

# 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>	
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: <b>P20W_2102_12_0018B</b>	
1.d Data Request Title: College and Career Readiness & Placement Analysis	
1.e Agencies from which data is being requested (Check all that apply):	
<input type="checkbox"/> OEC <input checked="" type="checkbox"/> SDE <input checked="" type="checkbox"/> CSCU <input checked="" type="checkbox"/> Department of Labor <input checked="" 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: SDE	Name: Ajit Gopalakrishnan
Agency:	Name:
Agency:	Name:
SECTION 2– REQUESTOR INFORMATION	
2.a Name (last, first)	Bill Gammell, Ph.D.
2.b Title	Director, Office of Research & System Effectiveness
2.c Organization	Connecticut State Colleges & Universities (CSCU)
2.d Mailing Address	61 woodland Street, Hartford CT 06105
2.e Email Address	gammellw@ct.edu
2.f Phone Number	860-723-0054
	Fax Number
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)	
J.D. Mathewson, CSCU, Senior Research Associate, <a href="mailto:jmathewson@commnet.edu">jmathewson@commnet.edu</a> , 860-723-0007 Oscar Rivera, CSCU, Senior Programmer Analyst, <a href="mailto:Orivera@commnet.edu">Orivera@commnet.edu</a> , 860-723-0015 Jan Kiehne, CSCU, Senior Associate for Decision Support Services, <a href="mailto:kiehnej@ct.edu">kiehnej@ct.edu</a> 860-723-0236 Chuck Martie, SDE, Research & Evaluation, <a href="mailto:Charles.martie@ct.gov">Charles.martie@ct.gov</a> 860-713-6809 Danielle Bousquet, SDE, Analyst, <a href="mailto:Danielle.Bousquet@ct.gov">Danielle.Bousquet@ct.gov</a> David Alexandro, SDE, Analyst <a href="mailto:David.Alexandro@ct.gov">David.Alexandro@ct.gov</a> Briana Hennessy, SDE Intern, <a href="mailto:briana.hennessy@ct.gov">briana.hennessy@ct.gov</a>	

Laura Yahn, UCONN, Research Analyst, [laura.yahn@uconn.edu](mailto:laura.yahn@uconn.edu)  
Dan Sokol, UCONN, Data Warehouse Administrator [daniel.sokol@uconn.edu](mailto:daniel.sokol@uconn.edu)  
Youyou Zheng, UCONN, IR Specialist, [Youyou.zheng@uconn.edu](mailto:Youyou.zheng@uconn.edu)  
Joe Daniels, UCONN, Data Warehouse Developer, [joseph.daniels@uconn.edu](mailto:joseph.daniels@uconn.edu)  
Lloyd Blanchard, UCONN, Associate Vice Provost, [lloyd.blanchard@uconn.edu](mailto:lloyd.blanchard@uconn.edu)  
Maura Provencher, CCIC, Vice President, [provencher@theccic.org](mailto:provencher@theccic.org) \*  
Liam McGucken, DOL, Research Analyst, [liam.mcgucken@ct.gov](mailto:liam.mcgucken@ct.gov) \*  
Dana Placzek, DOL Research Analyst, [dana.placzek@ct.gov](mailto:dana.placzek@ct.gov) \*

\*Analysts from CCIC & DOL will not have access to unit record data about CT high school graduates that SDE obtains from the National Student Clearinghouse.

### 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 of data and cohorts of students – if applicable. Provide enough detail so that the Participating Agencies can understand what they need to provide.

This data request is an evaluation of the criteria and models for determining college readiness. At a simple level, the concept of College Readiness can be defined as achievement of certain cut scores on standardized tests (e.g. SAT, Accuplacer) or placement into college level math and English courses. This analysis will look at those approaches and also evaluate the determination of college readiness by identifying and assessing factors that are correlated with operational definitions of college readiness such as 1) HS course-taking patterns, 2) number of college credits earned within the first year, and 3) first-year average GPA. Operational definitions to be explored include, but are not limited to the New England Secondary School Consortium (NESSC) definition of College Readiness. The NESSC operational definition states that students were college ready if they a) completed 24 credit hours with a GPA of 2.5 within 1 academic year **or** b) if they enrolled in a 3<sup>rd</sup> semester.

The analysis may also include an evaluation of career readiness by looking at whether students who chose not to pursue a postsecondary credential successfully entered the workforce. Factors such as employment retention and change in wages received over time would be relevant to this analysis.

The analysis enabled by this data request includes the work outlined for the SLDS grant research question about college and career readiness which is further described in the accompanying memo developed by the planning group, “Research Topic: Predictive models for College and Career Success.” SDE and CSCU will share responsibility for the analysis, and work will be divided according to agency relevance, staff expertise, availability and interest.

Data needed from SDE includes the following but are further articulated in the spreadsheet for Attachments A & B.

- high school graduates (e.g. including but not necessarily limited to the classes of 2016 and 2017)
- test scores (including 11<sup>th</sup> grade SAT scores)
- course history and grades received (4 years available)
  - o fall 2014- present have course data
- attendance
- mobility (movement between HS’s and districts)
- discipline (in school / out of school suspensions)
- repetition of grade
- demographics (time spent with non-disabled peers and disability types)
- SDE’s measure of college readiness (e.g. Indicator 5 and indicator 6)
  - o Are they taking AP, IB, workplace experience, CTE courses = Indicator 5 (take 2 in one of those categories)
  - o Do they meet CR threshold on ACT, SAT, AP test or IB test (e.g. indicator 6)

Data needed from CSCU, UCONN and CCIC includes the following. Detailed elements are further described in Attachments A & B.

- student enrollment from summer of 2015 through most recent current term available (e.g. spring 2018)
- Placement data (e.g. SAT scores)
- initial math and English courses taken and grades received
- credits attempted and earned by term
- course enrollment and grades for each term
- term and overall GPA by term
- FAFSA completion

Data needed from CT DOL includes the following. Data elements are further described in Attachment A & B

- Quarterly wages received from 1 quarter prior to graduation to most recent quarter available post-graduation
- Industry code
- Employer code – for determining employment retention

CSCU has already received one initial data set under 0018 which will be used for this request and shared with SDE after this data request and accompanying MOA is finalized. Additional data sets are expected to be necessary during the duration of the analysis. For example, it may be possible that CCIC does not have certain elements available at the time of the first data match that may become available in a subsequent year. It may also be possible that CSCU would need an additional match to wage records at the department of labor to capture more recent employment data. If additional data sets and matches are required for these or other reasonable circumstances, they would be for elements listed in Attachments A&B unless a formal modification is made.

**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).

Every term, colleges and universities make determinations about students' College Readiness, and these decisions have significant implications for both students and the institution. Placement determinations affect the student's trajectory towards completion by potentially affecting course selection, the length of time the student needs to be in school and the availability of financial aid. The testing and placement process itself can be a challenging or negative experience for the student, and it is administratively time consuming and costly for the institution. Given the importance of minimizing negative outcomes for students and facilitating smooth and accurate placement, this evaluation of state education programs is an important necessity.

We also need to evaluate the state education program's ability to prepare students for career readiness. For students who choose to enter the workforce directly after high school, the expectation is that their selected education program prepared them for their career; therefore, this analysis will also evaluate factors related to career readiness.

**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.

Once we can validate or improve upon current processes for determining College & Career Readiness, there will be a benefit to the Connecticut State Department of Education (SDE) and institutions of Higher Education as well. SDE will benefit from having clearer information about the high school factors that are correlated with positive performance in college and career entry after high school. These factors can be used to adjust, if necessary, the state's "On Track" system that provides feedback to school administrators about students who appear to be "on" or "off" track towards high school graduation and college/career readiness. Connecticut institutions of higher education will benefit by having clearer information about college readiness factors to make the most accurate placement determinations possible. By increasing efficiency and accuracy in this area, institutions will likely be able to spend less time on test administration and more time supporting students as they begin their postsecondary

education experience. Most importantly, students will benefit by being guided towards the courses that are most likely to promote their academic and career success.

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

Our current efforts to test the success of Connecticut Community College students placed into varying levels of English and math courses have already led us to construct a comprehensive, longitudinal student-by-student database wherein each identifier is paired with demographic data, placement test data, course placement, course results, GPA, completion, persistence and retention scores. Using ordinary least squares regression and binary logistic regression, we've built a model that can predict outcomes using student data we already have in the CCC Institutional Research Database.

We can expand this project to make it more relevant and reliable, by adding more readiness predictors, expanding the student universe and adding more dependent variables to measure student success. By including SAT score variables and secondary school readiness indicators, we can build models that predict the relationship not just between those variables and success in math and English, but how concurrent our existing placement measures are with SAT scores, thus allowing us to pick the most accurate possible SAT scores to establish placement cutoffs in these courses. By expanding the student universe to UConn and CCIC students we can further test the reliability of these models, fine-tune the relationship between readiness indicators and success variables, and generate predictions that are relevant not just to CCC students but also the full Connecticut postsecondary student population.

The use of ordinary least squares and binary logistic regression with the existing longitudinal dataset has already yielded deep insights into the success of developmental education policy at the CCCs. The agencies may pursue additional modeling techniques such as random forests and regularized regression models to improve prediction accuracy and may include other variables into the models such as attendance, mobility, discipline, etc.

### **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.**

Reports generated will be released publically, but the primary audience will be policy leaders at SDE and CT postsecondary institutions, especially those with influence in academic affairs. The analysis will support policy and operational decisions in several ways by 1) improving determinations of placement at CSCU, 2) fine-tuning SDE's "On-track" information system for high school counselors, and 3) informing a Direct Admissions process to CSCU institutions. Metrics will be generated to address the following and other related questions.

- To what degree are the official SAT test scores captured and benefitting students with placement now?
  - o What SAT scores does SDE have and do CT institutions of higher education have the same for students who enroll?
  - o How many students had multiple SAT scores on record with SDE? Would their placement determination have been different if scores other than the 'official scores' were used?
  - o How many students were placed into a developmental math or English course that they might have been able to skip if their SAT or other test scores were known?
- To what degree do the current cut scores for SAT place students in the right level of English or math?
  - o Do students who are above the cut score pass the English or math course into which they were placed?
  - o Do students below the cut score pass the developmental course into which they were placed?
  - o How do students with scores close to the cut off fair?
  - o What combinations of test scores are the best predictors of accurate placement?
- To what degree is our placement process guiding students to the initial math and English course that best matches their level of readiness and best supports college success.

- What high school factors correlate with postsecondary evidence of college readiness based on students who demonstrated college readiness through their initial academic performance?
- How are college persistence, retention, academic performance (GPA) affected?
- To what degree do high school academic history factors correlate with career readiness for students who do not pursue a post-secondary education after high school completion?

**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

All data will be provided in aggregate. Cell size suppression, grouping, the use of ranges and top/bottom categories will be the primary approaches used to protect personally identifiable information and minimize the risk of re-identification.

#### **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 analytical process for this data request is anticipated to take four years. During this time, CSCU and SDE will receive at least one initial combined data set. In the event that additional data sets and matches are required during the analytical time period, all analytical data sets will be fully destroyed by the end of the fourth year, from the point of signature on the accompanying P20 WIN MOA. For example, If the MOA is signed on 02/15/2021, then all analytic data received for this request will be fully destroyed no later than 02/14/2025.

**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 William J Gannell agree with each of the statements in section 4.b.

I Ajit Lopalakishua 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 William J Gannell Date 2/1/2021

Signature

Ajit Gopalakrishna

Date 2/1/2021