

Data Request for CASSIE Participating Institutions

This document outlines the data requested from institutions participating in CASSIE. This particular document is a technical overview. It presumes that readers are already familiar with and endorse the CASSIE mission to take a “big data” approach to documenting the value added to college success by international education. Implementing the data request will likely require collaboration across multiple offices on your campus, including but not limited to Offices of International Education and Institutional Research and Effectiveness. Please share this document with all involved stakeholders for review. It should serve as the first level source of reference for those at your institution that are executing the data request.

Overview: The focus of this project is on IPEDS First Time Freshmen (FTF) in Fall 2010 and Fall 2011 who were seeking either an Associate’s, Bachelor’s, or Bachelor’s with combined Master’s degree. For students in the Fall 2010 and Fall 2011 cohorts, CASSIE protocols request data on (a) prior academic preparation (e.g. SAT, HSGPA), (b) demographic characteristics (e.g. sex, race/ethnicity, Pell receipt), (c) international education experiences (e.g. education abroad, advanced foreign language study, Title VI program participation), and (d) academic progress (e.g. credit hours earned, degrees awarded). We are requesting that these data be provided for each term a student is enrolled at your institution from matriculation through Summer 2017. The [Data Template \(Excel format\)](#) details the exact variables requested, and the [Data Submission & Definitions \(Word\)](#) document provides submission instructions and definitions for all these variables.

Methods: In order to assess the value-added impact of international education experiences, this project uses a combination of exact matching and propensity score matching analysis. This statistical model requires that we sample both treated students (i.e., students who have had intensive international education experiences) and controls (i.e., students lacking defined international education experiences). The matching analysis compares students who attend the same institution (exact match) and are similar in observed characteristics (propensity score match) but who differ in their international education experiences.

Sampling: Participating institutions will provide term-by-term data specified in the [Data Template](#) on all students in the Fall 2010 and Fall 2011 entering cohorts¹. This includes students that did and did not participate in international education experiences. Matching is a data intensive procedure, and statistical inference improves dramatically with large samples and a robust set of pre-treatment covariates. Each of the elements specified in the [Data Template](#) were chosen carefully and will either serve as conditioning or outcome variables. By requesting and utilizing data on all IPEDS FTF in Fall 2010 in 2011, we will improve the quality of the matches between treated and control students. This will lead to stronger inference about the impact of international education experiences. Participating institutions will only need to provide the data; all matches and statistical analyses will be conducted by CASSIE researchers.²

¹ Do not include students in your sample who are Non Resident Aliens. This stipulation is due to various regulations and data privacy practices regarding the collection of data on these students.

² For schools that prefer to conduct their own sampling rather than providing the full census of 2010 and 2011 entering cohorts, an alternative option would be that the schools construct the matches themselves. This would involve: (i) For each student that was treated, identifying 10 control students who did not have that international education experience but who enrolled in school up until the same term of treatment, and (ii) Once these observations are identified, performing propensity score matching utilizing a statistical software such as STATA. CASSIE researchers are willing to guide schools in this exact match sampling process, but it will require substantial work on part of the participating institution. To ease the data processing burden on participating institutions and to ensure uniformity in sampling, CASSIE recommends the first option--that is, providing the full cohort data and allowing CASSIE to create the matched treated and control samples.