Appendix B

This appendix describes our plans for addressing the research question:

- What is the impact of Noyce on teacher recruitment and retention and on student achievement?

The impact analyses will focus on teacher and student outcomes. Quasi-experimental research design approaches will be used to address each of the questions of the impact of Noyce on teacher recruitment and retention and the impact of Noyce on student achievement. First, we describe how we will analyze state teacher certification and employment data to examine teacher outcomes. In a subsequent section, we describe how we will analyze district data on student standardized test scores to examine student outcomes.

Teacher Outcomes

For teacher outcomes, we consider both teacher recruitment and teacher retention. The aim of the Noyce Program is to train more teacher candidates who are highly qualified in a STEM content area to teach in high-need districts. Thus we consider two aspects of teacher recruitment – teacher certification in a STEM field and teacher entry into a school located in a high-need district. We also examine retention of STEM certified teachers in high-need districts. The impact analyses on teacher outcomes therefore seek to address the following sub-questions:

1. The impact of Noyce on **the number of teachers certified**: How does an IHE's receipt of a Noyce grant affect its production of certified or licensed STEM teachers?
2. The impact of Noyce on **teacher recruitment into high-need districts**: How does an IHE's receipt of a Noyce grant affect its production of certified or licensed STEM teachers who take teaching jobs in high-need districts?
3. The impact of Noyce on **teacher retention in high-need districts**: How does an IHE's receipt of a Noyce grant affect the persistence in teaching in high-need districts among the STEM graduates of its teacher certification program?

Question 1: How does an IHE's receipt of a Noyce grant affect its production of certified or licensed STEM teachers?
Our approach to this question seeks to determine whether receipt of a Noyce grant causes IHEs to produce greater numbers of certified STEM teachers than the numbers the IHEs would have produced if they had not received Noyce grants. Our proposed quasi-experimental approach to addressing this question utilizes a difference-of-differences approach. This approach is also known as a "pre-post with comparison group" design, and is computationally similar to a "short-interrupted time series" design. The discussion that follows illustrates why the approach is called "difference-of-differences."

Imagine two similar IHEs from the same state that both have programs to produce certified STEM teachers. Suppose that IHE "A" received its Noyce grant in 2005, and IHE "B" received its Noyce grant in 2009. Imagine that we counted the numbers of certified STEM teachers produced from each IHE for each year from 1999 to 2009. This time period includes before and after Noyce years for IHE "A", but only before Noyce years for IHE "B." For this analysis, we can think of IHE "A" as the treatment IHE, and IHE "B" as the comparison IHE.

In IHE "A" we calculate the difference between the numbers of certified STEM teachers for the years *before* receipt of the Noyce grant (i.e. 1999 – 2005) and the numbers in the years *after* receipt of the grant (2006-2009). Let us denote the after-Noyce minus before-Noyce difference as $Post_{IHE\_A}$ - Pre_{IHE_A} . Next, we do the same calculation for IHE "B", using the same "before" and "after" periods. For this calculation the "before" period is 1999-2005, before both IHEs received their Noyce grants, and the "after period" is 2006-2009, after IHE "A" received its Noyce grant, but still before IHE "B" received its Noyce grant. Let us denote this difference as $Post_{IHE\_B}$ - Pre_{IHE_B}. We then calculate the difference-of-differences as $(Post_{IHE\_A}$ - $Pre_{IHE\_A}) - (Post_{IHE\_B}$ - $Pre_{IHE\_B})$. If the "after" minus "before" difference is greater at IHE A than for IHE B, the difference of differences will be a positive number that we can interpret as evidence that receipt of the Noyce grant caused an increase the IHE's production of certified STEM teachers.

The rationale for interpreting the difference-of-differences as evidence of causation is that, in IHE "A," the "after" minus "before" difference is assumed to be due to both 1) year-to-year changes in the hiring of STEM teachers in high-need districts within the state, and 2) the effect of receipt of the Noyce grant on the IHE's production of teachers, whereas the difference in IHE "B" is due only to year-to-year changes in hiring of STEM teachers in high-need districts within the state. Subtracting the two differences takes out the effects of year-to-year changes in hiring within the state, leaving only the effect of the Noyce grant on production of the teachers. The identifying assumption here is that year to year changes in IHE A and IHE B would have been the same if IHE A had not received its grant. This assumption would not hold if there was a factor (other than the receipt of the grant) that only affected either of the IHEs in the pre or post period.

The example of IHE A and B is illustrative of only two IHEs, but our analyses will include many IHEs from several states, with varying degrees of overlap in the years of grant receipt. The strength of evidence from this design is directly dependent on the assumptions stated above, and the extent to which other systematic influences on the production of teachers have not been adequately accounted for in the analysis models represents a threat to our ability make causal inferences from the data. We further discuss limitations to the approach in a subsequent section.

Analytic Model for Impact Analysis
The analysis requires data on matched sets of IHEs where the matching criteria are such that:

- All IHEs in the set must be from the same state to control for state-level year-to-year variation in hiring of teachers.
- All IHEs must have had programs designed to produce certified teachers for the entire study period, including pre-Noyce years.
- There must be variation within the matched set in the year of the award of the Noyce grant.
- For a year to be included in the analyses, there should be at least one IHE that did not have a grant that year, i.e., a year in which all IHEs had a grant should not be included in the analysis since the effect of a secular change that happened in such a year would confound the effect of Noyce.

The analysis model will include data from matched sets from different states, and the model will include terms to identify and make comparisons within the matched sets. In other words, the impact of Noyce will be estimated from <u>within</u> each matched set, and then aggregated over all of the matched sets.

The dependent variable is a time varying measure of an IHE's production of STEM certified teachers. The analysis model has an indicator variable for whether the data come from a pre-Noyce year or a post-Noyce year, and dummy variables for years and IHEs. For simplicity, the model is shown as if all IHEs in the analysis came from a single matched set (e.g. all from the state of Georgia). In a combined analysis using data from multiple matched sets across states, there will be separate sets of year dummies from each matched set (thus, the year effects will be assumed to be common only among states in the matched set). The analysis model is of the form:

<u>Model 4.3.</u>

$$Y_{ij} = \beta_0 + \beta_1(T_{ij}) + \beta_2(Inst_1) + ... + \beta_{2+M-1}(Inst_{M-1}) + \beta_{2+M}(Yr_{2000}) + ... + \beta_{2+M+10}(Yr_{2009}) + \varepsilon_{ij}$$

Where:

Y_{ij}	=	the number of individuals who completed teacher preparation at IHE *j*, and received a STEM certification in year *i*.
T_{ij}	=	1 if a Noyce grant had been received at least one year prior to year *i* in IHE *j*;
	=	0 otherwise (i.e., if a pre-Noyce year at IHE *j*)
Yr_{2000}	=	1 if year is 2000,
	=	0 otherwise (year 1999 is the omitted year)
…		… additional dummy variables for years 2002 - 2008
Yr_{2009}	=	1 if year is 2009,
	=	0 otherwise (the dummy for year 1999 is the omitted from the model)
$Inst_1$	=	1 if 1st of M institutions (IHEs) in the analysis
	=	0 else
…		… additional dummies for additional institutions
$Inst_{M-1}$	=	1 if 2nd to last of *M* institutions (IHEs) in the analysis
	=	0 else (the dummy for the *M*th institution is omitted from the model)

β_1	=	The covariate-adjusted average difference between IHEs' production of STEM certified teachers who teach in high-need districts before and after receipt of a Noyce grant.
ε_{ij}	=	residual error, assumed distributed normal with mean zero and variance $\sigma^2_{Y|T,Yr,Inst}$ (i.e., the variance is the variance of Y_{ij}, after the set of independent variables including the treatment dummy, the year dummies, and the institution dummies, have explained some of the variance. $\sigma^2_{Y|T}$ is the unconditional residual variance (i.e., if the only independent variable on the right hand side of the model was the treatment indicator), and $R^2_{Y|Yr,Inst}$ is the proportion of variance explained by the set of covariates including the year dummies and the institution dummies. $\sigma^2_{Y|T,Yr,Inst} = \sigma^2_{Y|T}(1- R^2_{Y|Yr,Inst})$. Furthermore, we decompose $R^2_{Y|Yr,Inst}$ into $R^2_{Y|Inst}$ (the proportion of $\sigma^2_{Y|T}$ that is explained by the institution dummies), and $R^2_{Y|Yr}$ (the proportion of $\sigma^2_{Y|T,Inst}$

explained by the year dummies, i.e. the proportion of variance of Y that is explained by the Year dummies, after the treatment and institution dummies have already explained some of the variance) and such that

$$R^2_{Y|Yr,Inst} = R^2_{Y|Yr} + R^2_{Y|Inst}$$

and

$$\sigma^2_{Y|T,Yr,Inst} = \sigma^2_{Y|T}(1-(R^2_{Y|Yr} + R^2_{Y|Inst})).$$

A model of the form above has been tested using simulated data and has performed well for various sets of assumptions, the model produced impact estimates and standard errors that appeared to converge on the true parameter values used for the simulations.

We describe the method for estimating the minimum detectable effects of the Noyce Program on teacher certification in Appendix B.

Question 2: How does an IHE's receipt of a Noyce grant affect its production of certified or licensed STEM teachers who take teaching jobs in high-need districts?
The analytical approach to addressing Question 2 is essentially the same as that described for Question 1. The impact model is of the same form as that previously described. The only thing that changes is that for Question 2, the dependent variable in the impact model becomes:

Y_{ij} = the number of individuals who completed teacher preparation at IHE *j*, received a STEM certification in year *i*, and took a job teaching in a high-need district within two years of year *i*

Note that if an IHE received their grant in 2006, and the grant was expected to have an impact on production of teachers starting in 2007, then we would assess outcomes for the following post-Noyce 2007 and 2008, but not 2009 or later. This is because, if an individual received their STEM certification in 2007 then we would assess whether they had taken in job teaching in a high-need districts using extant employment data from 2007, 2008, 2009. If they received their certification in 2008 then we would look at the extant employment data from 2008, 2009, 2010. If they received their STEM certification 2009 we would need to look into the future (2011) to whether or not the individuals that got a job within two years of receiving their STEM certification.

Question 3: How does an IHE's receipt of a Noyce grant affect the persistence in teaching in high-need districts among the STEM graduates of its teacher certification program?
In this section we discuss the approach to addressing Question 3, and describe an approach for answering a different (but related) question about persistence of Noyce grantees. We note that our ability to answer questions about the persistence of Noyce teachers to stay in teaching in high-need districts is hampered by the study's timeline. That is, because of the timing of the Noyce grants and the timing of the study, there will be few potential years of follow-up after Noyce teachers have taken jobs teaching in high-need districts.

Our approach to Question 3 requires that we examine "persistence." For the purpose of explaining the approach, we define a "persistent" teacher to be one who remains teaching in a high-need district for three or more years. We could easily re-define persistence as being two or more years, or four or more years. The approach uses the same data sources as described for Question 3. For each individual in the data set, an employment history would be constructed. For each person, we would

identify the year of first employment in a high-need district, and then create an indicator for employment in a high-need district for the next two years (i.e., worked in a high need district three or more years). Then for each IHE and each year, we will calculate the proportion whose first employment as a teacher was in a high-need district, and persisted for three or more years. That proportion will be the dependent variable (Y_{ij}) in the impact model.

Y_{ij} = the proportion of individuals who completed teacher preparation at IHE *j*, received a STEM certification, and whose first job was teaching in a high-need district year *i*, who persist in teaching a high-need district for three or more years

Similar to what was described for Question 2, here we will have to limit the data to years when there is a sufficient number of years of follow-up to determine whether someone was employed in a high-need district for three or more years.

Otherwise, the analysis approach is the same as described above for Question 1.

An alternative question is, "how does the persistence of teaching in high-need schools for Noyce grant recipient teachers compare to that of other teachers?" We could use the same data source as that described for Question 2 if we were able to identify which teachers in the database were Noyce grant recipients. The data requirements for Questions 2 and 3 require that we are able to link teachers to their IHEs, but does not require that we know which individual teachers within IHEs were Noyce grantees. We expect that we will be able to make these links for data from some states (where they provide teacher names, which can be linked to the names of teachers in the Noyce monitoring data base), but not all states.

If we obtain state databases where we can identify which teachers are Noyce teachers, we will construct an employment history for each teacher who began teaching in a high-need district within the time frame covered in the database (e.g. 1999-2009). The persistence in teaching will be measured as years between first date of hire in a high-need district, and departure (no longer working in any high need district). These data are right-censored meaning that, for many teachers, the study period will end before the teacher has stopped teaching in high-need schools. We will analyze these data in a survival analysis framework, as described in chapters 10-12 of Singer and Willet (2003). These models will compare survival rates of Noyce to non-Noyce teachers.

Limitations
Study limitations include the following:

1) While we will try to ensure that we use only good quality data, any errors or omissions in the state data sources used to measure the number of STEM certified teachers produced by IHEs in the pre-Noyce and post-Noyce periods will create measurement error which may reduce our ability to detect an impact.

2) Causal inferences rest on an assumption that there are common "year effects" that that have similar influences on the production of STEM certified teachers of all IHEs with a matched set. If local economic, political, or regulatory conditions affect production in some but not other IHEs within the matched set, then the impact model may produce biased estimates.

3) Causal inferences rest on an assumption that, after controlling for year effects, any pre-existing within IHE time trends have been correctly specified in the analysis model. The impact model assumes that within an IHE, and absent the receipt of a Noyce grant, after controlling for year effects, the production of STEM certified teachers would have remained the same in the post-Noyce period as had been observed in the pre-Noyce period. If, within an IHE, the production was on an upward or downward trend that would have existed even in the absence of receipt of the Noyce grant, than this assumption would be violated and the impact estimates may be biased[1].

4) Causal inferences rest on an assumption that we have correctly coded model term for "post-Noyce years" to indicate the years when the Noyce grant should have had an impact on production. If, for example, the Noyce grant was not intended to boost production until two years after receipt of the grant, then the "post-Noyce year" indicator variable in the analysis model should not be codes as a "1" until two years after receipt of the Noyce grant.

Student Outcomes

The logic model for the Noyce program hypothesizes that the impact of Noyce on the production and retention of STEM-qualified teachers teaching in high-need districts will lead to improved K-12 student achievement in math and science for students in high-need schools and districts. This hypothesis could motivate a very broad research question such as, "*what has been the impact of the Noyce Program on student math and science achievement in high-need schools?*" Since Noyce has touched relatively few students compared to the size of the national population of K-12 STEM students in high-need schools, the true impact of Noyce on the national population of students in high-need schools is likely to be very small, and not directly estimable. One or more narrowly defined research questions are clearly required. In the design phase of the project Abt, in collaboration with NSF, considered three designs to address three more narrowly defined research questions. After discussion of the design options, NSF has indicated a preference to focus the design on the third research question discussed below. In the text that follows, we briefly review the three sub-research questions we considered to examine student outcomes, and in the remainder of the appendix we discuss in detail our proposed study design to address the third question.

Three Potential Sub-Research Questions Related to Student Outcomes
An example of a more specifically defined research question is:
- *What is the impact of having one or more Noyce teachers in a high-need school on school-level average math (or science) achievement scores?*

We note that, if our study found that the answer to the question above was a positive impact, this finding, by itself, does not represent causal evidence that the Noyce Program caused the positive impact. The conclusion that the Noyce Program caused the impact would be possible only if we had established that the "Noyce teachers" in the high-need schools would not have been in high-need schools absent the Noyce Program. That is, the same teachers who are "Noyce teachers" because they received Noyce grant support may well have gone to teach in high-need schools even if there were no Noyce Program.

---

[1] We will test this assumption by fitting a model that includes IHE-specific time trends. We will treat this model as a specification test -- if the IHE-specific time trends are jointly significant, then this model will be used to obtain the impact estimates.

The teacher impact analyses described in the previous section are designed to answer the question of whether the Noyce Program causes IHEs to produce greater numbers of STEM-qualified teachers who teach in high-need schools or districts. If that analysis results in a significant positive impact estimate, then a link could be made between a relationship of "Noyce teachers" to student achievement and the impact of the Noyce Program on student achievement.

In order for Noyce to have an impact on student achievement, Noyce must have an impact on the production of STEM-qualified teachers. We therefore caution ourselves that if the results of the first set of analyses do not provide convincing evidence that Noyce has impacted the production of STEM-qualified teachers in terms of numbers or quality, then there would be no reason to believe that Noyce could have had an impact on student achievement, and results from analyses showing relationships between Noyce teachers and student achievement would be unconvincing.

We envisioned that an approach to addressing the research question above would involve a difference-in-differences design with an adjustment for the proportion of students exposed to Noyce teachers. For more details on the approach, see Appendix D. Ultimately, NSF expressed greater interest in learning the impact observed on students of having a Noyce versus a non-Noyce teacher in a math or science class.

A similar, yet more broadly defined research question is:
- *What is the impact of having one or more <u>STEM-qualified teachers</u> in a high-need school on school-level average math (or science) achievement scores?*

This question also attempts to learn about causal linkages. Among the impact analyses described in the previous section, the evidence will be strongest for the analysis that determines whether Noyce has an impact on the numbers of STEM-qualified teachers produced by IHEs that then teach in high need schools. Thus, a relevant chain of causal linkages would be of the form, "if Noyce causes increases in the numbers of STEM-qualified teachers teaching in high need schools, does the presence of STEM-qualified teachers in high-need schools then have an impact on student math and science achievement?"

There is a growing literature devoted to it (or to its more broadly defined cousin "*what is the impact of STEM-qualified teachers on student math and science achievement?*"). If this question could be addressed using evidence from the research literature, project resources could be allocated to other aspects of the project or to other NSF priorities.

This question could also be addressed using extant data. Specifically, the same state-level data sources that will be used to address the question of the Noyce Program's impact on IHEs production of STEM-qualified teachers who teach in high-need schools have indicators of whether teachers are certified in STEM fields, although they do not have indicators for whether teachers were recipients of Noyce grants.

The third, and most narrowly defined research questions is:
- *Among students in high-need schools, what is the impact of being taught by a teacher who has received a Noyce grant on students' math (or science) achievement scores?*

This question will be the focus of the remainder of this appendix. Our proposed approach to addressing this question involves a contrast of spring achievement score results between students who

began a school year assigned to a math or science class with a Noyce grantee, to students in the same school taking the same class content but assigned to a non-Noyce teacher. The research question may or may not be further narrowed to ensure that the comparison teachers have a comparable number of years of experience in teaching.

This approach can be thought of as testing the effect of a classroom intervention, specifically having the class taught by a Noyce teacher versus being taught by a non-Noyce teacher. This approach is usually implemented as a random assignment study with careful attention paid to eligibility of students for inclusion in the study, random assignment of students to teachers, and to sample attrition. For example, the U.S. Department of Education's Striving Readers grant program currently has 14 studies underway using this kind of design. This design was also used by the the Teach for America (TFA) evaluation (Decker, Mayer, and Glazerman 2004) and the Alternative Teacher Certification Evaluation (Constantine et al., 2009). When implemented as a randomized controlled trial, inferences about the effect of the classroom intervention are straightforward. When random assignment of students to teachers is not feasible due to budgetary, timing, or other reasons, effects of a classroom intervention can be estimated using quasi-experimental designs. For the question of the impact of Noyce teachers on student achievement, our proposed design would involve identifying schools with at least one Noyce teacher, identifying one or more comparison classes within each school that are taught by non-Noyce teachers, and making within-school comparisons of spring student achievement scores of the students with the Noyce teachers to those taught by non-Noyce teachers. The differences between the scores of the students of Noyce and non-Noyce teachers would be aggregated over schools to produce an overall impact estimate. The analytic models would control for student-level pre-test scores (scores from the prior year), and any other student-level demographic data that are available (e.g., free-reduced price lunch eligibility, limited English proficiency status, special education status).

An analytical challenge presented by this design is that the relatively small population of Noyce teachers means that the impact analysis may require the use of state test results from different states and different grade levels. We first discuss the approach as if the entire analysis could be conducted using state mathematic achievement data from single state and a single grade level (e.g., 7th grade math assessment results in the state of California). We subsequently consider how we could standardize scores to accommodate data from multiple states and grade levels.

Analytic Model for Impact Analysis

The model is specified as a three-level hierarchical linear model where students (level-1) are nested in teachers (level-2) and teachers are nested in schools (level-3)[2].

The level-1 model, or student-level model is:

$$Y_{ijk} = \pi_{ojk} + \sum_{m=1}^{M} \pi_{mjk} (X_{mijk} - \overline{X}_{m.jk}) + e_{ijk}$$

---

[2]   If the data set is such that the analysis includes data from multiple classes taught by the same teacher, we will fit an alternative model with four levels, with students (level-1) nested in classes (level-2), classes nested in teachers (level-3), and teachers nested in schools (level-4). But for the purpose of explaining the approach and building up assumptions for power calculation, we will use the three-level model specification. This simplification will have little if any effect on the power calculations, but is much more straightforward to explain and understand.

where

Y_{ijk}	=	is a spring math achievement test score from i^{th} student (i in 1,2,...,n) in the jth teacher, (j in 1,2,..., J teachers per school), nested in the kth school (k in 1,2,...,K schools).
$(X_{mijk} - \overline{X}_{m.jk})$	=	is the mth of M student-level baseline covariates (e.g. prior year test score, LEP status, special education status, free/reduced price lunch status, sex, race), centered at the teacher-level mean.[3]
e_{ijk}		is the student level residual, assumed distributed normal with mean zero and variance $\sigma^2_{Y|(X-\overline{X})}$.
$\sigma^2_{Y|(X-\overline{X})}$	=	is the variance of the level-1 residuals, after the set of M covariates $(X_{mijk} - \overline{X}_{m.jk})$ has explained some of the level-1 variance. $\sigma^2_{Y|(X-\overline{X})} = \sigma^2_Y (1 - R^2_{Y|(X-\overline{X})})$, and σ^2_Y is the unconditional level-1 variance (i.e., if there were no level-1 covariates), and $R^2_{Y|(X-\overline{X})}$ is the proportion of level-1 variance explained by the level-1 covariates.

The level-2 model, or teacher-level model is:

$$\pi_{ojk} = \beta_{00k} + \beta_{01k}(T_{jk}) + \sum_{m=1}^{M} \beta_{0mk}(\overline{X}_{m.jk}) + \sum_{p=m+1}^{P} \beta_{0pk}(Z_{pjk}) + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k}$$

$$\pi_{2jk} = \beta_{20k}$$

…

$$\pi_{mjk} = \beta_{m0k}$$

where

π_{0jk}	=	is the teacher level mean spring math achievement test score for the jth teacher, (j in 1,2,..., J teachers per school), nested in the kth school (k in 1,2,...,K schools).
T_{jk}	=	1 if teacher has a Noyce teacher and =0 if teacher has a non-Noyce teacher.
$\overline{X}_{m.jk}$	=	is the mth of M teacher-level means of student-level baseline covariates (e.g. prior year test score, LEP status, special education status, free/reduced price lunch status, sex, race).
Z_{pjk}	=	is the pth of P teacher characteristics used as teacher-level covariates (e.g. years of experience)
r_{0jk}		is the teacher-level level residual, assumed distributed normal with mean zero and variance $\tau_{\pi|T,\overline{X},Z}$. Note that if many schools only have one treatment and one comparison teacher than we will have to drop this term from the model because there will not be enough degrees of freedom to estimate this term.
$\tau_{\pi|T,\overline{X},Z}$	=	is the variance of the level-2 residuals, after the treatment dummy and the set of teacher-level covariates ($\overline{X}_{m.jk}$, and Z_{pjk}) has explained some of the level-2 variance. $\tau_{\pi|T,\overline{X},Z} = \tau_{\pi|T}(1 - R^2_{\pi|\overline{X},Z})$, and $\tau_{\pi|T}$ is the unconditional level-2 variance (i.e. if there was a treatment indicator in the level-2 model, but otherwise there were no level-2

---

[3] Centering at the teacher-level mean is convenient for power calculations because the teacher-mean centered covariates will explain level-1 residual variation only, and will not explain level-2 variance.

covariates), and $R^2_{\pi|\bar{X},Z}$ is the proportion of level-2 variance explained by the level-2 covariates (the Xs and Zs).

The level-3 model, or school-level model is:

$$\beta_{00k} = \gamma_{000} + \mu_{0k}$$
$$\beta_{01k} = \gamma_{010} + \mu_{1k}$$
$$\beta_{10k} = \gamma_{100}$$
$$\beta_{20k} = \gamma_{200}$$
$$\dots$$
$$\beta_{M0k} = \gamma_{M00}$$
$$\beta_{p0k} = \gamma_{p00}$$
$$\dots$$
$$\beta_{P0k} = \gamma_{P00}$$

where

β_{00k}	=	the school-level mean spring math achievement test score for the kth school (k in 1,2,...,K schools).
γ_{000}	=	the grand mean spring math achievement test score
μ_{0k}	=	the random intercept for each school mean, which is assumed distributed normal with mean 0 and variance $\tau_{\beta_{00}|\bar{X},Z}$
β_{01k}	=	the treatment effect at the kth school (k in 1,2,...,K schools).
γ_{010}	=	the overall, grand mean treatment effect
μ_{1k}	=	the random effect associated with each school's treatment effect (i.e. the deviation of school k's treatment effect from the grand average treatment effect, which is assumed distributed normal with mean 0 and variance $\tau_{\beta_{01}}$

Note: Since the schools in the sample will be selected as a convenience sample (i.e. schools are not selected at random from a defined population), we will set the term $\tau_{\beta_{01}}$ to zero. This means that we will estimate a single, common treatment effect across schools, and we will therefore limit the inferences from the analysis to the schools included in the study.

We describe the method for estimating the minimum detectable effects of Noyce teachers on student achievement in Appendix C.

Limitations

The major threat to the internal validity of quasi-experimental approaches to this kind of question is the selection process by which students are assigned to teachers. If that selection process is far from random, for example if students are grouped into classes according to behavioral characteristics, motivation, interest in science and/or math, or any other characteristic that is unmeasured and not controlled for in the analysis model, then the quasi-experimental impact estimates may be subject to bias.

Furthermore, if Noyce teachers have better qualifications or credentials than non-Noyce teachers, school administrators may decide to place the more challenging students in the classes with the Noyce teachers, or may decide to place the students who are most likely to make gains in those classes. These selection issues represent a threat to internal validity.

The proposed study is a relatively small pilot study. The schools and districts will be selected as a convenience sample, and not as a random sample from the full population of schools that have Noyce teachers. Thus, there will be no claims made that the result from this study are representative of the full population of schools with Noyce teachers.

The pilot study will not be powered to detect small effects. The study will demonstrate the feasibility of using the proposed design in a larger scaled study which could be powered to detect small, but educationally meaningful effects.