

CT Preschool through Twenty & Workforce Information Network (P20 WIN)

DATA REQUEST DOCUMENT

This form and all attachments are to be submitted to the P20 WIN Data Governing Board for consideration and approval before the execution of each Data Request.

Information provided here is required to fulfill the mandatory provisions for written agreements according to the Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. §1232g and to comply with Unemployment Compensation (UC) considerations under state (CGS § 31-254) and federal (20 CFR 603) law. PERSONS WHO ARE NOT STATE GOVERNMENT EMPLOYEES OR OFFICIALS MAY NOT HAVE ACCESS TO PERSONALLY IDENTIFIABLE INFORMATION (Pii) FOR DOL UC RECORDS.

SECTION 1 – BASIC DATA REQUEST INFORMATION	
1.a Application Submission Date: 12/4/2019	
1.B TYPE OF SUBMISSION: <input checked="" type="checkbox"/> Initial Application <input type="checkbox"/> Continuation* <input type="checkbox"/> Modification**	
*Initial applications are approved for a given duration. Select 'continuation' if this request is to extend that timeframe. ** If you are requesting a modification to a request that has already been approved and is underway, select 'Modification' and address only the sections of this form which are being adjusted from the prior request.	
1.c P20 WIN Data Request Number: P20W_1911_5_0021	
1.d Data Request Title: Student Success, STEM Enrollment and Academic Environment: Effects of Grading Policies, Peers and Probation by Student Demographics	
1.e Agencies from which data is being requested (Check all that apply):	
<input type="checkbox"/> OEC <input type="checkbox"/> SDE <input type="checkbox"/> CSCU <input checked="" type="checkbox"/> Department of Labor <input type="checkbox"/> Cicu-IRPS <input checked="" type="checkbox"/> UCONN	
1.f Has this Data Request been discussed with program employees at the involved agencies?: <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No	
If yes please indicate with whom:	
Agency: University of Connecticut	Name: Lloyd Blanchard
Agency: Department of Labor	Name: Andrew Condon
Agency:	Name:
SECTION 2– REQUESTOR INFORMATION	
2.a Name (last, first): Ross, L. Stephen	
2.b Title: Professor of Economics	
2.c Organization: University of Connecticut	
2.d Mailing Address: 346 Oak Hall, Fairfield Way, University of Connecticut, CT 06269	
2.e Email Address: stephen.l.ross@uconn.edu	
2.f Phone Number	+1 860 486 3533
Fax Number	
2.g Provide Full name, organization, email address, and project role for ALL other persons within the requestor's organization who need to work with raw data, derived data or data output for this project before the P20 WIN Data Governing Board Participating Agencies involved have approved the data as appropriately aggregated for public release. Add additional space as necessary. Please print and attach a signed Confidentiality Agreement for each individual needing access to the resultant data set. (See Attachment C) Note: Persons who are not state government employees or officials may not have access to Pii for DOL UC records.	

1. Stephen L. Ross (University of Connecticut, Email: stephen.l.ross@uconn.edu, Role: Principal Investigator)
2. Talia Bar (University of Connecticut, Email: talia.bar@uconn.edu, Role: Principal Investigator)
3. Nishith Prakash (University of Connecticut, Email: nishith.prakash@uconn.edu, Role: Principal Investigator)
4. Deepak Saraswat (University of Connecticut, Email: deepak.saraswat@uconn.edu, Role: Research Assistant)

2.h Provide full name, organization, email address and project role for all other persons from the Participating Agencies or Contributing Institutions who have been identified as having legitimate interests in the evaluation who need to access raw data, derived data or any data output for this project before the P20 WIN Data Governing Board Participating Agencies involved have approved the data as appropriately aggregated for public release. For these individuals, access to the unit record data is limited. For example, named individuals may a) need access to data to support the data matching process or b) function as a subject matter experts from the Participating Agencies to support the data matching process or authorized requestor's analysis, or c) they may only have access to their own organization's data linked to other non-education records. After each name, please include an explanation as to why these individuals need access to the unit record data. **If not already signed, please print and attach a signed Confidentiality Agreement for each individual needing access to the resultant data set. (See Attachment C). Note: Persons who are not government employees or officials may not have access to Pii for DOL UC records.**

1. Lloyd Blanchard (University of Connecticut, Email: Lloyd.blanchard@uconn.edu, Role: UConn P20Win Representative)
2. Ann Wilhelm (University of Connecticut, Email: ann.wilhelm@uconn.edu, Role: Prepare Data)
3. Andrew Condon (Department of Labor, Email: andrew.condon@ct.gov, Role: DOL P20Win Representative)
4. Liam McGucken (Department of Labor, Email: liam.mcgucken@ct.gov, Role: Prepare Data)

SECTION 3 –ALIGNMENT OF PURPOSE& SCOPE

3.a Provide a brief summary of the Data Request. Include a description of the categories of data, years or data and cohorts of students – if applicable. Provide enough detail so that the Participating Agencies can understand what they need to provide.

We are requesting the data related to course taking and major decisions of cohorts of entering Freshman at the Storrs campus of the University of Connecticut with regard to their decision to major or not major in a STEM field, whether to take additional STEM courses, continuation at the university and completion of degree; as well as to examine the effect of probation on student outcomes including students pursuing STEM degrees. In addition to that, we request the DOL data for students who left UConn (after graduating or before that) related to their employment and industry. To analyze heterogeneities in connection to student outcomes, we also request data on teaching responsibilities (classes taught, characteristics and demographics of cohorts taught and instructor demographics) of instructors of courses they students undertook while at UConn. We request these data for cohorts of students starting to enter UConn from Fall 2008-Fall 2016.

3.b Clearly state the purpose of this request and describe how the purpose is an audit or evaluation of federal or state supported education program(s) (See Audit or Evaluation Exception 20 U.S.C. 1232g(b)(1)(C), (b)(3), and (b)(5) and §§99.31(a)(3) and 99.35).

There exist large differences across student demographics in student success and pursuit of STEM fields that often contribute to greater financial return from college degrees. For example, many minority groups are underrepresented in major U.S. universities and often have lower completion rates, and females are significantly underrepresented in STEM fields (such as math, science and engineering). Further, STEM courses are very challenging sometimes contributing to lower grades and causing students to interact with the university

probation system. Understanding reasons for these gaps might guide policy changes that could contribute to student success among underrepresented groups, benefiting both the members of these groups and the society as a whole.

In this study, we examine four types of factors that might influence student success overall and student success overall and within STEM fields: the grading policies of instructors early in the student's academic career, the attributes of the instructors in their early classes, the demographic and academic background of their peers, and if they struggle during their first few semesters the impact of being placed on probation. In terms of grading policies, we hypothesize that having an instructor who assigns harder grades is likely to impact course taking and major enrollment, and will investigate the impact of instructor grading practices with a special focus on STEM majors and on the differential effects of women and minorities. We also hypothesize that having an instructor that shares the demographic traits with a student or being enrolled in a section where more peers share demographic traits may influence both course taking and major choice, particularly in the area of STEM enrollment. Finally, we will examine the impact of being placed on probation on retention at the University, persistence in current major or in STEM fields in general, on eventual completion of a four year degree and on labor market outcomes.

3.c Describe the benefit that this audit/evaluation will provide to a local or State Education Authority or Agency (34 C.F.R. 99.1) and the state of Connecticut. Note requests that only provide audit or evaluation of non-state Participating Agencies will not be approved.

The university of has made a commitment to the state in terms of increasing its contribution to the state STEM workforce. Therefore, the university has a strong interest in understand the factors that influence persistence within STEM fields. The university also has a specific interesting in understanding factors that might limit the participation in STEM of traditionally underrepresented groups. Further, the university has a broader interest in the operation of its probation system and programs like UConn connect, which is intended to facilitate the success of students who end up on probation.

3.d Provide a brief description of the method for analysis.

For all the research questions mentioned in Section 3b above, we aim to use a regression framework to analyze the impact of teacher characteristics and academic disciplinary measures (academic probation in particular) on student's STEM course choices and academic and labor market outcomes (graduation, grades and employment etc.).

For STEM course choice and teacher characteristics, our empirical model compares students who took the same fall semester STEM course (c) in the same freshman year (t), but had instructors (p) who differ in attributes (X_{sct}) and the grades given in the course c during all other observed semesters ($AveGrade_{sc-t}$). Specifically,

$$y_{spct} = \beta AveGrade_{pc-t} + \gamma X_{pct} + \alpha Z_{spct} + \delta_{ctg} + \epsilon_{spct} \quad (1)$$

The dependent variable (y_{sct}) is the number of courses taken in the same department as course c by student s . The main coefficients of interest are β_1 , which captures the effect of grading practices on males, and β_2 which captures how this effect differs for females. The course-term fixed effects (δ_{ct}) assure the desired comparisons. We include a vector of student demographics (Z_{sct}). We will also investigate whether these grade and faculty attribute effects vary across different types of students, e.g. male vs. female, or non-Hispanic white vs. minority students.

For analysis of academic probation on student's academic and labor market outcomes, we estimate a regression discontinuity model as follows:

$$y_{stn} = \beta P_{stn} + \gamma GPA_{stn} + \alpha Z_{stn} + \delta_{tn} + \epsilon_{spct} \quad (3)$$

where P_{stn} is a dummy variable for whether the student was placed on probation in semester n , and GPA_{stn} is the student's cumulative Grade Point Average (GPA) in the same semester, which acts to capture any linear relationship between GPA and outcomes. In the regression discontinuity framework, GPA is referred to as the running variable. The sample is typically restricted to a bandwidth around the threshold for in this case probation in order to maximize the comparability of the samples below and above the threshold.

3.e Provide a description of the documents or reports that will reference data from this Data Request. For each report indicate the audience for the report and expected indicators or measures to be included in each.

The PI's will write a series of papers that will be provided to the Office of Institutional Research and Effectiveness on both factors related to success in STEM fields and on the UConn probation and warning system. A draft paper using only UConn data has already been provided to OIRE.

3.f In addition to utilizing a minimum cell size as documented by the Data Governance Procedure, identify/describe the statistical methods that will be used to minimize the risk of re-identification of PII for data to be published. Some typical methods are listed below; however, this list is not comprehensive. Options should be used in combination for maximum security. See the [USED Technical Brief 3](#) as a resource.

- **Suppression by:** cell, row, sampling (present data for a portion of students: e.g. 80%)
- **Blurring through:** aggregation of groups, rounding, use of ranges, use of top/bottom categories (e.g. 'less than 5%, greater than 95%),
- **Perturbation:** data swapping, adding noise, use of synthetic data

The sample involves nine cohorts of incoming freshman at UConn where a typical incoming class includes over 4,000 students. Even just focusing on students who enroll in STEM classes, our sample sizes number in the 10's of thousands. Our labor market sample will be smaller because some cohorts will have not entered the labor market yet, but even the smallest samples should number in the thousands observations. In the case of infrequent variables in the sample like Native American or Pacific Islander, we are either not presenting or in many cases not estimating (by combining with other groups) means and regression coefficients associated for these subsamples, even when the cell counts are above the minimum cell size. In this way, we practice relatively aggressive suppression, and we will continue this practice for the labor market analysis.

However, one form of analysis output deserves additional discussion. In RD analysis, it is common to illustrate the discontinuity in treatment (being below the threshold for probation) and in outcomes (GPA in future semesters, graduation) using pictures centered on the admission score threshold. These pictures are typically scatterplots for a sample of the treatment or outcome variables against the admissions score. A number of features act to protect the confidentiality of data when using these plots. First, we never plot individual observations. Instead, we plot the treatment and outcome shares or means for observations that fall into the same GPA bin. Specifically, we have been presenting these plots using 10 bins above and 10

bins below the threshold. Finally, we will never actually publish the specific numbers associated with these graphical points, and additional noise and blurring arises because numerical values can only be obtained by manually measuring the value from the figure.

SECTION 4 – PROTECTION OF CONFIDENTIALITY

4.a By what date will the datasets and all paper or electronic copies will be destroyed by the requestor?

After the Approved Requestor has received data files and the matrix of generic unique identifiers, the Approved Requestor has 12 months to complete analysis and destroy the data files. Note, that the Data Governing Board may set a shorter or longer time frame before the data files must be destroyed, and this 12 month limitation may be extended with written approval from the Data Governing Board. Requestors need to submit the Query Management Document as a 'continuation' for approval of an extension.

However, federal granting agencies and research universities like UConn that work with these agencies typically require that de-identified data be archived for 10 years after the completion of the project. This data would plan pending approval to archive this data on on the University of Connecticut's secure Research Network Attached Storage (RNAS) system. This archival period is important for two key reasons. First, the PI's will submit the work for publication in scholarly journals. The resulting review process will provide the state with detailed and expert review of the work performed by the PI. These external reviews are important to UConn in order to assess the weight that should be placed on the findings when making policy decisions. The PI's cannot submit work for review if it is infeasible to make revisions to the evaluation. Second, when the results of this study or program evaluation are published, UConn or the PI's may receive important feedback and advise, which may lead to desired enhancements of the program evaluation that was originally conducted. The cost of such enhancements would be increased dramatically if UConn is not able to archive the data for a period of time. Please note that just because the data is archived at the University does not imply that the PI's will have access to the data following the conclusion of the contract period. Such access would only be made available after both a new contract for use of the data was signed with the State of Connecticut and the Human Subjects Protocols were put back into place.

4.b With the addition of your signature at the bottom of this section, indicate that you agree to each of the following statements:

- I will only utilize the data received through this Data Request to meet the purpose as described. The approval given to receive data through this Data Request does not confer approval to use it for another purpose.
- I will not re-disclose the data received through this Data Request approval process to any public official who has not been authorized by the Data Governing Board to receive it, and who has not also signed a Personal Statement of Confidentiality and Non-Disclosure.
- I will take the necessary and appropriate precautions to safeguard personal information and will comply with all state and federal laws concerning the safeguarding and disclosure of such information.
- I will not use the datasets to re-identify individuals.
- I agree that approval to receive data does not convey ownership of the data.

- I agree that prior to the public release of any documents or reports generated from this Data Request I will supply all reports or documents to the Data Governing Board for review and verification that the intended purpose has been adhered to.
- I agree to store all resultant data, in print or electronic form, in a locked receptacle that can be accessed by authorized persons only.
- I agree to store all resultant data on secure desktop computers and in secure files to which access is restricted to authorized persons only.
- I agree that no resultant data may be transmitted via email or placed or stored on a mobile computing or storage device. For purposes of this agreement, a definition of "mobile computing device" includes, but is not limited to, notebooks, palmtops, PDAs, iPods®, Blackberry® devices, and cell phones with internet browsing capability. A "mobile storage device" includes but is not limited to, mobile computing devices, diskettes, magnetic tapes, external/removable hard drives, flash cards (e.g., SD, Compact Flash), thumb drives (USB keys), jump drives, compact disks, digital video disks, etc.
- I agree to destroy all confidential information obtained through this agreement as soon as such information is no longer needed.
- I agree that the Data Governing Board will be allowed access to monitor all authorized users to ensure such users adhere to the confidentiality requirements of the information obtained under this Agreement.
- I agree to comply with all provisions of the P20 WIN Data Request Management Procedure.

I  agree with each of the statements in section 4.b.

SECTION 5—ADDITIONAL INFORMATION

5.a Provide any additional comments that would be useful to the Data Governing Board in considering this request.

SECTION 6 – SIGNATURE OF REQUESTOR AND DATE

6.a Signature of Requestor

I understand that the entities that are providing data to P20 WIN (hereinafter Participating Agencies) have made reasonable efforts to ensure that the data available through P20 WIN are up-to-date, accurate, complete and comprehensive at the time of disclosure. These records reflect data as reported to the Participating Agencies by their data-submitting organizations for the reporting period indicated. Changes or updates to the data may occur after the time of disclosure and may impact data that have previously been made available. The Participating Agencies are not responsible for data that are misinterpreted or altered in any way. Derived conclusions and analyses generated from this data are not to be considered attributable to the Participating Agencies or the participating organization(s) from which the data originated.

I certify that the information supplied in this form, with attachments, is complete, accurate. The analysis will be conducted according to the protocol approved by the Data Governing Board, applicable federal, state and local laws regarding the protection of education records and unemployment insurance records. I will ensure that all protocol changes will be prospectively reviewed by the Data Governing Board. I will request approval from the Data Governing Board for changes to the Data Request and will not implement proposed changes until I receive Data Governing Board approval. I will promptly report to the Data Governing Board any related complaints, problems, and/or breaches of confidentiality.

Signature



Date

12/4/19