	**2**	**3**	**4**	**5**	**6**

In addition to our outcome variables, we also collected data on a number of school-demographic variables for year 2, the year before the program was introduced:

1. Size (number of students)
2. English Language Learners (%)
3. Free/Reduced Price Lunch (%)
4. Underrepresented Minority Students (% identifying as Latina/o, African American, Native American, or two or more races)

To match GFSF schools with comparison schools, we first created separate groups of schools based on the year of implementation. Based on this process, we ended up with four different datasets that each included the schools that implemented GFSF as well as the full list of control schools with available information from that same year.[3] Note that the 2010-11 GFSF school dropped from our analyses since it did not have any available outcome data. Additionally, there were no schools that implemented the program in the 2012-13 school year. There were also a small number of schools in other years that were dropped for this same reason, resulting in a final sample of 29 schools. After creating these datasets, we next estimated trend lines for each school using the observed outcome measures for years 1-5. Schools were then matched using a propensity score matching technique based on the four school-demographic variables identified above, the baseline outcome in the year prior to program implementation (year 2) and the baseline trend in mean outcomes years 1-5. Under this matching technique, each school is given a propensity score for having adopted the GFSF program based on the selected variables. This propensity score can take any value between 0 (no chance of adoption) to 1 (certain adoption). Using these propensity scores, the treatment schools (GFSF adopters) were matched with control schools (non-adopters) that had the closest propensity scores. We chose to match each treatment school with

[3] Some schools were deleted because of incomplete data.

five control schools in order to increase precision in our estimates without introducing too much bias by matching with dissimilar schools. Ultimately by limiting our comparison to schools that are similar on a set of observed variables, we are attempting to simulate the effects of a random assignment study that would provide the most convincing evidence of a true effect.

The inclusion of school-level demographic variables allows us to be confident that we are comparing similar schools. For example, we want to be sure we don't match schools that have more than 3000 students to those with fewer than 300 students or some schools in which more than 75% of students received free/reduced price lunch to those with less than 25% free/reduced price lunch. We include the baseline measure of the outcome (i.e., the outcome the year before the program was implemented) to be sure we are looking at schools that started at relatively similar places as it would not be practical to compare a school with a baseline 85% graduation rate to one with 30% graduation rate. Finally, we included a measure identifying the overall trend to ensure we wouldn't potentially compare schools that had declining outcome measures to those with increasing measures as this could introduce an additional source of bias in our estimates.

To determine program impact, we first compared the actual outcome with the predicted outcome based on the trend line for both GFSF schools and comparison schools, referred to as the deviation from the baseline trends. The program impact is simply the difference between the average deviation from the baseline trend for GFSF schools and the average deviation from the baseline trend for comparison schools.

The deviation from the baseline trend for comparison schools is a critical part of this analysis because it reveals the likely outcomes for GFSF schools had they not implemented the program. Put another way, it provides a powerful "counterfactual" to more accurately measure the improvement in school performance due the adoption of the GFSF program above and beyond the impact of other reforms that the school could have adopted.

An Illustration

To illustrate how this technique works, we'll estimate the impact of the GFSF program on A-G completion rates for a subset of nine schools that implemented all four years of the program. While the main segment of the curriculum is offered as a stand-alone course in the freshman year, there are 3 available modules that are designed to be offered as individual units within an existing course. Therefore, this illustration compares schools that implemented both the freshman year course and the three follow-up modules to comparable schools that did not offer the GFSF program at all.

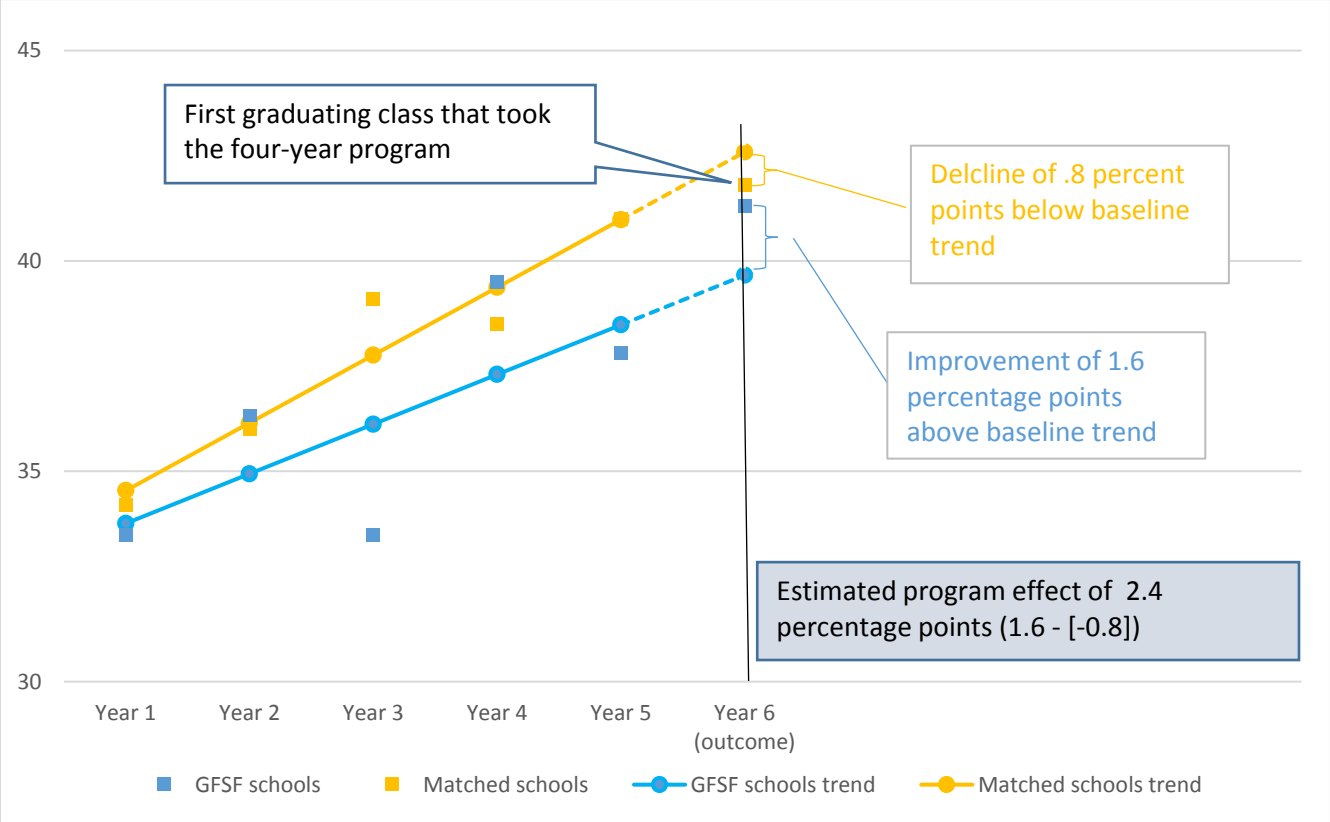
Table 2 shows the mean characteristics of the nine GFSF schools and a group of 39 comparison schools. All the matching variables except the baseline trend are measured the year before (year 2) the GFSF was implemented. The matching technique produced a sample of comparison schools that closely matched the GFSF schools on all of the variables. In particular, the A-G completion rate the year before the program was implemented was 36.3% for the GFSF schools and 36.0% for the comparison schools. The baseline trends were also similar: a 1.2 percentage point increase per year for the GFSF schools and 1.6 percentage points for the comparison schools.

Table 2
School Characteristics for GFSF Schools that Taught the Modules and Comparison Schools

	GFSF Schools (N=9)	Comparison Schools (N=39)
Enrollment	1196	1090
ELL (%)	9.4%	9.5%
FRL (%)	56.2%	64.4%
Minority (%)	61.2%	67.4%
Matched AG Rate	36.3%	36.0%
Baseline trend	1.2	1.6

The baseline trends for both groups of schools are shown in Figure 1 along with the estimated and actual values in year 6, the first graduating class that started the program four years earlier. The estimated A-G completion rate for the GFSF schools based on the baseline trend was 39.7% while the actual rate was 41.3%. Thus, GFSF schools performed 1.6 percentage points better than expected based on their past performance. The estimated A-G completion rate for the comparison schools was 42.6% while the actual rate was 41.8%. Thus, the comparison schools performed .8 percentage points worse than expected based on this past performance. The differences in the deviation from baseline trends yields an estimated program effect of 2.4 percentage points (1.6% - [-0.8]). Although this result only applies to this sample of schools, it does suggest that schools adopting the complete four-year GFSF program were able to improve their A-G completion rate by an average of 2.4 percentage points as compared to non-GFSF schools.

Figure 1
Percentage of 12th Grade Students Completing all UC/CSU Entry Courses (A-G)



A Summary of Results

To estimate the impact of the GFSF program, we created a number of matched GFSF school groups:

1. All schools that started the program over the five-year period from 2009-10 thru 2014-15 (N=33);
2. All schools that started the program over the two-year period from 2013-14 thru 2014-15, since those schools may have benefited from a more robust version of the GFSF program that included use of instructional models in grades 10-12 (N=24);
3. All schools that received medals[4] for at least two years, since those schools may have been more likely to implement the program with high fidelity (N=9);

[4] Academic Innovations certifies "medal" schools that implement the program with high fidelity. See: http://www.whatworkscareerchoices.com/medal1.html

4. All schools that taught the freshman course and the three modules in grades 10-12 (N=10).

We also created snapshots for a number of individual schools that program developers considered as exemplary programs:

1. Carpinteria High School (Carpinteria Unified School District), which began the program in 2009-10;
2. Desert Hot Springs High School (Palm Springs Unified School District, which began the program in 2009-10;
3. Indio High School (Desert Sands Unified School District), which began the program in 2009-10;
4. Bishop High School (Bishop Unified School District), which began the program in 2014-15
5. Arvin High School (Kern Union High School District), which began the program in 2013-14.

The results of these analyses are shown in a series of tables that examine a series of student outcomes. The first table shows the results for four-year graduation rates (Table 3). Across most of our analyses, GFSF schools exhibit higher outcome graduation rates than their matched schools. For example, among the sample of all schools (N=29), the four-year graduation rate is 92.6 percent compared to 90.1 percent for the matched school sample (N=135). The only exception is the for the sample of GFSF schools that taught the modules (N=10). Their four-year graduation rate is 92.5 percent compared to 96.2 percent for the matched schools (N=48).

To estimate the effect of the GFSF program we focus on how the outcome varies from the trend; that is, how much the actual graduation rate deviates from the trend or the expected graduation rate. In many samples, the deviation from the trend line is negative, which means the actual graduation rate is lower than the expected graduation rate. But what matters is how the deviation from the trend for GFSF schools compares to the deviation from the trend of the matched schools. If the deviation from the trend for the GFSF schools is larger than the deviation from the trend for the matched sample, then we can conclude that the GFSF program had a positive effect. The results show that only one sample, for Carpinteria high school, produced a positive program effect. The graduation rate of 91 percent for Carpinteria high school was 3.5 percentage points above its trend line, while the graduation rate of 78.5 percent for the matched schools was -3.2 percent points below its trend line. As a result, the estimated program effect of 6.7 percentage points. This means that ninth grade students who participated in the GFSF program at Carpinteria High School in 2009-10 had graduation rates that were 6.7 percentage points higher than if they had attended a similar high school that did not offer the program.

Table 3
4-Year Cohort Graduation Rates

	GFSF Schools			Matched Schools			
	Number	Outcome	Deviation from trend	Number	Outcome	Deviation from trend	Estimated program effect
All Schools	29	92.6	-1.1	135	90.1	-0.4	-0.7
Recent adopters	24	92.6	-1.5	120	92.2	0.8	-2.3
Medal schools	9	93.5	-0.4	45	88.5	1.6	-2.0
Schools that taught modules	10	92.5	-1.7	48	96.2	0.7	-2.3
Carpinteria High School	1	91	3.5	5	78.5	-3.2	6.7
Desert Hot Springs	1	83	2.7	5	74.5	5.6	-2.9
Indio High School	1	96.5	-4.6	5	82	0.3	-4.9
Bishop High School	1	91.2	1.4	5	95.6	2	-0.6
Arvin High School	1	99.3	-1.5	5	89.8	-1.7	0.2

The next table shows the results for completion of A-G courses required for entry to UC and CSU (Table 4). These results are generally more favorable. Medal schools (N=7) and schools that taught the modules (N=9) improved their A-G completion rates more than their estimated rates based on trends, whereas their matched schools had lower completion rates than their estimated rates based on trends. As a result, both samples of GFSF schools had positive program effects. This means that students who attended schools that implemented the GFSF program with high fidelity had higher improvements in A-G completion rates than students who attended similar schools that did not provide the program. Similarly, students who attended schools that implemented all four years of the program had higher improvements in A-G completion rates than students who attended similar schools that did not provide the program.

Among individual schools, Carpinteria, Indio, and Bishop Highs Schools all had positive program effects, while two high schools, Desert Hot Springs and Arvin, had negative program effects.

Table 4
A-G Completion Rates

	GFSF Schools			Matched Schools			
	Number	Outcome	Deviation from trend	Number	Outcome	Deviation from trend	Estimated Program Effect
All Schools	27	38.6	1.8	131	38.6	2	-0.3
Recent adopters	23	39.7	1.6	111	38.9	2.3	-0.7
Medal schools	7	38.3	1.6	35	45	-0.9	2.5
Schools that taught modules	9	41.3	1.6	39	41.8	-0.8	2.4
Carpinteria High School	1	46.8	9.1	29	38.6	-8.4	17.6
Desert Hot Springs	1	25.4	8.2	2	25.3	15.6	-7.4
Indio	1	25.4	-3.4	17	23.7	-8.9	5.5
Bishop	1	51.8	-5.7	2	56	-7.8	2.1
Arvin	1	30.5	0.9	2	38.2	9.1	-8.2

The final table shows results for suspension rates. In this analysis there are fewer matches because suspension data were only available beginning in 2011-12. The results show positive program effects in all three group samples, although the magnitude of the effects was small except for the sample of medal schools that showed a program effect of 2.4 percentage points (N=6). One high school, Bishop, also showed a robust program effect of 8 percentage points.

Table 5
Suspension Rates

	GFSF Schools			Matched Schools			
	Number	Outcome	Deviation from trend	Number	Outcome	Deviation from trend	Estimated Program Effect
All Schools	*Same as Recent Adopters due to availability of suspension data*						
Recent adopters	24	6	-0.3	118	6	-0.4	0.1
Medal schools	6	2.2	2.3	29	3.1	-0.1	2.4
Schools that taught modules	10	5.1	0.5	52	3.2	-0.3	0.8
Carpinteria High School							
Desert Hot Springs	*Suspension rates unavailable for analyses of early schools*						
Indio							
Bishop	1	0	7.9	5	1.2	-0.1	8
Arvin	1	6.8	-0.1	5	0.2	-0.1	0

Note: A positive value in the 'Deviation from trend' or 'Estimated Program Effect' column indicates a decrease in suspension rate, which is the desired outcome (i.e., a positive outcome)

Conclusion

This study uses a quasi-experimental research design to estimate the impact of the GFSF program on three student outcome measures: 4-year cohort graduation rates, A-G course completion rates, and suspension rates. The research design matched various samples of GFSF schools with schools with similar student populations and similar pre-program outcomes and trends. This helps determine whether schools implementing the GFSF program improved student outcomes compared to similar schools what did not implement the program. The overall conclusion is that the program did not consistently show positive program impacts. In some cases, there were positive effects and in others there were not. For example, schools that taught the program for all four years, meaning they taught the freshman course and all the modules in grades 10-12, improved their A-G completion rates by 2.4 percentage points above expected, but they lowered their expected graduation rates by 2.3 percentage points. One possible explanation for the negative program effects associated with graduation rates is that the graduation rates at these schools were already quite high, meaning that there was not much room for improvement. Therefore, graduation rate may not be the best metric by which to measure the effectiveness of the GFSF program. The same could be said for suspension rates, which are relatively low as well. This further emphasizes the need to focus on multiple outcomes and highlights the benefits in A-G completion rates for schools that have high fidelity of implementation or implement the full program.

A couple of individual schools produced more consistent results. Carpinteria High School produced positive effects on both graduation rates and A-G completion rates. This is even more remarkable considering program modules were not available when the school started the program, so these impacts are based solely on the freshman course (and perhaps other reforms the school may have implemented during the same period). Bishop High School showed positive effects of A-G completion rates and suspension rates, but a small decline in graduation rates.

The varying results suggest that the Get Focused, Stay Focused program can produce positive improvements in student outcomes, but not consistently. The effects appear to be heavily reliant on the school context and may be driven by factors beyond the scope of our evaluation (e.g., simultaneous implementation of other comparable programs, involved leadership, one extraordinary GFSF teacher, etc.). Of course, this study only examined three student outcomes. Additional research could examine impacts on other outcomes, such as college persistence, grades, and completion rates.