

*Instructions for Academic Performance Regressions of GPA against a variable for Gap Year and control variables for Sociodemographics, Academic Ability, and Curriculum.*

First, data should be collected for the following variables (variable names in *italics*):

- (i) GPA by class for each academic year, along with a first-semester GPA for everyone. For the Class of 2019 and all earlier classes for which you have gap year information, there should be academic-year GPAs for each of their four years, *not* cumulative GPAs, just GPAs for the courses in each given academic year (first-year *FYGPA*, sophomore *SOPHGPA*, junior *JRGPA*, and senior *SRGPA*, along with a first-semester GPA *FSGPA*). For the Class of 2020, there will be three academic-year GPAs plus *FSGPA*. For the Class of 2021, there will be two academic-year GPAs plus *FSGPA*, and for the Class of 2022 there will only be the first-semester GPA (*FSGPA*).
- (ii) For each student, define a binary variable  $GapYear = 1$  for students who took a year after high school graduation before college or university study; otherwise,  $GapYear = 0$ . We understand some schools code additional gap year information within their administrative systems. If you do, please include that information and consult with the PI about how to include it in the models.
- (iii) Academic Rating (*AR*). If your school doesn't use a numeric scale, you may turn the letter scale into a numeric scale. Please consult the PI if this is the case. If your school doesn't have an academic rating, but does have an overall rating, use that. If your school doesn't have a single AR, but multiple reader ARs, please average them together into one score. Alternatively, provide all to the PI who can average them before modeling.
- (iv) SAT/ACT scores (*SATEBRW*, *SATM*, or *ACTC*). If you have a mix of SAT and ACT scores, please convert one to the other using the College Board conversion tables so every student has one type of score. If your academic rating already contains standardized test scores as part of its formula, ignore this variable. If you are sending your data to the PI, don't worry about converting the SATs to ACTs, or visa versa. The PI has programming code to do that.
- (v) Sociodemographic variables (*Female*, *CaucAsian*, *QualifyNeedAid*, *Pelleligible*). Just like the *GapYear* variable, these variables are coded as a 0 or 1, where the group designated by the variable name (eg. Females, *Pelleligible*) get a 1 and everyone else is coded as 0.
- (vi) Curriculum variables (*PctCreditsNatSci*, *PctCreditsSocSci*, *PctCreditsHum*). This last set of variables is not required, but if it can be created, it greatly improves the models. For each academic year that a GPA is calculated you should also calculate the percentage of courses or credits (or the number of courses or credits) from each of the major divisions, natural sciences, social sciences, and humanities.

**Model 1:** For everyone with a *FSGPA*, regress *FSGPA* as the dependent variable and include *GapYear*, *AR*, and *SAT/ACT* as the independent variables. Repeat separately for those with *FYGPA*, *SOPHGPA*, *JRGPA*, and *SRGPA* in place of *FSGPA*.

**Model 2:** Repeat the 5 regressions from Model 1 but in addition to *GapYear*, *AR*, & *SAT/ACTs*,

add the sociodemographic variables to the models for each class.

**Model 3:** Repeat the 5 regressions from Model 2, but in addition to the list of independent variables from Model 2, add the curriculum variables if you are able to construct them.