

## **Data Description**

### **What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output**

by Diether W. Beuermann, C. Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo

March 2022

#### **Overview**

There are 7 raw administrative datasets used in this study. Below we detail each dataset and provide the information on how to access them:

1. School applications, performance on the Secondary Entrance Assessment (SEA), and school allocations were obtained from the Ministry of Education of Trinidad and Tobago. The datasets obtained under confidentiality agreements included the SEA cohorts from 1995 through 2012. Access is free of charge, but the process may take around 6 months. It can be obtained by submitting a research proposal to:

Division of Educational Research and Evaluation  
Education Towers  
Level 14, No. 5 St. Vincent Street.  
Port of Spain.  
Trinidad and Tobago  
Email: [directordre@moe.edu.tt](mailto:directordre@moe.edu.tt)  
Telephone: (868) 625-0806  
Telephone: (868) 628-7350  
Fax: (868) 628-9165  
Web: <https://www.moe.gov.tt/educational-research-and-evaluation/>

2. Performance on the National Certificate of Secondary Education (NCSE) was obtained from the Ministry of Education of Trinidad and Tobago. The datasets obtained under confidentiality agreements included the NCSE cohorts from 2009 through 2015. Access is free of charge, but the process may take around 6 months. It can be obtained by submitting a research proposal to:

Division of Educational Research and Evaluation  
Education Towers  
Level 14, No. 5 St. Vincent Street.  
Port of Spain.  
Trinidad and Tobago  
Email: [directordre@moe.edu.tt](mailto:directordre@moe.edu.tt)  
Telephone: (868) 625-0806  
Telephone: (868) 628-7350  
Fax: (868) 628-9165  
Web: <https://www.moe.gov.tt/educational-research-and-evaluation/>

3. Examination data at the individual level for the Caribbean Secondary Education Certification (CSEC) was obtained from the Caribbean Examinations Council. The datasets obtained under confidentiality agreements included the CSEC cohorts from 1993 through 2016. Access is free of charge, but the process may take around 6 months. It can be obtained by submitting a research proposal to:

Caribbean Examinations Council  
Prince Road  
Pine Plantation Road  
St. Michael BB11091  
Barbados, W.I.  
Telephone: (246) 227-1700  
Fax: (246) 429-5421  
Email: CXCEZO@CXC.ORG  
Web: [www.cxc.org](http://www.cxc.org)

4. Examination data at the individual level for the Caribbean Advanced Proficiency Examination (CAPE) was obtained from the Caribbean Examinations Council. The datasets obtained under confidentiality agreements included the CSEC cohorts from 2005 through 2016. Access is free of charge, but the process may take around 6 months. It can be obtained by submitting a research proposal to:

Caribbean Examinations Council  
Prince Road  
Pine Plantation Road  
St. Michael BB11091  
Barbados, W.I.  
Telephone: (246) 227-1700  
Fax: (246) 429-5421  
Email: CXCEZO@CXC.ORG  
Web: [www.cxc.org](http://www.cxc.org)

5. Arrest records were obtained from the Trinidad and Tobago Police Service covering the period between January 1990 and May 2017. Obtaining this database did not require confidentiality agreements and, therefore, the raw database is included in the replication package. The database is free of charge, but the process may take around 6 months. However, data is provided. It can also be obtained by submitting a research proposal to:

Crime and Problem Analysis Branch of the Trinidad and Tobago Police Service  
112 MATCO Building  
Henry Street, Port of Spain  
Trinidad and Tobago  
Phone: (868) 625-5148

Fax: (868) 625-4710

Email: [capa@ttps.gov.tt](mailto:capa@ttps.gov.tt)

Web: <https://www.ttps.gov.tt/Branches/Branches-A-E/Crime-And-Problem-Analysis>

6. Birth records were obtained from the Trinidad and Tobago Registrar General covering the period between January 2010 and September 2016. The datasets were obtained under confidentiality agreements. Access is free of charge, but the process may take around 6 months. It can be obtained by submitting a research proposal to:

Registral General's Department – Head Office

AGLA Tower, Government Campus Plaza

Corner of London & Richmond Streets

Port of Spain

Trinidad and Tobago

Phone: (868) 223- 2452

Web: <https://agla.gov.tt/registrar-general/registrar-general-civil-registry/>

7. Formal labor market participation records were obtained from the National Insurance Board of Trinidad and Tobago (NIBTT). This database included the active contributors registered in the system by May 2017. The dataset was accessed under confidentiality agreements within the premises of the NIBTT. Authorization to access is free of charge, but the process may take around 6 months. It can be obtained by submitting a research proposal to:

The National Insurance Board of Trinidad and Tobago

14-19 Queen's Park East

Port of Spain

Trinidad and Tobago

Phone: (868) 625-2171

Email: [customercare@NIBTT.net](mailto:customercare@NIBTT.net)

Web: [www.nibtt.net](http://www.nibtt.net)

## Data Sources

All the raw databases should be stored in folder 'IN'. This contains the following subfolders:

1. SEA: Includes the Secondary Entrance Assessment (SEA) databases. These exams are taken once a year, and the databases contain information about demographics, test scores, school choices and school assignment.
2. NCSE: Includes the National Certificate of Secondary Education (NCSE) databases. These exams are taken once a year, and the databases contain information about demographics, and test scores.
3. CSEC: Includes the Caribbean Secondary Education Certification (CSEC) databases. There are 2 additional subfolders, 'January' and 'June' since the exam is taken in both months and often

databases are separated for each of them. These exams contain information about demographics, subjects taken, grades per subject and school where the exam was taken.

4. CAPE: Includes the Caribbean Advanced Proficiency Examination (CAPE) databases. There are 2 additional subfolders, 'January' and 'June' since the exam is taken in both months and often databases are separated for each of them. These exams contain information about demographics, subjects taken, grades per subject and school where the exam was taken.
5. arrests: Includes the arrests databases. There are 2 databases, 'KO\_DATA\_all' for information for all years and 'KO DATA undp 2012 to 2017' for information starting 2012. This second one will be used to complement the first one in case there are some missing observations in it. These databases contain information about arrests made, demographics and date of arrest.
6. births: Includes the birth records. There are 4 databases that correspond to records from years 1980-1985, 1986-1990, 1991-2000 and 2001-2016.
7. NIBTT: Includes the National Insurance Board of Trinidad and Tobago (NIBTT) databases for registrations up until 2017.
8. additional\_databases: folder that has various databases such as school geolocation, school information, school name correspondances (when there are different versions of one school's name), default schools used for each community code, etc. These datasets are included in the replication package. Its construction was based on direct information given by the MOE, information based on the SEA, CSEC, CAPE databases or geolocations based on google map searches.
  - Schools3.dta: Contains different versions of each school name as they appear in the different databases. The objective is to identify the same school across databases.
  - Schools3\_single.dta: Identifies each school with one code and one name.
  - primary\_addresses.dta, secondary\_addresses.dta: Contains latitude and longitude for each school.
  - Secondary\_2013\_schoolcodes.dta: Contains information about each secondary school (single sex/mixed, private/assisted/public and region)
  - default\_schools\_used.dta: Default schools used per region
  - comm\_codes\_school.dta, comm\_codes\_default\_school\_wide.dta: Default schools used per community code.

Data file	Source	Notes	Provided
SEA datasets	MOE - TT	Confidential	No
NCSE datasets	MOE - TT	Confidential	No
CSEC datasets	CXC	Confidential	No
CAPE datasets	CXC	Confidential	No
Arrest datasets	TT Police Service	Available also by submitting a research proposal	Yes
Birth datasets	TT Register General	Confidential	No

NIB datasets	NIBTT	Confidential	No
School datasets	various		Yes

---

## Contents

As specified above, due to confidentiality agreements, the only raw dataset that we can post are the Police Arrets Records and some additional databases. This databases are included in the sub-folder “IN/arrests” and “IN/additional\_databases”.

The sub-folder “CODE” contains the programming files to replicate all Main and Appendix Figures and Tables of the paper. These include 60 dofiles and 1 R file in the main “CODE” folder, 3 dofiles and 12 matlab files in the subfolder “CODE/choice” and 4 dofiles in the subfolder “CODE/additional”. All the dofiles will be executed directly by running the following master file:

**0\_master\_file.do:** This program processes all raw files, merges them, and produces the final datasets and results used in the analyses. The final databases will be stored in the sub-folder “OUT”. Running the master file will run all code from start to finish, except for 2 instances: (i) running the R code ‘CODE/31\_size\_impacts\_correlations.R’ to produce table B3 and B5, and (ii) running the Matlab code ‘CODE/choice/0\_master\_matlab.m’ to run the choice model and its corresponding results. The master file will have breaks when this is needed. The general structure of the master file is:

PART A. Clean databases: Cleans and appends all years of each database.

PART B. Merge databases: Merges SEA database with NCSE, CSEC, CAPE, arrests, births and NIBTT.

PART C. Simulation of assigned school. Reproduces the school assignment algorithm based on students choices and SEA scores.

PART D. Final Database for Value Added estimation: Produces the final database used in value-added model.

PART E. Value-Added Estimation: Estimates the value-added model for the main test and the different robustness results.

PART F. Final database for choice model: Produces the final database used in the choice model considering schools that have an estimated value-added.

PART G. Choice Model: Estimates the choice model and all its tables and figures.

PART H. Other Tables and Figures: Estimates the remaining appendix tables and figures.

A detailed description of each dofile is included in the master file.

## Computational requirements

All the computer code used to generate the results in the article are included and instructions are in detail in the master file '0\_master\_file'. The versions of the programs used are as follow:

### *Stata programs:*

Reghdfe: version 5.9.0 03jun2020  
Strdist: version 1.2 09dec2017  
Ranktest: version 1.4.01 18aug2015  
Ftools: version 2.39.1 03jun2020  
Ivregdhfe: version 1.0.3 22jun2020  
Group\_id: version 1.0 15aug2009  
Rddensity: version 2.3 2021-02-28  
Rdrobust: version 7.6.0 09-25-2018  
Mdesc: version 2.1 25Aug2011  
Estout: version 3.24 30apr2021  
Sphdist: version 1.1.0 June 19, 2000

### *R programs*

Catchr: version 0.2.31  
Lmeresampler: version 0.2.0

## Memory and runtime requirements

The code for cleaning and computing the value added was last run on Stata 16 MP on a Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz 2.60 GHz, 16.0 GB RAM with a 64 bit Windows 10 OS. The code for computing the choice models was last run on Matlab R2016b on a iMac(R) with processor 3.6 GHz 10-Core Intel Core i9, 32 GB RAM with a macOS Monterey Version 12.0.1.

The overall runtime of the code is around 160 hours. Cleaning and merging the databases took 50 hours, computing the value-added and doing the required tests took 50 hours and estimating the choice model took 60 hours.

## Instructions to replicators:

- Edit 'CODE/0\_master\_file.do' to adjust the default path (line 25)
- Download the data files referenced above. Each should be stored in the prepared subdirectories of 'IN/'. Confidential data should be requested per instructions above and also stored in the corresponding folders.
- Run CODE/0\_master\_file.do to run all steps in sequence until first break (line 410). Continue with analysis in R on '36\_size\_impacts\_correlations.R' and then with 'CODE/0\_master\_file.do' until next break (line 492). Run matlab code 'CODE/choice/0\_master\_matlab' and finally continue with the last part of 'CODE/0\_master\_file.do'.

## List of tables and programs

Figure/Table	# Program	Line Number	Output File	Note
Table 1	32_data.do	259	direct output - tex file included	
Table 2	36_size_impacts_correlations.R	241,253	direct output - tex file included	It is a subset of table B3
Table 3	3_creation_results.do	943	inference_tests.tex	
Figure 1	33_figure1.do	69	figure1_combined.gph	Produced with random data
Figure 2	3_creation_results.do	554	academout_va_fs_probflexible.gph academout_vavg_fs_probflexible.gph	
Figure 3	3_creation_results.do	571	nonacademout_va_fs_probflexible.gph nonacademout_vavg_fs_probflexible.gph	
Figure 4	3_creation_results.do	745	comparison_va_probflexible.gph	
Table A1	35_robustness_appendix_A.do	432	direct output - tex file included	
Table A2	35_robustness_appendix_A.do	114	direct output - tex file included	
Figure A1	34_likelihood.do	21	discontinuity.gph	
Figure A2	35_robustness_appendix_A.do	450	figureA1_5	
Table B1	38_spread_cutoffs.do	73	direct output - tex file included	
Table B2	23_outcomes.do	523	direct output - tex file included	
Table B3	36_size_impacts_correlations.R	241,253	direct output - tex file included	
Table B4	37_corr_bootstrap.do	239	direct output - tex file included	
Table B5	36_size_impacts_correlations.R	272	direct output - tex file included	
Figure B1	38_spread_cutoffs.do	54	cutoff_spread.gph	
Figure B2	39_poly_specification.do	113	poly.gph	
Figure B3	40_graph_va_vs_hs.do	93	corr.gph	
Figure B4	41_outcome_score_ventiles.do	58	figure_outcomes_score_b.gph	
Table C1	42_n_schools.do	100	direct output - tex file included	
Table D1	43_2sls_others_validation.do	300	direct output - tex file included	
Figure D1	44_extrapolation.do	27	scores_by_selectivity_sim_p10.gph	
Figure D2	45_scatter_weights.do	51	distribution_SEA_score_weights.gph	
Figure D3	43_2sls_others_validation.do	268	figureD3_ci.gph	
Figure E1	46_selectivity.do	128	selectivity_all.gph	
Figure F1	48_real_predicted_validation.do	491	figureF1.gph	
Table G1	49_additivity_and_match_test.do	692	direct output - tex file included	
Figure H1	49_additivity_and_match_test.do	686	figureH1.gph	
Figure I1	50_admin_prob_example.do	35	figureI1.gph	
Figure J1	3_creation_results.do	554	academout_va_rolog.gph academout_vavg_rolog.gph	
Figure J2	3_creation_results.do	571	nonacademout_va_rolog.gph nonacademout_vavg_rolog.gph	
Figure J3	3_creation_results.do	745	comparison_va_rolog.gph	
Figure J4	3_creation_results.do	554	academout_va_noprob.gph academout_vavg_noprob.gph	
Figure J5	3_creation_results.do	571	nonacademout_va_noprob.gph nonacademout_vavg_noprob.gph	
Figure J6	3_creation_results.do	745	comparison_va_noprob.gph	
Figure K1	3_creation_results.do	646	figure_k1.gph	

Figure L1	51_choice_likelihoods.do	152	figureL1.gph
Figure L2	3_creation_results.do	297	figure_l2.gph
Figure M1	3_creation_results.do	368	figure_m1_a.gph figure_m1_b.gph
Table M1	3_creation_results.do	943	inference_tests.tex
Figure N1	3_creation_results.do	554	academout_va_probflexoos.gph academout_vavg_probflexoos.gph
Figure N2	3_creation_results.do	571	nonacademout_va_probflexoos.gph nonacademout_vavg_probflexoos.gph
Figure N3	3_creation_results.do	745	comparison_va_probflexoos.gph

---

## Contact Information

Diether W. Beuermann

[dietherbe@iadb.org](mailto:dietherbe@iadb.org)

C. Kirabo Jackson

[kirabo-jackson@northwestern.edu](mailto:kirabo-jackson@northwestern.edu)

Laia Navarro-Sola

[laia.navarrosola@ies.su.se](mailto:laia.navarrosola@ies.su.se)

Francisco Pardo

[fpardo@utexas.edu](mailto:fpardo@utexas.edu)

## References

Crime, Problem Analysis Branch of the Trinidad, and Tobago Police Service. Trinidad and Tobago's Arrest Records, 1990-2017. 10.5281/zenodo.6456606.

Caribbean Examinations Council (CXC). Caribbean Secondary Education Certification (CSEC), 1993-2016.

Caribbean Examinations Council (CXC). Caribbean Advanced Proficiency Examination (CAPE), 2005-2016.

Ministry of Education. Secondary Entrance Exam (SEA), 1995-2012.

Ministry of Education. National Certificate of Secondary Education (NCSE), 2009-2015.

National Insurance Board of Trinidad and Tobago. Registrations in the National Insurance Board, 1980-2017.

Trinidad and Tobago Registrar General. Trinidad and Tobago's Birth Records, 2010-2016.