Appendix B

Impact Analyses Using Extant Data

The research questions that will guide the impact analyses are motivated by the intended impacts, but constrained by the study design and the availability of suitable data to address the questions. The impact analyses will seek to address the following questions:

1) What is the impact of Noyce on ***recruitment of STEM teachers***: How does an IHE's receipt of a Noyce grant affect its production of certified or licensed STEM teachers that take teaching jobs in high-need districts?
2) What is the impact of Noyce on ***teacher retention***: How does an IHE's receipt of a Noyce grant affect the average number year teaching in high-need districts among the STEM graduates of its teacher certification program?
3) The impact of Noyce on ***STEM content knowledge of STEM teachers***: How does an IHE's receipt of a Noyce grant affect the average scores on assessments of STEM content knowledge among the STEM graduates of its teacher certification program?

Quasi-experimental research design approaches will be used to address the questions above. Due to the availability and sources of relevant data, and the timing of measurement, we expect that the strength of evidence will be strongest for the first question, weaker for the second question, and weakest for the third question. Below, we describe our approach to addressing each of these questions.

How does an IHE's receipt of a Noyce grant affect its production of certified or licensed STEM teachers that take teaching jobs in high-need districts?

Our approach to this question seeks to determine whether there is evidence that receipt of a Noyce grant causes IHEs to produce greater numbers of certified STEM teachers who take jobs in high-need districts, 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 from each IHE that take teaching jobs in high-need districts 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 teacher that take jobs in high need districts 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 and will be interpreted as evidence that receipt of the Noyce grant caused an increase the IHE's production of certified STEM teachers that take jobs teaching in high-need districts.

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, year-to-year changes hiring of STEM teachers in high-need districts within the state, and the effect of receipt of the Noyce grant on the IHE's production of the teachers. Whereas the difference in IHE "B" is due only to year-to-year changes 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 the effect of the Noyce grant on production of the teachers. The actual analysis will involve 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 be adequately accounted for in the analysis models represents a threat to our ability make causal inferences from the data. We discuss threats to validity 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 – this is 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.
- All must have received a Noyce grant at some point – this is so that comparisons are made among IHEs similar in that they are the type of IHEs that would go after and win a Noyce grant.
- There must be variation within the matched set in the year of the award of the Noyce grant.

The analysis model will include data from matched sets from different states, but 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 within each matched set, and then aggregated over all of the matched sets.

The dependent variable is a time varying measure of an IHEs production of STEM certified teachers that take jobs teaching in high-need districts within two years of receiving a STEM certification. 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:

Model 1.1.

$$Y_{ij} = \beta_0 + \beta_1(T_{ij}) + \beta_2(Yr_{2000}) + ... + \beta_{12}(Yr_{2009}) + \beta_{13}(Inst_1) + ... + \beta_{13+M-1}(Inst_{M-1}) + \varepsilon_{ij}$$

Where:

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*
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 that teach in high-need districts before and after receipt of a Noyce grant.

A model of the form above has been tested using simulated data and has performed well in the sense that 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.

Data Sources for Difference-of-differences Impact Analysis

We know that some states maintain databases where IDs are assigned to individuals at the time that they receive a teaching certification within the state, and records are added to the database, linked by ID, for each year that the individual is employed in any public school within the state. We have begun investigating whether states have, and will give us access to the kind of data that are needed for these analyses. We have, for example, learned that Georgia has the necessary data and is willing to give us access to the data. We have preliminary indications that New York, Massachusetts, and Missouri will be able to provide required data. Investigations for these and other states are ongoing. The critical data elements are described below:

- Whenever a person receives a teaching certification in the state, a record is entered into the database.

- The following certification data elements can be linked, by person, within the data base:
    - The IHE that recommended (or graduated) the person – this will allow us to link certified individuals to IHEs that have received Noyce grants
    - The date that the certification was received – this will allow us to determine if this person's certification is a product of the IHE before or after receipt of the Noyce grant
    - Certification field – this will allow us to determine if the person was certified in a STEM field.
    - Changes and additions to certifications are maintained in the database
- The following employment data elements can be linked, by person, within the database:
    - For each year, whether the individual was employed as a teacher in a public school in the state
    - The name and ID of the school district where teacher is employed – this will allow us to determine whether the individual was teaching in a high-need district.
    - The name and ID of the school where teacher is employed – this will give us the option of determining whether the individual was teaching in a high-need school.
- The data go back to back to at least three years before the year of the first receipt of an Noyce grant by an IHE in that state. (In Georgia, the first Noyce grant was awarded to an IHE in 2005, and the data in the state database go back to 1998).

Using Georgia as an example, for each of the IHEs in Georgia that has ever received a Noyce grant, and each year from 1998 through the present, the data elements described above will allow us to calculate the number of STEM certified teachers produced by the IHE that have taken jobs teaching in high-need districts.