

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: 1/22/2021

1.b TYPE OF SUBMISSION: Initial Application Continuation* 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:

1.d Data Request Title: The effects of affordable pre-Kindergarten on school diversity, student outcomes, and parental labor supply

1.e Agencies from which data is being requested (Check all that apply):

OEC SDE CSCU Department of Labor Cicu-IRPS UCONN

1.f Has this Data Request been discussed with program employees at the involved agencies?: Yes No

If yes please indicate with whom:

Agency: Department of Education

Name: Ajit Gopalakrishnan

Agency: OPM

Name: Scott Gaul, Katie Breslin

Agency: DOL

Name: Patrick Flaherty

Agency: OEC

Name: Rachel Leventhal-Weiner

Agency: Office of the Governor

Name: Mohit Agarwal

SECTION 2– REQUESTOR INFORMATION

2.a Name (last, first) Seth Zimmerman

2.b Title Associate Professor of Economics

2.c Organization Yale University School of Management

2.d Mailing Address PO Box 208200, New Haven, CT 06520-8200

2.e Email Address seth.zimmerman@yale.edu

2.f Phone Number 413-478-0466

Fax Number N/A

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.

Seth Zimmerman, Yale University, seth.zimmerman@yale.edu, primary investigator

John Eric Humphries, Yale University, johneric.humphries@yale.edu, co-investigator

Xiaoyang Ye, Brown University, xiaoyang.ye@brown.edu, co-investigator

Christopher Neilson, Princeton University, cneilson@princeton.edu, co-investigator

Adam Kapor, Princeton University, akapor@princeton.edu, co-investigator

Manuel Martinez, Yale University, manuel.a.martinez@yale.edu, research professional

Maria Elena Guerrero, Yale University, mariaelena.guerreroamezaga@yale.edu, research professional

Jan Knuf, Yale University, jan.knuf@yale.edu, research professional

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.**

Marquelle Middleton, New Haven Public Schools, MARQUELLE.MIDDLETON@new-haven.k12.ct.us

Michele Bonanno, New Haven Public Schools, MICHELLE.BONANNO@new-haven.k12.ct.us

Mr. Middleton is the director of school choice at the New Haven Public Schools and Ms. Bonanno is the magnet school coordinator. They will be providing data from NHPS for the merge to P20 WIN records.

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.

This data request will enable a comprehensive analysis of the benefits of free pre-Kindergarten to students, localities, and the state. We are requesting SDE-SLDS, OEC, and DOL-UI data from P20 WIN Data Agencies for the following groups: 1) students who applied to NHPS pre-K lotteries between 2003 and 2020; 2) students enrolled in NHPS schools over this period; 3) students who enrolled in other Connecticut pre-K programs over this same period. For each data set, we are requesting the period 2003-2020 or the largest time-window available in this range. For the DOL data, we are requesting 2000-2020 as this will provide three years of earnings and employment trends for parents prior to the first lottery in 2003.

P20-WIN data are crucial to this project for three reasons. First, SDE data include records for students who attend districts other than NHPS. This is critical to our research design, since many students who apply to or start out in New Haven Pre-K programs eventually end up in other Connecticut districts. A credible study of NHPS pre-K programs must include outcomes for these students. We will then use the data to study future school performance, graduation, and college enrollment.

Second, DOL data will allow us to see if labor-market attachment increases shortly after high school graduation for the earlier cohorts and, if linkage is possible, how parental labor supply changes in response to receiving a slot in the NHPS pre-K programs. We are also interested in outcomes related to benefits receipt for both parents and children, if such data are available. We provide more detail on programs of specific interest below. We look forward to communicating directly with DOL about data availability and merge procedures and are happy to provide any support we can in this process.

Third, we are requesting OEC data on Care4Kids and the more recent Early Childhood Information System to evaluate if receiving a NHPS pre-K slot causes families to substitute away from other state-funded programs.

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).

We propose to evaluate the impacts of free pre-Kindergarten (pre-K) on short- and long-run outcomes for students and families by combining data from the New Haven Public Schools (NHPS) with data from P20-WIN. The goal is to provide a comprehensive picture of the benefits of pre-K programs to individuals, localities, and the state.

This is a question of broad policy interest, and New Haven is an ideal setting in which to investigate it. NHPS has multiple magnet schools, which offer 1- or 2-year pre-K programs. For students who receive a slot, these programs provide free high-quality pre-K with few restrictions on eligibility. Ordinarily, understanding the causal effects of free pre-K on students and families is hard to do because students who enroll in free pre-K may be different than those who do not. In this case, however, NHPS' system of awarding pre-K spots using lotteries allows us to conduct a convincing evaluation of the effects of pre-K access. We will compare outcomes for families and students who receive randomly-allocated offers of pre-K placement to those who do not.

The project will address four research questions:

1. How does access to free pre-K affect students' academic, labor market, and social safety net outcomes?

2. How does access to free pre-K affect students' use of other state- and federally-subsidized pre-K programs, such as Head Start, Care 4 Kids, School Readiness, and Smart Start?
3. How does access to free pre-K affect parents' labor force participation, earnings, and participation in safety net programs in the short- and long-run?
4. How do free pre-K programs affect the racial and ethnic integration of school districts?

Overall, we are providing a rigorous evaluation of the pre-K programs offered by NHPS's magnet schools. These programs are supported by both state and federal funds.

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.

Broadly speaking, answers to the proposed research questions will support P20 WIN's research agenda, particularly with reference to understanding (1) students at the lowest levels of readiness and (2) overcoming barriers to success.

Concretely, this research will provide insight into four pathways through which free pre-K may affect student outcomes, parental outcomes, and utilization of other state resources: (1) we will evaluate whether pre-K programs affect student enrollment, academic achievement, and graduation; (2) we will evaluate whether free pre-K programs affect parents' labor supply. Receiving a slot may reduce parents' time out of the labor force or cause some parents to enter the labor force; (3) we will evaluate whether pre-K programs affect college enrollment, employment, and earnings for affected students; and (4) we will evaluate whether access to free pre-K decreases use of other childcare subsidies and other social safety net programs among parents and children.

Findings from this research will help state and local policy makers understand the benefits of New Haven's free pre-K programs, how these programs change labor supply, and potentially how free pre-K helps families substitute away from other state and federal programs. We believe the findings of this research will be informative to Connecticut as a whole and are relevant to the larger national debate on early childhood education. In particular, we anticipate being able to provide a comprehensive account of the benefits of pre-K to different stakeholders. In addition, the research will help determine if the pre-K programs increase the diversity of the students who proceed to enroll in magnet schools for primary education.

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

The proposed research would utilize the variation in pre-K placement associated with the pre-K lotteries to estimate the causal effect of receiving a pre-K offer on student and family outcomes. The research design allows us to estimate the benefits of pre-K for those who applied and received an offer as well as those who applied, received an offer, and enrolled in a program.

One technical challenge we anticipate is that families may apply to multiple pre-K magnet programs, so applicants who list multiple schools are more likely to get a slot. This can be addressed by controlling for application sets. In addition, applications provide useful information to the researchers, as they reveal information on families' underlying preferences or potential needs. For example, a family that only applies to a single magnet school pre-K program likely has different preferences or different outside options than a family that applies to every magnet pre-K program available. This unique institutional setting allows us to learn about the effects of the pre-K programs, and how those effects may depend on the preferences and needs of the families.

As part of the on-going research relationship with NHPS, we have generated preliminary evidence relating to research questions (1) and (2) above. First, we find that among those who apply, students who are offered pre-K slots score 0.2 standard deviations higher on math tests between grades three and eight, and more than 0.1 standard deviations higher on reading tests. Second, families who are offered slots are 20 percentage points more likely to then matriculate into NHPS for kindergarten. This 20 percentage point increase is also true for suburban applicants, who are an important population for increasing diversity within NHPS. These promising short-run findings suggest the possibility of long-run gains.

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.

We expect to submit a research paper for publication in a scholarly journal, the audience of which will be researchers as well as policy makers around the world. We will also submit a policy report to the New Haven Public Schools and relevant state agencies.

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

It is expected that suppression and blurring will be the primary approaches used to protect personally identifiable information (PII) and minimize the risk of re-identification. We will not report any individual or personally identifiable information. Cells containing or referring to a count of less than 10 will be suppressed along with the cell with the next to smallest value. Additional strategies that may be used included providing multiple year averages for key data points. Some of the output will be model-based estimates and we will suppress un-needed coefficients (for example, district and year fixed effects from an OLS regression).

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.

All unit record data resulting from the query to the system will be destroyed by the research team within 12 months of the date of the receipt of the analytical data sets by the researchers.

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 Seth Zimmerman

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 proposed project relies on combining NHPS data on school choice, school enrollment, and academic outcomes with P20-WIN data on academic and labor market outcomes. NHPS will provide their part of the data as part of an ongoing partnership with the researchers. *Attached to this proposal is a letter from Assistant Superintendent Ivelise Velazquez describing NHPS' support.* The current partnership builds on a productive working relationship between the researchers and NHPS. The results of this partnership include both research published in top academic journals and concrete changes in NHPS' school choice policies. We build on relationships, infrastructure, and personnel already in place. We would be happy to extend our research partnerships to other participating agencies to help support policy relevant questions that this project may be able to answer.


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 January 22nd, 2021