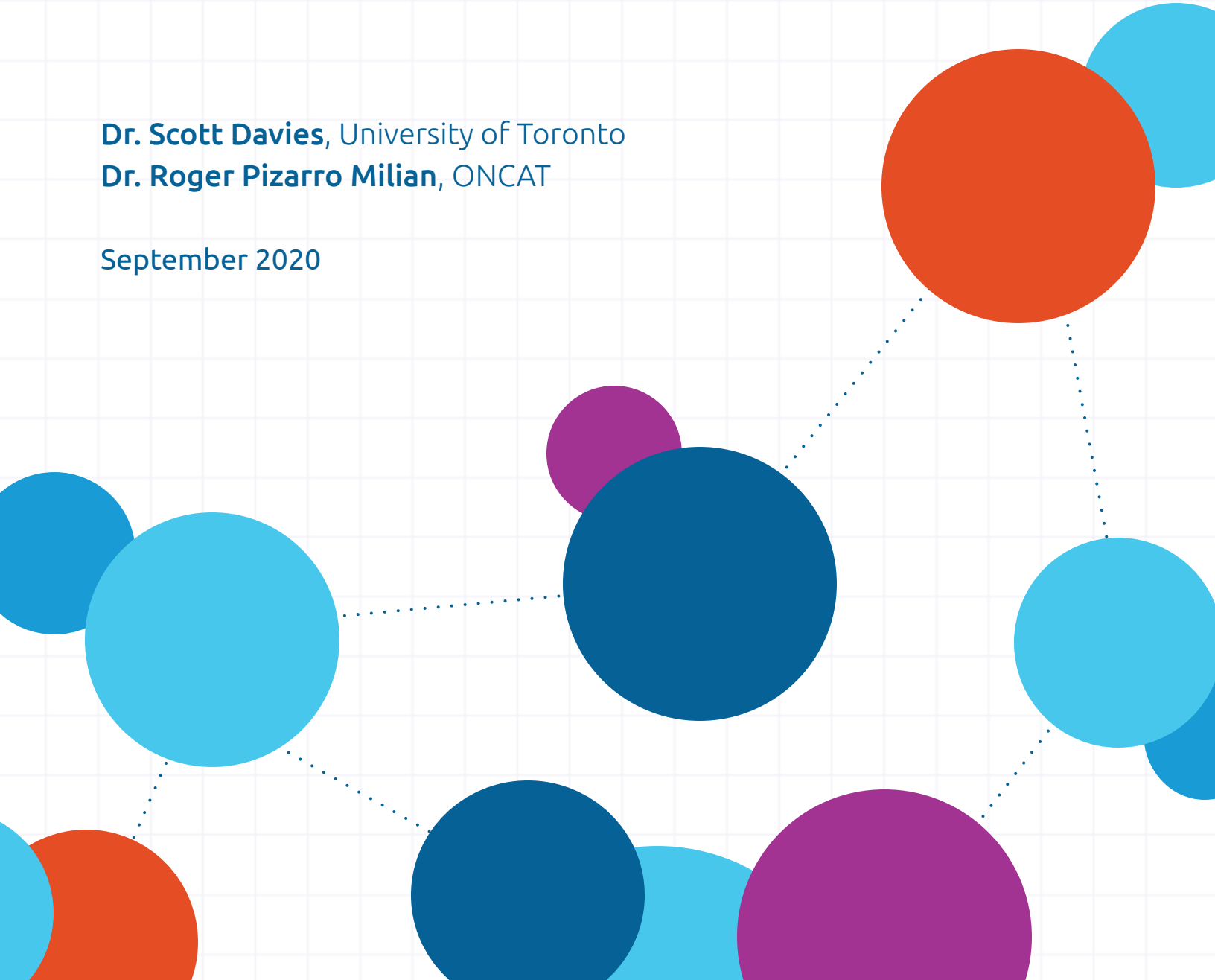
oncat caton

RESEARCH BRIEF

A Look at the TDSB-UT Linkage and Transfer Project

Dr. Scott Davies, University of Toronto
Dr. Roger Pizarro Milian, ONCAT

September 2020



In our last brief, we discussed the significant data requirements for the study of transfer and student mobility research more broadly. In particular, we emphasized the utility of longitudinal datasets that capture early academic performance and socio-demographic information, along with details on later-stage student trajectories (e.g., transfer) and outcomes (e.g., access to STEM, GPA, and graduation rates). These types of datasets constitute the 'holy grail' in educational research.

Today, we provide an overview of an exciting data source: the Toronto District School Board-University of Toronto (TDSB-UT) administrative linkage, which checks off many (but not all) of the required boxes that we laid out before. The TDSB-UT linkage is part of a wave of innovative linkages that are being performed in the field of education across many jurisdictions. As we will discuss with more detail later, this data source has several advantages over the sources typically used to study transfer in our province. Some notable qualities include:

Population-Level Coverage: The linkage captures the **entire** population of students—32,302 in all—that traveled one of Canada's most trafficked and demographically diverse K–12 to PSE corridors over almost two decades. Such sample size allows us to segment our analysis in a number of useful ways, including according to transfer types and student demographics.

"To our knowledge, this constitutes the first direct linkage of its kind between a K–12 school board and university administrative records in the province of Ontario."

Extensive Longitudinal Window: The broad time frame (2000–2019) that this dataset covers allows us to follow the trajectories of non-traditional PSE students, which are often missed by cross-sectional designs, including: those who drop out of high school, but complete later on as mature students; those who take gap years to work between high school and university; and those forced to take lengthy breaks during university due to unforeseen circumstances.

Comprehensive Range of Controls: The dataset contains an extensive set of variables reflecting academic characteristics, covering Grade 9 to university

completion (e.g., grades, field of study, etc.), along with high school behavioural measures (e.g., absenteeism, suspension, etc.), and diverse demographics captured via the TDSB's Student Census. This allows us to estimate the predictors and outcomes of transfer accurately later in the life-course.

Mobility Tracking: Most importantly for ONCAT, the linkage captures (via a previous institution field) every TDSB student who, during this period, entered **any** college or university prior to enrolling at UT.

To our knowledge, this constitutes the first direct linkage of its kind between a K–12 school board and university administrative records in the province of Ontario. Previous linkages have successfully merged administrative records from universities to Ontario University Application Center (OUAC) and Ministry of Education records (see Dooley, Pain, & Robb, 2012). However, our school board data provides more in-depth coverage of both demographics (e.g., sexual orientation, parental education) and other measures (e.g., absenteeism). Our linkage constitutes another step towards closing the gap between Ontario and other jurisdictions where these sorts of linkages are common.

How It's Made

Before we dive into details about this data source and how we are using it for transfer research, let's first go over the behind-the-scenes work that made it possible. In the beginning, there were a series of research ethics board (REB) protocols, legal department reviews, and other institutional sign-offs at both TDSB and UT that made student-level data available to researchers for linkage. In addition, the procedures used to handle data through this project were reviewed by the Ontario Privacy Commission to ensure that they were compliant with privacy laws and regulations. Luckily, the TDSB is an experienced research organization and has developed proven matching and associated procedures through other projects[1] that directly informed this one. We are fortunate that the TDSB has the expertise and willingness to engage in this sort of work, as their records offer an incredible degree of detail on student characteristics.

1. For example, see: Oreopolous & Brown, (2017).

After that lengthy process was complete, administrative records from approximately 200,000 students who started Grade 9 within the TDSB (2000–2010) were cross-referenced to 225,000 student records at UT (2002–2019) using both Ontario Education Numbers (OEN) as well as alphanumeric identifiers for earlier cohorts.[2] This process produced the overall sample of 32,302 perfect matches: students that we have high confidence are the same in both administrative records.

Once the matching process was complete, all unique identifiers (e.g., names, OEN, etc.) were removed from the dataset. This protects the privacy of the individuals and at the same time ensures that researchers have access to granular data. It is consistent with best practices used by other organizations, such as Statistics Canada, to de-identify our personal records for use by researchers.

The resulting dataset allows, as we highlight below, analysts to ask an array of key policy-relevant questions in educational research. For an organization like ONCAT, it allows us to ask: At what rate do TDSB students transfer into UT? What type of pathways do they travel to UT? What high school metrics predict transfer? And, perhaps most importantly: Do transfer and direct-entry student outcomes differ once we statistically account for differences in their high school academics and demographics?

A More In-Depth Look at the Metrics

The TDSB-UT linkage has four main groups of variables that are pertinent to transfer research:

High School Records: This includes academic variables like average grades, EQAO scores in math, English, and science, the results of the first OSSLT exam (pass/fail), and the types of courses taken (e.g., academic, applied, etc.). It also includes what could be considered as "behavioural" metrics, including their number of absences and whether they were ever suspended. Lastly, it includes student status in the TDSB five years after Grade 9 (e.g., completed, dropped out, etc.).

Demographics: This includes a long list of variables captured through

2. The OEN was introduced in Ontario in 2003 but was not fully implemented across Ontario PSE until later in the 2000s. As such, an alphanumeric was developed for students in earlier cohorts utilizing their gender, birthday, and first and last name (for full details, see: Davies, Brown & Chakraborty, 2019, appendix 3).

the TDSB's detailed student census, including gender, sexual orientation, self-identified race, world region of birth, language spoken at home, parental level of education, and occupational category. Many of these variables are not present in the administrative records of colleges or universities, limiting the sorts of transfer student analyses that previous researchers have conducted.

Previous Institution: At UT, the administrative records contain a text field with the name of the student's most recent institution prior to enrolling at the university. We manually coded this field into a series of dummy variables representing particular types of PSE institutions, including Ontario colleges and universities, other Canadian colleges and universities, and international institutions. These metrics provide more detail than the traditional Credit Transfer flag, which only identifies those transfer students as originating from another Ontario PSE institution.

UT Outcomes: Tapping into the UT administrative records, we can also observe if a student graduated during our long window, whether they gained access to a STEM field of study, their cumulative grade point average (GPA), the number of credits they accumulated, and their time to completion.

Combined, this set of data fields present a wide range of opportunities to model statistically both the predictors and outcomes of transfer students compared to their direct-entry counterparts. It also allows us to analyze trends across transfer student types. We can do this type of analysis with greater precision than studies that rely on PSE-level administrative records only, given that we can control for more early predictors of transfer.

Limitations

The focus on a single school board to university corridor obviously makes these findings difficult to generalize or apply to other corridors across the province, particularly in rural and northern institutions. There is no large provincial picture to be drawn here, despite this being the most trafficked corridor in Ontario. Furthermore, this dataset only captures those TDSB students that eventually transfer into UT since we cannot track outgoing transfers. We cannot say anything about which students are most likely to leave UT for other PSE institutions.

As with nearly every dataset, there are also some notable patterns in missing data. Most TDSB academic fields are nearly 100% complete (with the exception of absenteeism,

which has 30% missing data), and most UT student records are at least 95% complete. However, some demographic variables have sizeable rates of missing data. In particular, the TDSB student census has only 67% coverage because it was not widely distributed among early cohorts in the dataset (pre-2004). As a result, some measures, like parents' occupation, had sizeable rates of missing data, leaving only 50% valid cases. What this effectively means is that researchers using it must be mindful of the potential bias introduced by dropping these cases from their analysis.

Next Up

Next week, we will begin to publish the results of some exploratory transfer analyses we conducted with the TDSB-UT linkage. This includes both descriptive information, as well as the results of logistic regression models. Our goal through these types of briefs is to present findings in a way that is digestible by stakeholders who—like us—are not trained as statisticians. In light of such a goal, we do away with the convention of relying on readers to interpret vast regression tables with coefficients, standard errors, p-values, and other statistical details.[3] Instead, we present as much of what we can in plain language, visuals, and a few simple tables.

As always, we welcome further feedback or questions about this work.

3. The full regression model output is available upon request from Roger Pizarro Milian (rpizarromilian@oncat.ca).

Dr. Scott Davies

Dr. Scott Davies is Professor and the Canada Research Chair in Data, Equity and Policy in Education (Tier 1) at the Department of Leadership, Higher and Adult Education at the University of Toronto.

Dr. Roger Pizarro Milian

Dr. Roger Pizarro Milian is the Senior Researcher at the Ontario Council on Articulation and Transfer (ONCAT).

oncat caton

oncat.ca/en/projects/tdsb-ut-linkage-and-transfer-project

Established in 2011, the Ontario Council on Articulation and Transfer (ONCAT) was created to enhance academic pathways and reduce barriers for students looking to transfer among Ontario's public colleges, universities, and Indigenous Institutes.

The authors acknowledge the financial support provided by the Ontario Council on Articulation and Transfer (ONCAT), funded by the Ontario Ministry of Colleges and Universities, to develop this publication.
The views expressed in this publication are the views of the authors and do not necessarily reflect those of the Government of Ontario. © Copyright, ONCAT, 2020