2019-2020 Funded Site
Graduation Rate Evaluation

Stephanie Chasteen (External Evaluator)
with analyses by Joseph Taylor

January 18, 2020 (finalized April 1, 2020; substantially revised December 11, 2020)
SUMMARY REPORT

About the study

One page visual

Findings

Recommendations

FULL REPORT

Methods

Findings: Descriptive statistics
- Graduation rates across all sites
- Comparison across funding tracks
- Sustainability and net long-term impacts
- Cost-benefit ratio analysis

Findings: Regression studies
- Best definition of department size
- How was “impact period” defined?
- What was the analytic approach?
- Predictors of graduation rates
- Sustainability (Regression study)

Recommendations
- Portfolio of funding tracks
- Screening applications for funding
- Supporting funded sites
- Tracking and calculating graduation numbers
- Next steps for disseminating and using the work

Appendix 1: Cost-benefit analysis

Appendix 2: Study of Physics Teacher Graduate Production in PhysTEC Sites
Graduation rates increase on average across all sites

**Increases during funding**
- +0.9 average increase in number of physics teacher graduates* (p<0.001).
- 70% of sites increased the number of physics teacher graduates during funded period
- +1.5 average increase in graduates at those sites with gains (p<0.001)

**Long-term increases post funding**
- +0.7 average increase in number of physics teacher graduates long-term pre-to-post (p<0.001)
- 59% of sites increased the number of physics teacher graduates long-term pre-to-post funding
- +1.8 average increase in graduates at those sites with gains (p<0.001)

*Strongly limited by pre-funding graduation rates, but not related to department size

**More gains for Comprehensives**

**Robust long-term increases for Comprehensives**
- Average pre-post increase in teachers at Comprehensive sites.
  - +1 of Comprehensive sites increased the number of physics teacher graduates
  - +1.7 average increase in graduates at Comp. sites with long-term gains (p<0.001)

**Small grants cost-efficient**
- $1664 cost per year of teaching for Recruiting grants (1/3 of that at Comprehensives.)

**Methods**
Analyses completed with “Funding +1” period which includes funded period plus one year post-funding, to account for lagging indicators. Graduates are rounded to nearest whole graduate. See full report.

Sites included were all N=43 sites funded since 1998 (26 Comprehensive, 8 Targeted, 9 Recruiting). In analyses addressing the sustainability of the grants, these were reduced to include only those with information in the sustainability period (N=25, 8, and 7 respectively.)
SUMMARY REPORT

About the study

I engaged statistician Joseph Taylor to conduct systematic analyses of physics teacher graduation rates at funded sites since 1998. Graduation rates were available through 2019-2020. The purpose was to allow interpretation of the impact of different funding streams, as well as planned studies of the sustainability of Targeted and Recruiting grants. Later, I conducted additional descriptive analyses of graduate numbers.

Evaluation questions

1. What is the impact of funding, and funding track, on the number of physics teachers?
2. How do background variables (department size, prior graduation rates) affect the number of physics teachers?
3. How are these changes sustained?
4. What are good procedures for PhysTEC in data collection and monitoring to track impact in the future?

Analytical approach

This work used descriptive statistics to provide comparisons between funding tracks, and a regression model tested the effect of different variables (department size, pre-funding graduation rates, and funding track) on graduation rates. Limitations of this study include small numbers of sites, and of physics teacher graduates, often limiting the ability to test for significance. Another limitation is that different sites have data for different numbers of years.

Whole number graduates

Because of small number fluctuations, for this report I define “+1 graduate” as being a gain of between 0.5 and 1.5 graduates. In this way, the number of graduates is rounded to the nearest graduate, and gains of less than 0.5 graduates are considered to be 0 gain, helping to account for floor effects. This is a reasonable approach, but not backed up by any specific analytics.
Sample size

Sites included were all N=43 sites funded since 1998 (26 Comprehensive, 8 Targeted, 9 Recruiting). In analyses addressing the sustainability of the grants, these were reduced to include only those with information in the sustainability period (N=25, 8, and 7 respectively).

Time period definition: “Funded plus one” and “impact”

All graduation rates were analyzed with respect to when funding began (i.e., “funding year 1”) allowing comparison of impacts across different cohorts. We use two time periods of interest, due to changes in approach during analysis:

1. **Funded period plus one.** This is the funded period, plus the first post-funding year; Year 1 of funding through Year 1 post-funding. This is usually 4 years, and captures the fact that graduates in post-funding Year 1 were likely recruited during the funded period.

2. **Impact Period.** Joe Taylor used an “impact period” for the regression studies, which is shifted forward by one year compared to the “funded plus one” period: Year 2 of funding through Year 2 of post-funding. This is a more conservative estimate of lagging indicators.

*Figure 1 (left):* Graduates averaged across all sites for pre-funded (F-3- F-1), funded (F1-F4), and post-funded (PF1-PF3) years. This chart displays only years for which data exists for most sites: F4 is colored light grey to indicate only some sites had a 4th year of funding.

*Figure 2 (right):* Graduates averaged across all sites for pre-, during- and post- periods for the three different possible analysis periods. In each case, only sites for which post-period data exists are shown, and data are averaged across all available data.

The choice of the period has implications. As shown in Figure 2 above because graduation rates continued to climb during PF1 and PF2, where those years are categorized has impacts on the “during” and “post” numbers. Choosing “Funded +1” provides a reasonable compromise between two extremes based on an understanding of what occurs at funded sites.
Findings

Findings regarding graduation rates

PhysTEC funding is associated with significant gains in graduation rates during funding, and most sites experience gains, across all tracks. The average gain in physics teacher graduates was nearly one graduate (+0.9 graduates); this gain is statistically significant (p<0.001). In the later regression study (using the “impact” period), the significant effect remained even when department size and pre-funding teacher graduation rates are taken into account. Most funded sites (N=30; 70%) experience gains of at least 0.5 teachers. Among the 70% with gains, the average gain is +1.4 graduates. We can estimate that ~1 site’s gains are due to floor effects (when starting at 0 teachers, the rate can only increase).

Comprehensive sites generate greater gains during funding than other funding tracks, but all funding tracks demonstrate success. Comprehensive sites experience greater gains in numbers of teacher graduates (+1.1 graduates), and a greater fraction of sites experience gains (76%) than other funding tracks, and the increase for those sites with gains is greater for Comprehensive sites (+1.7 graduates) than for other types of sites (+1.1 graduates). However, all funding tracks experience positive gains from pre-to-during funding, and more than half of all sites experience positive gains. We considered what exemplary performance looks like for different funding tracks; For Comprehensive sites, it is unusual to graduate as many as 4-5 teachers/year (or to increase graduation numbers by 2-3 teachers/year), and for Targeted sites, it is unusual to graduate as many as 2-3 teachers/year (or to increase graduation numbers by 1 teacher/year). Thus, these might be considered as teacher graduation rates that merit special recognition.
The number of teachers graduated prior to PhysTEC funding has a strong influence on the numbers, and gain, during PhysTEC. The more teachers that a site graduates prior to PhysTEC funding, the more graduates it is likely to have during funding, but the lower the gain in teacher graduation numbers ("ceiling effect").

Department size is not a strong influencer of physics teacher graduation rates. No matter how the effect of department size was tested, it had no influence on graduation or change in graduation rates.

Findings regarding net long-term impact and cost-benefit analysis

About half of the Comprehensive sites sustain their gains and 72% show long-term gains in graduation rates; net long-term impacts are mixed for other types of grants. Among those sites which experienced gains in graduation rates, 40% of Comprehensive sites maintained the gains achieved during the funding period after funding ended. However, 72% experienced net long-term gains in graduation rates compared to their pre-funding rates (because even a decrease in graduation rate during-to-post-funding still represented an overall improvement compared to pre-funding). On average, Comprehensive sites increase their graduates by +1.0 graduates from the pre-funding to the sustainability period, and this gain increases to +1.7 graduates when we only include the 72% which experienced long term gains. Results are similar for Comprehensive I and II grants. However, net long-term impacts are closer to 0 graduates for smaller grants and only few of the other types of sites maintained gains made during the funding period. A few Targeted and Recruiting grants do demonstrate net long-term impacts on graduation rates, resulting in potentially cost-effective interventions (gains of +1 graduate among those with longer-term gains), though at least a few of these might be due to floor effects. Across all grants, these results have the potential to address the national teacher shortage.
Smaller grants may be more cost-efficient, and less risky investments

While Comprehensive grants provide more robust results, the increased cost of these grants results in a cost per year of teaching that is about 3 times that of the Recruiting grants ($5400 vs. $1600). Targeted grants had a higher cost, but similar rates of success as Recruiting grants in the current analysis, resulting in a cost-benefit ratio that is less favorable than Recruiting grants (cost per year of teaching is $4700; similar to the the Comprehensive grants), demonstrating the importance of grant cost, long-term gains in graduation rates, and percent of sites with gains, in cost-benefit analysis. The investment risk factor of the larger grants is also increased: based on the percent of grants which do not result in annual graduation increases, for each funding decision about $186,000 is at stake for Comprehensive grants (compared to $128,000 for Targeted and $69,000 for Recruiting grants). Important caveat: These cost/benefit calculations do not include less tangible benefits (e.g. increased site capacity and infrastructure), which are also important. See Appendix I for details on these calculations.
Findings regarding data collection and analysis

The period of analysis strongly affects results. When we conducted these analyses for funded period, funded plus one, or impact periods, we reached different conclusions. The “Funded +1” period appears to offer a reasonable compromise, and I believe it avoids binning students who are part of the direct grant impact period within the sustainability period, or vice versa.

Floor effects occur in about 10% of sites. About 10% of sites experienced decreases in graduation rates, usually of 1 teacher/year. Thus, statistical fluctuations can easily account for changes of 1 graduate/year, and for those sites starting at 0 graduates, about 10% of them might be expected to increase their graduation rate by 1 teacher/year by random chance.

Round to whole-numbers to help address small fluctuations. Results are strongly affected by small-number fluctuations. In this study, we accounted for this by rounding to the nearest whole-number graduate. In the recommendations, we suggest not counting a change of 1 teacher as an impact if it occurs for a minority of sites, due to these small fluctuations. Rolling three-year averages are a way to address these fluctuations in a visual presentation.

The number of Bachelor’s degrees is a good measure of department size. The best variable to use for department size appears to be the number of Bachelor’s degrees granted per year (available in AIP Uber spreadsheet of IPEDS data).

Recommendations

Portfolio of funding tracks

- Continue to offer a range of funding tracks. PhysTEC funding has a strong influence, and these attract a variety of institution types. Concepts of Scaling Science (such as moral justification and acceptable risk) are discussed in the report which may be valuable for future grants.
- Offer more small grants. Small grants are a good investment in terms of the cost per graduate and low risk. One could also consider funding Comprehensive sites at a lower level to reduce risk.

Screening applications for funding

- Carefully screen Comprehensive sites. This analysis suggests that 4X as much effort should go into screening these sites given the financial risk.
- Consider how to use prior graduation rates in screening. There might be an argument to be made to fund sites that are not already graduating large numbers of teachers, to increase gains.
Supporting funded sites

- **Flag sites that are on track for null gains and intervene.** Effectiveness of funding is strongly increased when all sites experience positive gain during, and post funding.
- **Reward graduation numbers, and gains, appropriate to the funding track.** Exceptional performance for Comprehensive sites is 4-5 teachers (or gains of 2-3 teachers/year), whereas for smaller institutions it is graduating 2-3 teachers/year (or gains of 1-2 teachers/year).
- **Use these results to counsel leaders on appropriate expectations.** While it is reasonable for sites to hope to increase and sustain their teacher graduation rates, it is also not extraordinary to experience some losses. To maintain motivation, site leaders should set appropriate expectations.

Tracking and calculating graduation numbers

- **Require more than 3 years of baseline and post-funding data.** This would improve the quality of impact analysis.
- **Use funded + 1 periods to report outcomes, rather than funded periods.** This helps account for lagging indicators.
- **Remove or interpolate data for outliers.** This can help provide more easily interpreted graphs and data which are based on small numbers.
- **Use a whole-number of graduates to report outcomes.** Rounding graduates to the nearest whole number allows clearer interpretation of impacts.
- **Do not count changes of +/- 1 teacher/year as an impact if it occurs for a minority of sites.** This helps account for floor effects and small fluctuations.
- **Use regressions for impact analyses.** The current regression analysis, while costly, allowed accounting for effects such as pre-funding graduation rates, floor effects, and department size.

I suggest that these results be used within the project and the PhysTEC community to better understand the effects of funding and use this information. I also suggest disseminating through PhysTalk, the year-end NSF report, and the website.
FULL REPORT

Methods

What were we trying to learn?

How is the physics teacher graduation rate affected by PhysTEC funding, versus other variables?

1. To what extent does the average physics teacher graduation rate change from pre-funding to during-funding for different funding tracks?
2. How is this change in graduation rates affected by background variables (pre-funding graduation rates, changes in department size), using linear regression modeling? (analysis conducted by Joe Taylor; see “Regression study”).
3. For Comprehensive sites, how are these changes maintained through the sustainability period? (Note that we only had adequate data to address this question for Comprehensive sites).
4. Which sites are “standout sites,” performing better than would be predicted by the model?
5. What are the implications for PhysTEC in terms of data collection and monitoring?

What were the data sources?

Below is a list of the relevant variables and data sources.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Funding track</td>
<td>Which funding track each site was awarded. *Source: PhysTEC data.</td>
</tr>
<tr>
<td>Teacher graduation rate</td>
<td>The average number of annual physics teacher graduates during the relevant period.* *Source: PhysTEC data through 2019-2020 academic year.</td>
</tr>
<tr>
<td>Change in teacher graduation rate</td>
<td>The difference between teacher graduation rates from one period to another. *Source: Calculated number</td>
</tr>
<tr>
<td>Department size</td>
<td>The average number of Bachelor's degrees awarded by the department, averaged across the period of interest (e.g., pre-impact).* *Source: IPEDS data in “Uber” spreadsheet.</td>
</tr>
</tbody>
</table>
Below is an outline of the main funding tracks in this study, based on information from PhysTEC.

<table>
<thead>
<tr>
<th></th>
<th>Comprehensive</th>
<th>Targeted</th>
<th>Recruiting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus</strong></td>
<td>Large institutions</td>
<td>Small institutions</td>
<td>Small institutions</td>
</tr>
<tr>
<td><strong>RFP years</strong></td>
<td>Multiple years 2001-2015 (2019 is not included in this study)</td>
<td>2010, 2011, 2012, 2013</td>
<td>2014 (2020 is not included in this study)</td>
</tr>
<tr>
<td><strong>Award amount</strong></td>
<td>$300,000</td>
<td>$75,000</td>
<td>$30,000</td>
</tr>
<tr>
<td><strong>Consistent elements</strong></td>
<td>3 years funding (though 2001, 2003 and 2004 were 4+)</td>
<td>3 years funding</td>
<td>3 years funding</td>
</tr>
<tr>
<td><strong>Support</strong> (italicized items vary across sites)</td>
<td>Annual site visit Monthly calls PD for TIRs PhysTEC conference Meeting at conference</td>
<td>Annual site visit Monthly calls PD for TIRs PhysTEC conference Meeting at conference</td>
<td>Annual call PhysTEC conference Meeting at conference</td>
</tr>
</tbody>
</table>
Findings: Descriptive statistics

These findings all use the “funded period plus one” period and were conducted by Stephanie Chasteen in December 2020 using updated data.

Graduation rates across all sites

PhysTEC funding is associated with significant gains in teacher graduation during the funded period.

Table 1 and Figure 2 provide information about graduation rates across all 43 sites studied. The average gain in physics teacher graduates was nearly one graduate (+0.9 graduates); this gain is statistically significant (p<0.001). The results of the regression model (which removes effects of pre-funding graduation rates, department size, and changes in department size, but used the “impact period”) upheld these findings, providing reassurance that graduation rate changes are due to funding rather than other local factors.

PhysTEC funding is associated with positive gains for most sites. Most funded sites (N=30; 70%) experience gains of at least 0.5 teachers. Among those 70% with gains, the average gain is +1.5 graduates.

Floor effects may impact about 10% of sites which start at 0 teachers

There is a concern among PhysTEC staff that institutions which are producing 0 teachers are only able to demonstrate growth, whereas those which are already producing 1 teacher may experience growth or decreases. This is termed a “floor effect.” We do find evidence of a floor effect in the histograms of teacher graduation rate changes (not shown); if we consider that it is unlikely for PhysTEC funding to cause a decrease in graduation rates, we could consider that all negative changes in graduation (1-2 teacher decreases for 3 sites overall; 7% of sites) represent the somewhat normal fluctuation of graduation rates, and that this fluctuation is not possible for those starting at 0. So we might consider that out of the 11 sites which began by producing 0 teachers, a small number of them (1; 7% of 11) might just as well have experienced decreases, and so their increases in teacher graduation rates are suspect. There is not an easy way to account for floor effects, unfortunately. One approach is to do as was done in this study, and to use a regression model. In a regression model, a site’s gain is predicted based on where they are expected to perform, given the average pre-funding level. Thus, all sites are put at the same average pre-funding level.
Comparison across funding tracks

In this section, we provide comparative data across funding tracks to understand the differential effects of different amounts of funding and different levels of support. The conclusions in this section draw from Table 1 and Figures 3, below.

### Table 1. Annual graduation changes by funding track (gains, sustainability, and total gains), using “funded + 1” period.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Comp.</th>
<th>Targeted</th>
<th>Recruiting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N with data through sustainability period</strong></td>
<td>40</td>
<td>25</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td><strong>Increase pre-to-during</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ave. increase during funding</td>
<td>+0.9 graduates*</td>
<td>+1.1** graduates</td>
<td>+0.6 graduates</td>
<td>+0.6 graduates</td>
</tr>
<tr>
<td>“Hit rate” of gains***</td>
<td>70% (N=28)</td>
<td>76% (N=19)</td>
<td>63% (N=5)</td>
<td>57% (N=4)</td>
</tr>
<tr>
<td>Ave. increase (sites with gains)</td>
<td>+1.5 graduates**</td>
<td>+1.7 graduates**</td>
<td>+1.1 graduates*</td>
<td>+1.1 graduates*</td>
</tr>
<tr>
<td><strong>Change during-to-post</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ave. change from funded (sustainability)</td>
<td>-0.1 graduates</td>
<td>-0.1 graduates</td>
<td>-0.4 graduates</td>
<td>-0.2 graduates</td>
</tr>
<tr>
<td>“Hit rate” of sustaining gains***</td>
<td>32% (N=21)</td>
<td>40% (N=10)</td>
<td>25% (N=2)</td>
<td>14% (N=1)</td>
</tr>
<tr>
<td><strong>Net change from pre-to-post: These are the ultimate impacts of the grant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ave change pre-to post</td>
<td>+0.7*</td>
<td>+1.0**</td>
<td>+0.2</td>
<td>+0.4</td>
</tr>
<tr>
<td>“Hit rate” of gains**</td>
<td>59% (N=23)</td>
<td>72% (N=18)</td>
<td>38% (N=3)***</td>
<td>29% (N=2)</td>
</tr>
<tr>
<td>Ave. increase (sites with gains)</td>
<td>+1.7 graduates**</td>
<td>+1.8 graduates**</td>
<td>+1.3 graduates</td>
<td>+1.8 graduates</td>
</tr>
</tbody>
</table>

* Items flagged * are statistically significant at the p<0.05 level. ** are statistically significant at the p<0.001 level.

** I define “hit rate” as the % sites with positive gains pre-to-during (more than +0.5 graduates) or sustainability of gains during-to-post (changes greater than -0.5 graduates for those which experienced pre-to-during gains).

*** In the Targeted Sustainability study I interpolate a graduation rate for SUNY Geneseo’s outlier year, resulting in an N=4 for net long-term gains, but this does not affect the average increase for sites with gains.
Figure 2. Longitudinal graduation rates, by funding track. Grey circles indicate years for which data is not available for all sites.

Data is available for 15 (out of 23) sites past PF4; while this is included in that site’s average sustainability numbers, it is not plotted here.

Data is available for 6 (out of 8) sites for PF5, and for only 2 sites beyond PF5.

Data is catalogued for all of these years, however 3 sites report no data for PF1, and 2 report no data for PF2 and PF3. The reasons for this lack of reporting are not clear.
Figure 3. Aggregated graduation rates per period, by funding track
Data plotted is for the N=40 sites with data into the post-impact period

The results for Comprehensive I versus II sites are nearly indistinguishable; see below.
All types of sites experience positive gains during the funded period

(Note that these results differ from the previous report, as new data and modified period of impact changed the conclusion as to whether Targeted sites outperformed Recruiting sites). Using the “funded +1” period we see that all types of sites increase their graduate numbers on average by +1.0 graduate when rounded to the nearest whole graduate (Table 1). The percent of sites experiencing positive gains is 70% across all funding tracks, and is more than 50% for any one funding track. These increase in graduates have potential to contribute meaningfully to solving the teacher shortage.

Comprehensive sites outperform other types of funding tracks on all measures.

Compared to other funding tracks, Comprehensive sites experience greater increases in the number of graduates (+1.1 on average), more Comprehensive sites achieve increases in teacher graduation rates (72%), and more of these increases are greater than +1 graduate (Table 1). Compared to Targeted and Recruiting grants, Comprehensive grants are at least 2 times more effective at boosting graduation rates, on average, when considering gains during funding and total gains pre-to-post funding (Table 1, Figure 3). Looking at line graphs of graduation rates over time (Figure 2), Comprehensive sites show a more obvious change over the funded +1 period. Thus, Comprehensive grants are generally more successful, but these grants also have a higher cost, and a cost-benefit calculation is done later.

This analysis also has implications for awards. Based on the regression and standard deviation, larger institutions (like comprehensive sites) may be rewarded for graduating 4-5 teachers per year, or increasing their graduation numbers by 2-3 teachers/year. Smaller institutions, however, may be rewarded for graduating 2-3 teachers per year, or increasing their graduation numbers by 1 teacher/year.

Increasing the proportion of sites with positive gains would increase impacts

Targeted and Recruiting sites both experience average positive increases of about 0.5 graduates, but these are only sustained for Recruiting sites. As more data is collected over time these conclusions are likely to shift. The difference between Comprehensive and the smaller grants are due to two factors: (1) fewer sites with positive gains pre-post (28% and 38% for Targeted and Recruiting vs. 72% for Comprehensive) and (2) a smaller gain for those sites with positive pre-post gains (+1.1 for both Targeted and Recruiting sites vs. +1.7 for Comprehensives). The smaller increase in graduates might be appropriate for this type of institution getting a smaller grant, but increasing the proportion with positive gains to be commensurate with Comprehensive sites would boost the average graduation rate to be closer to +1.0 graduates for Targeted sites.
Sustainability and net long-term impacts

We investigated the sustainability of graduation rate gains. We were only able to do this for the sites for which there was graduation data in the sustainability period (Year 2 post-funding and beyond). While all funding tracks maintain their during-funding graduation rates, on average (the average change in graduation rates from during-to-post “funding +1” period was 0 graduates when rounded to the nearest whole graduate), this is not a valid measure of sustainability. This is because the gain during-to-post includes maintaining a zero rate of graduation or maintaining a graduation rate during-funding which is actually lower than the pre-funding rate. Thus, I investigate whether the sites which experienced gains in graduation rates maintained those gains.

Almost half of Comprehensive sites with graduation gains sustain those gains; sustainability is lower for other site types.

Among those sites which experienced gains (N=4, 5, and 19 for Recruiting, Targeted, and Comprehensive, respectively), 40% of Comprehensive sites maintained those gains, but only 1-2 of the Recruiting and Targeted sites maintained those gains (14-25%). Thus, sustainability rates are also superior for Comprehensive sites, and helping other legacy sites maintain their graduation gains would increase the long-term impacts of the grants.

Sites with graduation gains maintain long-term gains of 1 or more teachers

Among those sites with sustained gains in teacher graduation rates, overall long-term increases were +1.3 graduate for Targeted grants, and +187 graduates for Comprehensives and Recruiting (Figure 4). However, the Recruiting long-term gains are based on N=2 sites that achieved pre-post gains.

Figure 4. Net pre-to-post gains in graduation rates for those sites with gains, on average.
Across all sites, Comprehensive sites show the most consistent net long-term gains in graduation rates; results from other sites are more mixed.

The ultimate goal of the grants is to sustainably increase the number of teachers. Thus, I investigated the total gain\(^1\), pre-to-post funding, using the “funded +1 period”. Comprehensive grants, on average, increase their graduation rates by +1.0 graduates (p<0.001; Figure 5). Figure 5 shows that the most common result for Comprehensive sites is an increase in teachers from pre-to-post, and the most common increase is 1 teacher, but some sites experience greater increases. This increase is an important step towards addressing the national teacher shortage. For Recruiting and Targeted grants, however, the most common result is 0 change pre-to-post (Figure 3); only a few sites in each track experienced sustainable increases for an average increase close to 0 graduates. These grants have reduced costs, however, so the cost-benefit analysis below examines whether lower success rates are still cost-effective.

Figure 5. Histogram of long-term gains by funding track

Cost-benefit ratio analysis

A large part of the rationale for offering Targeted and Recruiting grants are their low cost. Thus, the impact analysis above should not be taken to suggest that lower-level funding opportunities should not be offered; rather, the risk-to-reward ratio for these grants is lower. Below, in Table 2 I provide a simple cost/benefit analysis to help compare different funding tracks. These analyses are very crude, and do not take into account factors such as development of local expertise and structure, sustainability, and institution size, and selection effects. The calculations are outlined in [Funded sites cost-benefit analysis spreadsheet](#) and in the Appendix.

---

\(^1\) (Note that the number of sites experiencing gains pre-to-post is greater than that with sustained gains because this metric takes into account only the pre-funding graduation rate, ignoring any change during the period of funding.)
Table 2. Cost-benefit analysis of funding tracks, per site, using Funded +1 period.

<table>
<thead>
<tr>
<th></th>
<th>Comprehensive</th>
<th>Targeted</th>
<th>Recruiting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost-benefit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost per grant (direct grant</td>
<td>$663,392</td>
<td>$206,778</td>
<td>$95,681</td>
</tr>
<tr>
<td>cost plus APS staff time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average total new teachers</td>
<td>4.3</td>
<td>2.5</td>
<td>2.4</td>
</tr>
<tr>
<td>during &quot;funding + 1&quot; period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average total new teachers</td>
<td>4.0</td>
<td>0.9</td>
<td>1.7</td>
</tr>
<tr>
<td>during 4 years post-funding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated total new teachers</td>
<td>2.78</td>
<td>0.6</td>
<td>1.2</td>
</tr>
<tr>
<td>during years 6-10 post-funding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Estimated total new graduates</td>
<td>11.1</td>
<td>4</td>
<td>5.3</td>
</tr>
<tr>
<td>Estimated cost per year of</td>
<td>$5,432</td>
<td>$4,702</td>
<td>$1,664</td>
</tr>
<tr>
<td>teaching*, rounded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Risk of unproductive grants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of sites with no gains</td>
<td>0.28</td>
<td>0.62</td>
<td>0.72</td>
</tr>
<tr>
<td>pre-post</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investment risk**</td>
<td>$186,000</td>
<td>$128,000</td>
<td>$69,000</td>
</tr>
</tbody>
</table>

* Estimated 10.9 years in profession is based on PhysTEC data, extrapolated. This “cost per year of teaching” was recommended by Michael Marder and is the grant cost / (total number of new graduates * years in profession).

** Investment risk is the cost of the grant, times the % of sites without gains (Table 2). It is a gross estimate of the amount of money at risk for any particular funding decision.

**Recruiting and Targeted grants are cost-efficient and less risky**

Even though the number of graduates is lower for Targeted sites and Recruiting sites, the lower cost makes them more cost-effective. Realize, however, that long-term data on these grants is sparse, and the number of sites is small; while we can be confident of the direction of relationships (e.g. that they are more cost effective than Comprehensive grants), do not place too much stock in the exact numbers. Smaller grants result in a cost per year of teaching about 1/3 that of the Comprehensive sites, and the investment risk is similarly about 1/3 that of the Comprehensive sites (based on the number of sites without any gains in graduation rates). Targeted sites’ cost is similar to that of the Comprehensive sites because of the greater cost of the grants, and smaller long-term gains, compared to Recruiting sites -- but these results are tentative due to low sample size.
Efficiency can be achieved by reducing the number of sites which experience no gain in graduation rates

Cost-benefit ratio can be increased by increasing graduation rates at sites or by reducing the proportion of sites without gains or reducing grant cost. The Recruiting and Targeted grants’ success was hampered by low “hit rates”; less than 50% of sites achieved positive gains in the long-term (though more than half achieved pre-during funding gains). For those who achieved long-term gains in graduation rates, on average they achieved gains of +1 teacher. Thus, maximal grant efficiency might be provided by offering small grants (e.g. $30,000) but doing what is necessary to ensure that more of the funded sites increase their graduation numbers. Careful pre-screening might reduce failure rates, as well as more vigorous non-financial support (video conferences, mentoring). This is the current plan for Recruiting sites, and future measurement will indicate if this approach bore fruit.

Summary of descriptive analyses

I summarize these results in graphical form below (Figure 6), showing that Comprehensive sites result in greater graduation rates, a greater fraction of sites with gains, but a higher cost per year of teaching, and greater investment risk per site.

Figure 6. Summary of analyses across funding tracks.

<table>
<thead>
<tr>
<th></th>
<th>Average increase</th>
<th>Percent sites with gains pre-post funding</th>
<th>Cost per year of teaching</th>
<th>Investment risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0</td>
<td>0.5</td>
<td>1.0</td>
<td>1.5</td>
</tr>
<tr>
<td>Comprehensive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Targeted</td>
<td>6.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruiting</td>
<td>6.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Discussion: Scaling science**

To help interpret the cost-benefit analysis of these different funding tracks, and their risks, I consulted a discussion of scaling science for social impact. Here I pose some key questions from that discussion:

- **Inclusive coordination.** How can you include the various actors who would make it possible to address a problem at scale? PhysTEC’s various interventions (e.g. Comprehensive sites, Fellows sites) fundamentally provide engagement and directed coordination, and is a strong indicator that PhysTEC can achieve impacts on physics teacher education at scale.

- **Optimal scale.** What is the optimal scale to achieve solutions? Due to trade-offs, an optimal scale is likely to be an intermediate scale. Additionally, small scale interventions can have a variety of impacts (such as accessibility to underserved populations) apart from pure increases in graduation numbers. And as the proportion of physicists involved in physics teacher education increases, the culture of U.S. physics departments changes, making further impacts easier.

- **Moral justification and risk.** How certain must you be that your innovation will achieve positive impacts and avoid negative ones before you scale? What are the acceptable impact risks of scaling? Social innovators should seek to achieve an acceptable level of impact risk based on the urgency of the problem and the cost of failure. The graphic below illustrates this idea.

![Diagram showing the relationship between scale, certainty of impact, and acceptable impact risk.](image)

This brief discussion of scaling science suggests that offering a variety of funding tracks is well-justified. Such funding tracks are likely to address the problem of educating future physics teachers at scale because:
1. Various funding levels are likely to engage multiple actors in a wide variety of U.S. physics departments, increasing inclusive coordination.
2. Intermediate scaled interventions are likely to have a variety of impacts (including cultural change in U.S. physics departments) which help ease scaling.
3. The certainty of impact, based on these analyses, is acceptably high -- the certainty of impact is high for large-scale Comprehensive sites (upper right of graph), whereas the risk is low for Targeted sites, with less certain impact (lower left of graph).

Recommendations

Portfolio of funding tracks

Continue to offer a range of funding tracks
PhysTEC funding has a significant effect on funded sites’ physics teacher graduation rates, and these cannot be explained by “floor” effects (i.e., that a site starting at 0 graduates can only increase and not decrease their graduation numbers). These funding tracks are a good investment in solving the national shortage of physics teachers. You might justify these various funding tracks in future grants using the Scaling Science discussion in this report, and include concepts of scale in future Theories of Change; see Scaling Science article.

Offer more small grants
Because they are more cost effective, and less risky (lower left of Impact/risk graph), it may be prudent to offer more smaller grants in the future, and fewer Comprehensive grants. Comprehensive grants are likely still necessary to attract the attention of larger institutions. One could also consider funding Comprehensive sites at a lower funding level to reduce risk. You might justify these various funding tracks in future grants using the Scaling Science discussion in this report, and include concepts of scale in future Theories of Change; see Scaling Science article.

Screening applications for funding

Carefully screen Comprehensive sites
Because of the high investment risk of Comprehensive sites (2X that of Targeted, and 4X that of the current Recruiting grants). While such sites are already carefully screened, is the time investment of this screening adequate for that financial risk? One might expect 4X as much
effort in screening Comprehensive sites than Recruiting sites. This will help further reduce the risk for Comprehensive sites.

**Consider how to use prior graduation rates in screening**

Given that pre-funding graduation rates were associated with strong during-funding graduation rates, but limited subsequent gains, there might be an argument to be made to fund sites which are *not* already graduating large numbers of teachers.

---

**Supporting funded sites**

**Flag sites that are on track for failure and intervene**

The effectiveness and efficiency of all funding tracks is strongly impacted by the existence of sites which experience no positive gain in physics teacher graduation. Identifying sites which are not on track to generate gains and increasing interventions at those sites is recommended to ensure the full benefit of PhysTEC funding. This could be done in an annual review of site data, with some pre-identified interventions to assist.

**Reward graduation numbers, and gains, appropriate to funding track**

Based on these analyses, Comprehensive sites may be rewarded for graduating 4-5 teachers, or increasing graduation numbers by 2-3 teachers/year. Smaller institutions, like Recruiting grants, may be rewarded for graduating 2-3 teachers/year, or increasing their graduation numbers by 1-2 teachers/year.

**Be careful in messaging about department size (and whether this may impact teacher graduation rates)**

The data does not support the existence of a relationship between department size and graduation rates during funding, and so it may not be fully justifiable to indicate that growing the department will lead to increases in teacher graduates, though the converse is likely to be true; offering a teacher education track may help to grow the department.

**Counsel sites to expect only small decreases in graduation rates post-funding.**

Based on results from both Comprehensive and Targeted sites, sites may expect to reduce physics teacher graduation rates by 0 or 1 teachers/year by around Year 3 post-funding. Setting reasonable expectations (and perhaps incentives for maintaining current graduation rates) would be appropriate. Because sites which maintained graduation rates also experienced increases in department size, boosting the size of the department (i.e. number of physics majors) may help protect gains made during funding.
Tracking and calculating graduation numbers

Consider requiring more than 3 years of baseline and post-funding data

The project will benefit from more years of baseline data (more than 3 years) if possible. This will provide more evidence toward the nature of the graduation rate trends prior to funding, and subsequent insight into the true impact of funding and when that impact begins to manifest. Similarly, the project has been collecting graduation numbers from some sites for quite a few years beyond formal funding periods. This is a good practice toward continued estimation of impact sustainability -- especially given the fact that in the current study, the “sustainability” period did not begin until post-funding Year 3.

Consider accounting for lagging indicators using “funded +1” (rather than funded periods) in data analysis

Longitudinal impact analyses are subject to residual and lagged effects. Specifically, the effects of funding cannot be detected immediately, being lagged by the time necessary for those effects to manifest in higher graduation rates. Similarly, investigations of sustainability are confounded by the residual effects of funding that are still detectable for some time after funding officially ends. The recommendation is thus to err on the side of face validity; if one cannot conceive of funding impacts occurring prior to a certain year, that year should still be considered in the baseline period. Likewise, if the result of funded activities could not be realized until some years after funding, those additional years should be included in the outcome data associated with funding impact. This was the approach used to define the impact period of these analyses, and the funded+1 period, and should be continued until shown to create more bias than it corrects.

Account for small numbers (of years, and graduates) to provide more trustworthy averages:

- **Remove or “smooth” outlier years**

  These and future analyses will necessarily draw upon average values for various periods of time, and often average values are derived from just a few years of data. Whether this is a problem for analyses will depend as much on the variation in data values as it does on the quantity of data. For example, a 4-year average is always better than a 3-year average in terms of representativeness and resistance to outlier bias, but that principle is only true if the standard deviation is non-trivial. That is, a three-year average could be trustworthy if the data are consistent longitudinally. Perhaps the project could adopt an outlier identification scheme, where a value that is more than 2 standard deviations above or below the mean for that variable is considered an outlier and removed from the analysis or is replaced by the mean of the other values for that site during the relevant period.
• **Use nearest whole-number of graduates**
  When averaging across graduation rate changes of only a few graduates, one often ends up with averages of partial graduates (i.e. 1.2 graduates, -0.3 graduates). I recommend using nearest whole-number graduates to avoid over-interpreting these partial graduates. Using this cutoff, a site must have at least +0.5 graduates gain to count as a positive gain, and must have at least -0.5 graduates loss to count as a loss. This also helps to account for floor effects (those with 0 graduates cannot experience loss), since 15% of sites experienced losses of 1-2 graduates that can’t be reasonably explained by the impact of PhysTEC.

• **Only count changes of +/- 1 teacher as a clear impact if it occurs for a majority of sites**
  About 10% of sites experienced decreases in graduation rates, usually of 1 teacher/year. Thus, statistical fluctuations can easily account for changes of 1 graduate/year, and for those sites starting at 0 graduates, about 10% of them might be expected to increase their graduation rate by 1 teacher/year by random chance. If a change of +/- 1 teacher occurs for 50% or more of sites, one can reasonably assume that this is an effect of funding rather than statistical fluctuation.

**Use AIP IPEDS major graduation rate as a proxy for department size**
Controlling for department size change while estimating changes in physics teacher production will continue to be important from the standpoint of face validity of the findings. The IPEDS department size data (number of physics major graduates, averaged over the period of interest) from the AIP “Uber” spreadsheet had the strongest relationship with graduation rates. While those data were less correlated with physics graduation rates once other factors were included in the regression model, it is still useful to include department size as a predictor variable in the future for face validity (i.e., so a reasonable person might believe that changes in graduation rate are due to funding, rather than department growth).

**Next steps for disseminating and using the work**
Since this report has many implications for the project (and was a significant investment) I consider a few possible ways to disseminate these results:

1. Bring to the Project Management Team
2. Discuss in detail with any PhysTEC staff or interns who will be analyzing graduation data
3. Include in year-end NSF report
4. Include in PhysTalk
5. Possibly create an evaluation section of the website
Appendix 1: Cost-benefit analysis

Here I outline how the cost-benefit analyses were conducted.

Risk factor
The investment risk factor is calculated by:
- Investment risk = (grant cost in dollars)*(Fraction of sites with pre/post gains less than 0.5 teachers).

Cost per year of teaching
For this calculation we needed to estimate the following parameters
- Average years in profession (extrapolated from PhysTEC data for years 1-5 post-graduation): 10.9 years
- New teachers graduated during years 6-10 post-funding (extrapolated from data for the 7 sites with data past PF7): 0.55 of the teachers graduated year 1-5 post funding.
- Staff cost for supporting the funded sites (estimated by PhysTEC).

Note that all calculations remove the baseline teacher graduation rate and use only gains.

The calculation was then:
- Total new teachers = (additional new teachers Year 1 of funding through Year 1 post-funding)*4 years + (additional new teachers year 2-5 post-funding) * 4 years + (additional new teachers year 6-10 post funding)*5 years
- Cost per year of teaching = (grant cost in dollars)/(total new teachers)*(estimated years in profession).