Analysis Plan

Project Name: Using Post 9/11 GI Bill Benefit Balance Notifications to Proactively Increase Student Persistence and Degree Completion

Project Code: 1810

Date Finalized: November 14, 2018

Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

Outcome Variables to Be Analyzed:

We expect to analyze enrollment intensity using rate of pursuit. This variable, generated by our agency partner, ranges from 0 to 1 and indicates rate of enrollment relative to full-time. It is calculated by dividing the number of course credits for which a student is enrolled by the number defined as full-time for the school's academic calendar. For a semester-based academic calendar, full-time is defined as 12 credits — so a student taking 12 credits on a two-semester calendar would have a rate of pursuit equal to 1, and a student taking 6 credits on the same calendar would have a rate of pursuit equal to .5. Rate of pursuit is capped at 1 for students taking more than the number of credits defined as full-time.

We will also examine the impact of the treatment on a dichotomous indicator of full-time enrollment (transformation described below).

Transformations of Variables:

We will generate a binary outcome indicating full-time enrollment by recoding any rate of pursuit (described above) less than 1 as 0, and full-time enrollment as 1.

All other data provided by our agency partner will be used in the raw form provided.

Imported Variables:

We will match our treatment and block indicators to our dataset using a unique study ID created during the initial treatment randomization process.

To complete some of our exploratory analyses, we will match data from the partner with publicly available postsecondary institution-level data from the Integrated Postsecondary Education Data System (IPEDS). This will require us to use a crosswalk that links unique identifications for postsecondary institutions between the two data systems. The crosswalk is provided by the partner agency and includes a direct match between agency data and publicly available data.

Transformations of Data Structure:

The data for analysis will be delivered in a cross-sectional structure with each individual beneficiary represented by a single record. We do not anticipate having to make any data structure changes to complete the analysis.

Data Exclusion:

This analysis will include individual beneficiaries assigned to the control and treatment conditions at the beginning of the study.[1] New beneficiaries (those accessing their benefits for the first time) for the Fall 2018 term will be excluded from the analysis.

Within our specifications that include institution-level covariates, we will exclude any institution that does not not report through the Integrated Postsecondary Education Data System (IPEDS), operated by the National Center for Education Statistics (NCES). This will include any international postsecondary institutions as well as any domestic institutions that chose not to report their data. Based on preliminary data matching, we expect that this will exclude less than two percent of individuals.

Treatment of Missing Data:

Since benefit reimbursement is tied to submission of data, we expect no (or very limited) missing data for our primary outcomes or covariates.

We will not impute institution-level covariates for those not reporting through the IPEDS data system.

Statistical Models & Hypothesis Tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to analysis.

Statistical Models:

Randomization Test

Before continuing with analysis, we will check the initial randomization by conducting d^2 omnibus balance tests using observable characteristics — in particular, beneficiary type (i.e. veteran, dependent, etc.), institution type (i.e. public two-year, public four-year, etc.), prior enrollment intensity, location/urbanicity, and benefit level remaining.

[1] Originally this study was designed to have two treatment groups. The treatment group referenced in this analysis plan is the one initially designated as the “general email” group.

Planned Analysis: OES Abstract

The planned analysis for the OES abstract will examine the intent-to-treat (ITT) effect between the control group (no email) and the focal treatment group (email), using ordinary least squares (OLS) regression:

$$RP_i = \propto + \beta_2 GE_i + \varepsilon_i \tag{1}$$

where RP_i is our continuous measure of enrollment intensity (rate of pursuit). GE_i is our binary treatment indicator that signifies whether the beneficiary was sent the treatment email or not, and ε_i is an idiosyncratic error term.

Using a linear probability model (LPM), we will also estimate an ITT effect on the binary indicator for full-time enrollment (FT_i) with identical specifications as in Eq. 1. We will also run a logistic regression model to test the robustness of the LPM results. If the LPM and logistic regression yield substantively different results, the LPM results will take precedence.

Planned Analysis: Publication and Other Outlets

In addition to the planned analyses for the OES abstract, we are planning on conducting a series of additional analyses to support possible peer-reviewed publications and distribution through other external outlets. First, we will extend our ITT specification (Eq. 1) to include a set of baseline and enrollment covariates:

$$RP_i = \propto + \beta_2 GE_i + X'_i + \phi'_s + \delta'_s + \lambda'_b + \varepsilon_i \tag{2}$$

where we add the following to Eq. 1: X', which is a vector of individual covariates that are known to impact school enrollment; ϕ'_s, which is a vector of time-varying institution-level covariates that may impact enrollment intensity for our sample population; δ'_s, which is a vector of institution-level fixed effects to account for time-invariant and unobserved institution influences; and λ'_b, which is a vector of block-group fixed-effects that were used during the randomization process. Both our individual-level factor and randomization block indicators are included to increase the precision of our estimates.

Second, we will capitalize on data provided by GovDelivery (the platform through which email will be sent) on email open rates and click-through indicators to estimate a treatment-on-the-treated (ToT) effect. Using two-stage least squares (2SLS), we will instrument for an indicator variable for accessing the treatment (i.e. read the sent email) using an indicator variable for assignment to treatment (i.e. sent an email).

Follow-Up Analyses:

Given the timing of the intervention (towards the end of the summer 2018 term), we would expect that any identified changes in enrollment intensity would be primarily concentrated in institutions with open or rolling course enrollment — predominantly non-selection and open access institutions. Accordingly, we will explore heterogeneous treatment effects by institutional type. We will analyze the effects of the treatment by the following institution types:

- Public four-year institutions
- Private four-year institutions
- Public two-year institutions
- For-profit four-year institutions
- Highly selective (based on admissions criteria) institutions
- Open access (no admissions criteria) institutions

We also hypothesize that beneficiary type (i.e. veteran, spouse, or child dependant) and remaining benefit level would further influence the likelihood of responding to the treatment. To this end, we will also explore heterogeneous effects by beneficiary type and levels of benefits remaining.

Finally, we will also examine the effect of the intervention by prior year's enrollment intensity category. Based on data provided previously, we will examine heterogeneous effects of the treatment by FY 2018 enrollment:

1) Less than part-time (Rate of pursuit less than 0.50)
2) Part-time (Rate of pursuit equal to 0.50)
3) More than part-time, less than full-time (Rate of pursuit greater than 0.50, but less than 1)
4) Full-Time (Rate of pursuit equal to 1)

Inference Criteria, Including Any Adjustments for Multiple Comparisons:

We will use standard inference criteria. We will use two-tailed tests and three threshold *p*-values: 1%, 5%, and 10%. Given the very small cost of the intervention and the large educational benefits of increasing postsecondary enrollment intensity, any measurable effect is likely policy relevant.

Limitations:

The primary limitation of this study is the use of categorical enrollment intensity outcome variable. Analyzing changes in the actual number of credits enrolled would allow for a more precise estimate of the treatment effect.

Exploratory Analysis:

We will examine categorical (less than part-time; part-time; greater than part-time, etc.) changes in enrollment intensity from the prior year to the current.

We will explore more downstream and long-run outcomes of the intervention. Specifically, we plan to explore the impact of the intervention on enrollment intensity for the spring semester as well as the likelihood of completing a degree program. Finally, we will examine the potential impact of institutional mobility. Each of these outcomes represent a potential indirect impact of the intervention.