Datasets from an impact evaluation of a targeted prekindergarten program

Bianca Montrosse-Moorhead a, *, Shaun M. Dougherty b, Tamika P. La Salle a, Jennie M. Weiner c, Hannah M. Dostal d

a Department of Educational Psychology, University of Connecticut, USA
b Department of Leadership, Policy, and Organizations, Vanderbilt University, USA
c Department of Educational Leadership, University of Connecticut, USA
d Department of Department of Curriculum and Instruction, University of Connecticut, USA

Article history:
Received 21 March 2019
Received in revised form 18 November 2019
Accepted 18 November 2019
Available online 29 November 2019

Keywords:
Prekindergarten
Evaluation dataset

ABSTRACT

Given general trends in extant research on the impact of prekindergarten and that the structure and implementation of prekindergarten programs vary by state, researchers, educators, and policymakers are raising questions about what works, for whom, under what conditions, and the cost-benefit of such endeavors. Yet not all states have formally examined program impacts and few datasets have been expressly collected to evaluate effects. The current data article represents empirical examinations in the state of Connecticut on the comparability of treatment and control groups, tests of the robustness of impact estimates, and the psychometric properties of outcome measures. Stata code for replication purposes is included.

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
2. Experimental design, materials, and methods

2.1. Ethical statements

This study was approved by the University of Connecticut’s Institutional Review Board (Protocols H14–193 and H15–247), and carried out in accordance with a Memorandum of Agreement with the Connecticut State Department of Education and Office of Early Childhood and the Family Educational Rights and Privacy Act of 1974.
### Table 1
Data codebook for the datasets used by authors.

| P20 WIN Index # | Source | Data Category | Data Element Name | Data Element Definition |
|----------------|--------|---------------|-------------------|------------------------|
| 17             | SDE-SLDS | Student Reporting | EntryDate | Date of student’s entry into last facility in PSIS, as YYYYMMDD. |
| 18             | SDE-SLDS | Student Reporting | ExitDate | Date of student’s exit into last facility in PSIS, as YYYYMMDD. |
| 19             | SDE-SLDS | Student Reporting | ReportingDistrictName | Name of district responsible for reporting the student’s enrollment. |
| 21             | SDE-SLDS | Student AmericanIndianOrAlaskaNative | | Y-Yes if child is American Indian or Alaska Native. Otherwise N—No or NS-missing. |
| 22             | SDE-SLDS | Student Asian | | Y-Yes if child has origins in any of the original peoples of the Far East, Southeast Asia, or the Indian Subcontinent. Otherwise N—No or NS-missing. |
| 23             | SDE-SLDS | Student BirthYear | | Year portion of date of birth. |
| 24             | SDE-SLDS | Student BlackOrAfricanAmerican | | Y-Yes if child has origins in any of the black racial groups of Africa. Otherwise N—No or NS-missing. |
| 26             | SDE-SLDS | Student FreeReducedLunchEligible | | Indicator of student eligibility for federal free/reduced lunch program at any time. |
| 27             | SDE-SLDS | Student Gender | | Gender of the child. M = Male, F = Female. |
| 30             | SDE-SLDS | Student HispanicOrLatino | | Y-Yes if child’s ethnicity is Cuban, Mexican, Puerto Rican, south or central American, or other Spanish culture or origin regardless of race. Otherwise N—No or NS-missing. |
| 35             | SDE-SLDS | Student LastFacility1_SchoolName | | Name of last facility. |
| 44             | SDE-SLDS | Student NativeHawaiianOrOtherPacificIslander | | Y-Yes if child has origins in any of the original peoples of Hawai‘i, Guam, Samoa, or other Pacific Islands. Otherwise N—No or NS-missing. |
| 54             | SDE-SLDS | Student White | | Y-Yes if child has origins in any of the original peoples of Europe, Middle East, or North Africa. Otherwise N—No or NS-missing. |
| 55             | SDE-SLDS | Student Assessment Administered_FallOfYear | | Year of fall of school year in which the test was administered. |
| 56             | SDE-SLDS | Student Assessment Administered_SIFYear | | Year of spring of school year in which the test was administered. |
| 58             | SDE-SLDS | Student Assessment LevelScore | | Student’s performance level. |
| 59             | SDE-SLDS | Student Assessment ScaleScore | | Conversion of a student’s raw score on a test to a common scale that allows for a numerical comparison between students. |
| 60             | SDE-SLDS | Student Assessment TestType | | Name of standardized assessment. |
| 129            | SDE-SLDS | Student SASID | | State assigned student identification number |
| 393            | OEC | Early Childhood Enrollment ID | | Enrollment ID is an ID number associated with SASID and specific funding type and dates for an enrollment record. |
| 394            | OEC | Early Childhood FacilityCode | | Facility code that is associated to the Early Care and Education Facility or school (7-digit code). |
| 395            | OEC | Early Childhood FacilityName | | Name of the Early Care and Education Facility or school. |
| 397            | OEC | Early Childhood LastName | | Child’s legal first name that appears on the child’s birth certificate or document indicating legal name change. |
| 398            | OEC | Early Childhood FirstName | | Child’s legal last name that appears on the child’s birth certificate or document indicating legal name change. |
| 399            | OEC | Early Childhood MiddleName | | Child’s legal middle name that appears on the child’s birth certificate or document indicating legal name change. |
| 401            | OEC | Early Childhood DateOfBirth | | Child’s date of birth. Format MM/DD/YYYY. |
| 402            | OEC | Early Childhood GenderCode | | Gender of the child. |
| 408            | OEC | Early Childhood IsHispanic | | Y-Yes if child’s ethnicity is Cuban, Mexican, Puerto Rican, south or central American, or other Spanish culture or origin regardless of race. Otherwise N—No or NS-missing. |

(continued on next page)
2.2. Collection of student educational record data

The Connecticut State Department of Education and Office of Early Childhood maintain monthly and yearly educational records for all students enrolled in public schools (prekindergarten through grade 12). This data includes, for each student, a unique identification number, name, date of birth,
Table 2
Covariate balance at birthday cut-off.

|                      | Male        | White       | Black       | Latino      | Asian       | Other       | Low Income  |
|----------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| **IK bandwidth**     | 0.066 (0.137) | 0.062 (0.115) | 0.075 (0.145) | -0.197 (0.141) | 0.028 (0.027) | 0.039 (0.058) | 0.016 (0.121) |
| **N**                | 236         | 285         | 244         | 266         | 286         | 285         | 304         |
| Bandwidth = 120 days | 0.106 (0.194) | 0.157 (0.167) | -0.022 (0.195) | -0.164 (0.198) | 0.024 (0.021) | 0.122 (0.093) | 0.110 (0.161) |
| **N**                | 139         | 139         | 139         | 139         | 139         | 139         | 139         |
| Bandwidth = 150 days | 0.063 (0.168) | 0.123 (0.145) | 0.035 (0.170) | -0.181 (0.174) | 0.023 (0.028) | 0.098 (0.075) | 0.041 (0.150) |
| **N**                | 178         | 178         | 178         | 178         | 178         | 178         | 178         |
| Bandwidth = 180 days | 0.079 (0.151) | 0.098 (0.131) | 0.071 (0.152) | -0.209 (0.157) | 0.022 (0.029) | 0.070 (0.067) | 0.020 (0.139) |
| **N**                | 233         | 233         | 233         | 233         | 233         | 233         | 233         |
| Bandwidth = 210 days | 0.066 (0.138) | 0.066 (0.121) | 0.081 (0.137) | -0.201 (0.143) | 0.025 (0.028) | 0.047 (0.061) | 0.026 (0.130) |
| **N**                | 259         | 259         | 259         | 259         | 259         | 259         | 259         |
| **Control Mean**     | 0.529       | 0.235       | 0.382       | 0.441       | 0.000       | 0.029       | 0.735       |

**Note.** Heteroskedasticity robust standard errors clustered by birthdate are in parentheses. Each coefficient is the reduced form estimate of the relationship between being eligible for state-funded prek based on birthdate and the listed covariate. The coefficients shown are generated by local linear regression using a triangular kernel and specified bandwidth. The center of the specified bandwidth is the cut-off point. Also listed is the mean of the covariate for students just before the threshold for qualifying for prekindergarten. *p < .10, **p < .05, ***p < .01.
race/ethnicity, whether the student participates in the federal free and reduced priced lunch program, the school attended, and the school district in which the school is located. These data were provided to the research team through a Memorandum of Agreement.

More specifically, datasets are available through Connecticut’s P20 WIN. To access datasets, researchers must apply to become an authorized representative. The request is then reviewed by members of the P20 WIN Data Governing Board, and if approved, a Memorandum of Agreement (MOA) specific to the data request is executed. Once executed, the authorized representative receives a matched dataset. All requestors are required to pay a fee. Under special circumstances the fee may be waived or reduced at the discretion of the Data Governing Board. The authorized representative and anyone on their team who has signed Personal Statements of Non-Disclosure is able to work with the data. All publicly shared information as a result of dataset use must be sufficiently aggregated to maintain confidentiality. A detailed articulation of how to access datasets is included in the P20 WIN Data Request Management Procedure manual available at http://www.ct.edu/files/pdfs/P20WIN-

### Table 3
Testing assumptions about functional form and bandwidth of impact estimates.

|                      | Broad Math (SE) | Oral Language (SE) | Basic Reading (SE) | Picture Vocabulary (SE) |
|----------------------|----------------|-------------------|--------------------|-------------------------|
| Linear, BW = IK      | 10.152***      | 5.459             | 10.828***          | 4.425                   |
| N                    | (3.252)        | (7.364)           | (3.079)            | (4.906)                 |
| N                    | 273            | 199               | 190                | 251                     |
| N                    | 11.080***      | 12.954            | 13.125***          | 6.224                   |
| N                    | (4.069)        | (9.794)           | (4.221)            | (7.588)                 |
| N                    | 273            | 199               | 190                | 251                     |
| N                    | 10.302***      | 23.600**          | 20.585***          | 15.858*                 |
| N                    | (5.217)        | (11.041)          | (6.743)            | (9.575)                 |
| N                    | 425            | 425               | 425                | 418                     |
| N                    | 10.601***      | 4.842             | 10.319***          | 4.696                   |
| N                    | (3.370)        | (5.750)           | (3.381)            | (5.111)                 |
| N                    | 219            | 219               | 219                | 215                     |
| N                    | 10.878**       | 12.113            | 13.432***          | 8.217                   |
| N                    | (4.418)        | (9.524)           | (4.014)            | (7.929)                 |
| N                    | 219            | 219               | 219                | 215                     |
| N                    | 10.242***      | 4.028             | 10.352***          | 4.426                   |
| N                    | (3.099)        | (5.248)           | (3.152)            | (4.629)                 |
| N                    | 255            | 255               | 255                | 251                     |
| N                    | 11.346***      | 10.144            | 12.568***          | 6.213                   |
| N                    | (4.167)        | (8.969)           | (3.853)            | (7.586)                 |

Note. *p < .10, **p < .05, ***p < .01.

### Table 4
Test of robustness of impact estimates using different specifications with optimal bandwidth.

|                      | Broad Math (SE) | Oral Language (SE) | Basic Reading (SE) | Picture Vocabulary (SE) |
|----------------------|----------------|-------------------|--------------------|-------------------------|
| Baseline IK bandwidth model | 7.198*** | 3.000             | 13.028***          | 1.216                   |
| N                    | (2.865)        | (6.005)           | (4.189)            | (4.817)                 |
| Covariate-adjusted model | 6.618*** | 1.349             | 12.658***          | 0.130                   |
| N                    | (2.361)        | (4.684)           | (4.014)            | (3.892)                 |
| School clustered SE model | 6.618*** | 1.349             | 12.658***          | 0.130                   |
| N                    | (2.631)        | (5.124)           | (4.027)            | (4.075)                 |
| School FE model      | 12.133**      | 5.674             | 8.524              | 9.038                   |
| N                    | (5.865)        | (14.081)          | (15.916)           | (6.443)                 |
| N                    | 299            | 236               | 113                | 262                     |

Note. The coefficients shown were generated using OLS with the stated bandwidth and a triangular kernel. Models specified with covariates include indicator variables for gender, race/ethnicity, and eligibility for subsidized meals. *p < .10, **p < .05, ***p < .01.
2.3. Collection and measurement of student outcome data

Student outcome data were collected by a team of 58 trained and certified assessors using the Woodcock Johnson IV Tests of Achievement (WJ-IV) and the Peabody Picture Vocabulary Test 4th Edition (PPVT-4). Data from these assessments were then scored and composite measures of student achievement were entered into a datafile by the research team. Four composite measures in total were used and entered into the datafile.

The Basic Reading composite included two subtests from the WJ-IV, specifically Word Attack and Letter-Word Identification. The Letter-Word Identification subtest measures pre-reading skills including letter and word recognition and identification skills. The Word Attack subtest measures phonics and decoding skills. The reliability statistics for individual tests range from 0.90 to 0.94 [2].

The Oral Language composite included two subtests (the Picture Vocabulary and Oral Comprehension) from the WJ-IV. The Picture Vocabulary subtest primarily assesses expressive vocabulary, though early items provide some information about receptive vocabulary skills. The Oral Comprehension subtest measures the ability to understand short, orally presented, passages. Subtest reliability statistics range from 0.82 to 0.88 [2].

The PPVT-4 measured participants’ picture vocabulary, or students’ word knowledge ability. Spearman-Brown corrected, split-half, and alpha reliabilities consistently fall above 0.94, indicating solid internal consistency evidence [3].

The Broad Mathematical Skills composite consisted of three measures from the WJ-IV, specifically Applied Problems, Calculation, and Math Fluency subtests. The Applied Problems subtest was used to measure mathematics problem solving skills. The Calculation subtest measures students’ ability to complete items ranging from writing numbers to performing numerical operations. And, the Math Fluency subtest measures students’ ability to solve problems containing numerical operations quickly. The reliability statistics for individual tests range from 0.96 to 0.97 [2].

2.4. Statistical analysis

All statistical analyses were conducted using Stata. For all student outcomes, the standard score with a mean of 100 with a standard deviation of 15 was used. A multilevel regression model with
cluster standard errors was used because it does not force the same parametric assumptions about the distribution of the error terms at either the school or student level, or the correlation among them. The robustness of analysis decisions were evaluated through other specifications, including adding covariates, using an alternate clustering approach, and using fixed effects for each site. Unless otherwise indicated, \( p < 0.05 \) was considered statistically significant. The Stata code for analyzing data is included as a supplementary file.

**Acknowledgments**

This paper was prepared using datasets provided under contracts with the Connecticut Academy of Science and Engineering (CASE), the Connecticut State Department of Education (CSDE), and the Connecticut Office of Early Childhood (OEC). The content does not necessarily reflect the views or policies of CASE, CSDE, or OEC, and should not be construed as an endorsement by entities.

**Conflict of Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.104881.

**References**

[1] B. Montrosse-Moorhead, S.M. Dougherty, T.P. La Salle, The overall and differential effects of a targeted prekindergarten program: evidence from Connecticut, Early Child. Res. Q. 48 (2019) 134–145.

[2] K.S. McGrew, E.M. LaForte, F.A. Schrank, Technical Manual: Woodcock-Johnson IV, Riverside, Rolling Meadows, IL, 2014.

[3] Pearson Education, Inc, Peabody picture vocabulary test (4th ed.) (PPVT™-4). http://www.pearsonclinical.com/language/products/100000501/peabody-picture-vocabulary-test-fourth-edition-ppvt4.html?PId=PAa30700#tab-scoring, 2011.