

# 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 UNIT LEVEL WAGE DATA.

SECTION 1 – BASIC DATA REQUEST INFORMATION	
<b>1.a Application Submission Date:</b>	
<b>1.b TYPE OF SUBMISSION:</b> <input type="checkbox"/> Initial Application <input type="checkbox"/> Continuation* <input checked="" 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.	
<b>1.c P20 WIN Data Request Number:</b> <a href="#">P20W_1712_5_0017</a>	
<b>1.d Data Request Title:</b> Evaluation of the Labor Market Effects of Technical High Schools in Connecticut	
<b>1.e Agencies from which data is being requested (Check all that apply):</b>	
X <input type="checkbox"/> SDE <input type="checkbox"/> CSCU                      X <input type="checkbox"/> Department of Labor <input type="checkbox"/> Cicu-IRPS <input type="checkbox"/> UCONN	
<b>1.f Has this Data Request been discussed with program employees at the involved agencies?:</b> <input type="checkbox"/> Yes    X <input type="checkbox"/> No If yes please indicate with whom:	
Agency:	Name:
Agency:	Name:
Agency:	Name:
SECTION 2– REQUESTOR INFORMATION	
<b>2.a Name (last, first):</b> Stephen Ross	
<b>2.b Title:</b> Professor of economics	
<b>2.c Organization:</b> University of Connecticut	
<b>2.d Mailing Address:</b> Dept. of Economics, 341 Mansfield Rd, Storrs CT 06269	
<b>2.e Email Address:</b> Stephen.l.ross@uconn.edu	
<b>2.f Phone Number:</b> 860-486-3533    Fax Number	
<b>2.g Provide Full name, institution, email address, and project role for ALL other persons working with data, derived data or data output for this project. 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)</b>	
Rick Cloud, Performance Office, Connecticut State Department of Education, <a href="mailto:Richard.Cloud@ct.gov">Richard.Cloud@ct.gov</a> , Data management: Mr. Cloud will be involved in preparing the data file of student data to be sent to DOL for the matching process.	
Eric Brunner, Department of Public Policy, University of Connecticut, <a href="mailto:Eric.Brunner@uconn.edu">Eric.Brunner@uconn.edu</a> , Principal Investigator (PI): Research Design, Coordinate Data Analysis and Analysis of Results using anonymized data for the evaluation of the effects of attending one of the schools in the Connecticut Technical High School System (CTHSS).	

Shaun Dougherty, Department of Public Policy, University of Connecticut, [Shaun.Dougherty@uconn.edu](mailto:Shaun.Dougherty@uconn.edu), PI: Research Design, Coordinate Data Analysis and Analysis of Results using anonymized data for the evaluation of the effects of attending one of the schools in the CTHSS.

Stephen Ross, Department of Economics, University of Connecticut, [Stephen.L.Ross@uconn.edu](mailto:Stephen.L.Ross@uconn.edu), PI: Research Design, Coordinate Data Analysis and Analysis of Results using anonymized data for the evaluation of the effects of attending one of the schools in the CTHSS.

Andrew Ju, Department of Economics, University of Connecticut, [Andrew.Ju@uconn.edu](mailto:Andrew.Ju@uconn.edu), Co-author and Visiting Faculty: Research Design, Coordinate Data Analysis and Analysis of Results using anonymized data for the evaluation of the effects of attending one of the schools in the CTHSS.

Erdal Asker, Department of Economics, University of Connecticut, [Erdal.asker@uconn.edu](mailto:Erdal.asker@uconn.edu), Research Assistant: Manage and Analysis of anonymized data under direction from the Principle Investigators.

### SECTION 3 –ALIGNMENT OF PURPOSE& SCOPE

**3.a Provide a brief summary of the Data Request.** Match longitudinal labor market and apprenticeship data data from Department of Labor (DOL) to the SASID's of a population of students who applied to one of the schools in the Connecticut Technical High School System (CTHSS) for admission for the fall semesters of the years 2006 to 2017. This data will include detailed administrative data on students including course taking from the State Department of Education (SDE). The list of students with data elements for matching attached (the first seven elements listed in P20Win Attachment AB SDE Elements tab) will be provided to the DOL by SDE.

**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). To evaluate the causal effects of attending one of the schools in the CTHSS on labor market outcomes over time including apprenticeships, employment, earnings and industry of employment, as well as educational outcomes including Career and Technical Education program completion and completion of programs in higher education. The State of Connecticut expends considerable resources running an independent high school system for the purpose of providing career and technical training within a formal school setting, but outside of traditional school district based high schools. This study will evaluate the impact of this opportunity on the post-high school outcomes of students who apply to participate in this system.

**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: An evaluation of the effectiveness of the CTHSS will help guide the resource allocation decision of policymakers, especially in the difficult budget environment that the State of Connecticut will face over the next several years. Further, the evaluation will aid the CTHSS mission directly by identifying areas where the system is having a successful impact on student outcomes and where the system is failing to enhance student outcomes.

**3.d Provide a brief description of the method for analysis.** Applicants to one or more the CTHSS schools submit applications in 8<sup>th</sup> grade and are given a numeric score by the school based on a series of factors that are available in the SDE longitudinal student database. This numeric score is recorded for all applicants and has been made available to Principal Investigators (PI's). For each year, every school established a threshold based on this scoring system, and applicants above the threshold are offered admission to the school. This structure allows us to conduct a Regression Discontinuity (RD) analysis where applicants just above and just below the threshold are compared controlling for the student's actual admission score. Analyses of the effects of school choice are typically biased because systematic factors determine who is admitted to the school and who is not. However, an RD analysis compares individuals who are very similar on the underlying factors that determined admission, and then tests for whether there is a discontinuity in average outcomes when comparing those individuals who have very similar scores with some just above the threshold (treatment group) and some just below (control group). RD techniques are well established in the Program Evaluation literature for estimating the causal effects of programs, especially within education.

**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 prepare both a formal report for the SDE, as well a more traditional article intended for distribution through working paper series and academic journals. The outcomes to be examined include employment in each year, overall incidence of unemployment post-school, earnings by year and overall, completion of apprenticeship programs, and indicators for whether working in Science, Technology, Engineering and Mathematics (STEM) intensive industries. On the education side, outcomes will include standardized test scores, completion of Career and Technical Education programs, high school graduation, and college completion and degree type. We will also analyze outcomes separately by subsets of applicants based on free and reduce price lunch status, minority status, English Language Learner status, pre-high school test scores, average test scores of the sending school or district, and average test scores of the Technical High School to which the student applied.

**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 10's of thousands of applicants. Our regression sample is reduced by failures to match applicants to SSAID, missing outcome variables, a bandwidth restriction where we limit the observations in our sample to be within a certain distance of their particular year and school cutoff and dropping school-year combinations where the data appear to violate the standard conditions for an RD analysis. After these restrictions, the estimation samples fall between 8 and 20 thousand observations. As a result, the means and regression coefficients presented typically represent thousands of observations or at least several hundred for variables like Asian or English Language learners. Our labor market sample will be smaller, but even the smallest samples should number several thousand 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 (attending at CTHSS school) and in outcomes (high school graduation or college attendance in our current analysis) 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 admissions score bin. Specifically, we have been presenting these plots using 10 bins above and 10 bins below the threshold. Second, the results are effectively blurred in the process of conducting the RD analysis. We are conducting what is referred to as a fuzzy RD because the assignment rule does not perfectly predict treatment, and so for each school and year we must identify the effective admission score threshold. This threshold has never been made public (and we will not make public), and is only identifiable using the confidential admissions data. Then, all observations for every school and year are centered on the appropriate threshold by calculating relative scores that are designed to take a value of zero at the threshold, and then the re-centered data with relative scores is pooled across all schools and years. Therefore, resulting bins are based not on individuals who share a specific admission score, but rather being the same number of points below or above an unknown admissions threshold, which varies over school and year. Therefore, a point on the graph that identifies average rate of high school graduation for individuals who are approximately 10 points above the threshold cannot be related to people whose admissions score fell into a specific bin or even to students who actually attended a technical high school. 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. For example, on our figure for high school graduation, the width of the dot used to illustrate the percent graduating from high school is greater than one percentage point in graduation rates so it is impossible to measure graduation rates from this figure with a high level of precision.

## 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.* The Connecticut Technical High School System evaluation project is conducted by the University of Connecticut at zero monetary cost to the SDE because the research is funded by a Federal grant from the Institute for Education Sciences (IES) in the U.S. Department of Education. The intention of the Principle investigators (PI's) is to obtain an additional grant to support the analysis of the labor market data. The PI's expect the duration of this grant to be five years, and so they would require access to the data for this five year period plus one additional year to complete the work. If the grant started on July 1, 2020, then the PI's would be expected to delete the provided data and any derived data sets on June 30, 2026 from any user accounts.

However, per IES policy, SDE and the PI's developed a Data Management Plan that requires that the de-identified dataset that was derived from FERPA protected data be archived for 10 years after the completion of the project. The final Data Management Plan specified that the data would be archived on the University of Connecticut's secure Research Network Attached Storage (RNAS) system. That specific archival requirement only applies to the SDE data, but we would hope to establish a limited archival period following the conclusion of the grant period during which the de-identified student level data with labor market variables attached would be archived either at the University or SDE.

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 SDE 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, SDE or the PI's may receive important feedback and advice, which may lead to desired enhancements of the program evaluation that was originally conducted. The cost of such enhancements would be increased dramatically if SDE 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 observe 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 Stephen L Roos, 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.

The Principle Investigators (PI's) at the University of Connecticut are currently analyzing educational outcomes for Connecticut Technical High School System applicants under a contract between SDE and the University of Connecticut. The contract is attached to this request. The contract allows for the option of including Department of Labor Data in the evaluation project if that data becomes available, and we would like to provide the labor market data to the PI's under the current contract if possible.

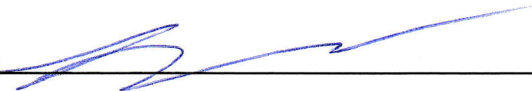
### SECTION 6 – SIGNATURE OF REQUESTOR AND DATE

#### 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

9/15/20