



INEQUALITIES IN TEST SCORES BETWEEN INDIGENOUS AND NON-INDIGENOUS YOUTH IN CANADA

Michael Barber

Department of Economics, Queen's University
Kingston, ON Canada K7L 3N6

&

Maggie E. C. Jones

Department of Economics, University of Victoria
Victoria, B.C., Canada V8W 2Y2

July, 2019

Abstract

This paper documents a robust achievement gap between the math scores of Indigenous and non-Indigenous youth in Canada between 1996 and 2008. Using data from the restricted-access National Longitudinal Survey of Children and Youth we show that after controlling for a rich set of observables, students who self-identify as Indigenous perform 0.31 standard deviations lower on a standardized math test compared to their non-Indigenous counterparts. We find that this test gap emerges by the age of 12, and it did not decline between 1996 and 2008, despite the recommendations of the 1996 Royal Commission on Aboriginal Peoples to ameliorate the public education system for Indigenous students. Counterfactual estimates from the decomposition method of Lemieux (2002) suggest that the test gap among the lowest performing students would have been eliminated if Indigenous students faced the same level of and returns to observable characteristics as non-Indigenous students. This exercise does not result in a narrowing of the test gap in the upper tail, suggesting that unobservables, rather than observables, are driving the majority of the test gap among high achieving students.

Keywords: test gap, Indigenous peoples, decomposition methods

JEL Classifications: I21, I24, J15

Author Contact:

Maggie Jones, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; E-mail: maggie.ec.jones@gmail.com.

INEQUALITIES IN TEST SCORES BETWEEN INDIGENOUS AND NON-INDIGENOUS YOUTH IN CANADA*

Michael Barber[†]
Queen's University

Maggie E.C. Jones[‡]
University of Victoria

July 23, 2019

Abstract

This paper documents a robust achievement gap between the math scores of Indigenous and non-Indigenous youth in Canada between 1996 and 2008. Using data from the restricted-access National Longitudinal Survey of Children and Youth we show that after controlling for a rich set of observables, students who self-identify as Indigenous perform 0.31 standard deviations lower on a standardized math test compared to their non-Indigenous counterparts. We find that this test gap emerges by the age of 12, and it did not decline between 1996 and 2008, despite the recommendations of the 1996 Royal Commission on Aboriginal Peoples to ameliorate the public education system for Indigenous students. Counterfactual estimates from the decomposition method of Lemieux (2002) suggest that the test gap among the lowest performing students would have been eliminated if Indigenous students faced the same level of and returns to observable characteristics as non-Indigenous students. This exercise does not result in a narrowing of the test gap in the upper tail, suggesting that unobservables, rather than observables, are driving the majority of the test gap among high achieving students.

JEL Codes: I21, I24, J15,

Keywords: test gap, Indigenous peoples, decomposition methods

*We thank all individuals who provided valuable feedback on this work, including seminar participants at Queen's University and the University of Victoria, as well as conference participants at the 2018 CEA meetings and the 2016 ACEA meetings. This research was supported by funds to the Canadian Research Data Centre Network (CRDCN) from the Social Science and Humanities research Council (SSHRC), the Canadian Institute for Health Research (CIHR), the Canadian Foundation for Innovation (CFI) and Statistics Canada. Although the research and analysis are based on data from Statistics Canada, the opinions expressed do not represent the views of Statistics Canada or the Canadian Research Data Centre Network (CRDCN). All errors are our own.

[†]Department of Economics, Queen's University. E-mail: barberm@econ.queensu.ca

[‡]Corresponding Author: Department of Economics, University of Victoria. E-mail: maggie.ec.jones@gmail.com

Education is an important determinant of labour market outcomes (Becker, 1964; Card, 1999; Rose and Betts, 2004), and given that learning is a cumulative process, understanding the determinants of academic success throughout a student’s school experience is crucial for understanding gaps in socioeconomic outcomes later in life (Gagné, 1968; Todd and Wolpin, 2003; Ding and Lehrer, 2014; Heckman, Pinto, and Savelyev, 2013; Heckman, Gertler, Pinto, Zanolini, Vermeerch, Walker, Chang, and Grantham-McGregor, 2014; Heckman and Conti, 2014). This paper studies the test score gap in math between minority and non-minority groups in Canada to further our understanding of the origins of socioeconomic disparities among traditionally marginalized demographics. We use the restricted-access National Longitudinal Survey of Children and Youth (NLSCY), a nationally representative longitudinal survey of approximately 23,000 Canadian youth between 1996 and 2009,¹ to quantify the extent and determinants of the test gap between white students and black, South Asian, Chinese, and Indigenous students, paying particular attention to the Indigenous-white test gap.²

We focus on the results of a math component administered to students during each cycle of the survey. The NLSCY includes information on a number of factors that have been shown to be important inputs into the education production function (Hanushek, 2008), including household income, parental education, household characteristics, school type, and location. Our estimates suggest that, conditional on these inputs, the Indigenous-white test gap among children aged 8-15 is approximately 0.31 standard deviations.³ This figure is remarkably similar to the Indigenous-white test gap estimated for grade 7 students in Friesen and Krauth (2010) using administrative data from the province of British Columbia. By splitting the sample into age cohorts, we show that the majority of this test-gap emerges after students begin middle school and splitting into survey wave cohorts reveals that the test gap changed very little between 1996 and 2008. That

¹The NLSCY began in 1994; however, due to issues with the math component of the survey that we discuss in our empirical section, we focus on the years 1996 onwards.

²In both Canada and the United States, Indigenous peoples have lower educational outcomes compared to their non-Indigenous counterparts, yet there has been little focus among economists on understanding the origins of these socioeconomic disparities.

³The NLSCY is not administered on Indian reserves, where education falls under federal jurisdiction, as opposed to off-reserve schools, which fall under provincial jurisdiction. If quality differs between on- and off-reserve schools, our estimates of the test score gap will either over- or under-represent the gap depending on the differences in school quality. Despite the lack of an on-reserve sample, our results are still important for understanding socioeconomic disparities faced by Indigenous peoples in Canada, as recent statistics suggest that only one in five Indigenous students now attend schools on-reserve.

being said, we acknowledge that comparisons over time are sensitive to scaling choices and should therefore be interpreted with caution (Bond and Lang, 2013).

The marked worsening of the Indigenous-white test gap at around the same time students begin junior high school suggest that students at this age face additional adversities that merit consideration. For instance, Friesen and Krauth (2010) show that having a large number of Indigenous peers tends to increase the achievement of Indigenous students, so if many Indigenous students in Canada are located in schools where their peers are predominantly white, they may face discrimination, bullying, or racism, which could in part explain the timing of the test gap.⁴ Unfortunately, due to the data limitations of the nationally representative sample, we cannot speak to peer composition to the same extent as Friesen and Krauth (2010); however, the fact that the test gap we estimate is similar in magnitude to theirs suggests that their insights may also be applicable to our sample.

To shed light on some of the socioeconomic mechanisms underlying the Indigenous-white test gap, we use the methodology of Lemieux (2002) to decompose differences across the entire distribution of test scores.⁵ A comparison of the Lemieux (2002) framework to the more traditional Oaxaca-Blinder (Blinder, 1973; Oaxaca, 1973) decomposition method reveals that the latter masks important heterogeneity in terms of the factors driving test gaps in the tails of the distributions. We construct counterfactual test score distributions for Indigenous students to show that the test gap in the lower tail is more than eliminated once we account for differences in both the returns to and levels of observables characteristics. We find almost no change in the test gap after computing

⁴Another possible explanation for the decline in test scores when students enter junior high school is that some reserves only have elementary schools and students must travel outside of their home community to attend junior high school or high school. If reserve schools are providing a lower quality of education or if travelling away from one's home adversely affects educational attainment, then this could mechanically decrease the test score gap in the later years of schooling. This is unlikely in our setting because students are selected to participate based on their home community, and reserve communities were not included in the survey.

⁵Decomposition methods have been used to study differences in Indigenous-non-Indigenous outcomes in Canada. For instance, George and Kuhn (1994) use an Oaxaca-Blinder decomposition to study earnings gaps and find that differences in observable characteristics can explain up to 50 percent of the wage gap. More recently, Feir (2013) uses an Oaxaca-Blinder decomposition to understand whether the reasons behind earnings differentials between Métis and North American Indians living off-reserves and non-Indigenous Canadians, have changed over time, as well as the extent of the on-reserve earnings penalty. The advantage of the Lemieux (2002) framework over other decomposition methods typical in the labour literature, like Juhn, Murphy, and Pierce (1993), is that it directly accounts for differences in the distribution of covariates. A limitation of the Lemieux (2002) methodology is that accounting for differences in the distribution of covariates must be done using a logit or probit model when there is a large number of covariates, as in our case. These estimates are not consistent when the error terms are heteroskedastic. An alternative methodology was developed in Melly (2005), which we do not consider in this analysis.

our counterfactual distributions for those in the upper tail of the distribution, suggesting that unobservables, rather than observables, are driving the test gap among high achieving students.

This study provides the first nationally representative estimates of the off-reserve Indigenous-white test gap in Canada and focusses the analysis on a population that has traditionally been excluded from the test-gap literature in North America more generally.⁶ Exceptions include the work of Clotfelter et al. (2009) whose study includes an analysis of achievement gaps between American Indians and whites in North Carolina and the Canadian studies of Richards, Howe, and Afolabi (2008) , Friesen and Krauth (2010), and Richards, Vining, and Weimer (2010) that focus on Indigenous-white achievement gaps in the province of British Columbia. Given that the political economy of Indigenous groups in British Columbia and North Carolina may be different than elsewhere—e.g., British Columbia had relatively fewer historical treaties, has a comprehensive process of modern treaty-making, and more representation of Indigenous material in provincial curricula, among other differences—the insights from existing studies may not extrapolate to other Indigenous groups. We also extend our analysis to include other Canadian minority groups, where existing Canadian studies have focussed only on the test gap between Indigenous and non-Indigenous students (Richards, Howe, and Afolabi, 2008; Friesen and Krauth, 2010; Richards, Vining, and Weimer, 2010).

In concordance with the literature focussed on the black-white test gap in the United States (Fryer and Levitt, 2004, 2006; Clotfelter et al., 2009; Sohn, 2012; Bond and Lang, 2013, 2018), we also find a sizeable difference in test scores between black and white students. As a comparison, our estimates of the black-white test gap in Canada are about half the magnitude of those inferred by Fryer and Levitt (2004) from extrapolating their analysis of kindergarten and grade one students to later grades. While Chinese and South Asian students in our sample systematically outperform whites, we do not find any discernible trends in test scores across ages or time for minority groups other than Indigenous students.

⁶We are aware of at least two studies that have quantified the Indigenous-white test gap outside of North America—in Australia (Leigh and Gong, 2009), and Peru (Sakellariou, 2008).

2 Data and Background

Our empirical analysis uses the restricted-access National Longitudinal Survey of Children and Youth (NLSCY). The NLSCY is a nationally representative longitudinal dataset that was first administered to students in 1994-1995, and then students were re-interviewed every two years ending in 2008-2009. In addition to the original students included in the longitudinal component, new students were added to each wave of the survey to generate a nationally representative cross-sectional component. The youngest students in the NLSCY are under the age of 1 and the oldest are up to 25 years old.

Questions in the survey cover a wide range of topics including health, education, physical and cognitive development, and physical and social environments. In addition to interviewing students themselves, in some years the surveys also include a parent component, school and teacher component, and a module for the interviewer to fill out that addresses neighbourhood characteristics.⁷ The NLSCY was administered in all 10 provinces, but excludes students living in any of the three territories, on Indian reserves or Crown lands, residents of institutions, full-time members of the Canadian Armed Forces, and some remote regions.⁸ Since schools on Indian reserves fall under federal jurisdiction, while those elsewhere fall under provincial jurisdiction, the test gap we estimate in this paper may over or underestimate the full extent of the test gap in Canada.⁹ The lack of comparable data for Indigenous peoples on reserves is a pervasive problem occurring in almost all socioeconomic datasets in Canada.

The NLSCY contains a variety of questions regarding ethnic background and ancestry. In particular, in the 2000-2001 wave, a question was added to the survey, which explicitly asked the racial identity of the individual.¹⁰ Since the identity-based questions were only included after 2000,

⁷Each survey wave contains the student component, but only a small number of waves contain the school, teacher, and neighbourhood components.

⁸Remote areas excluded from the study are those with fewer than 10 households or 25 persons, or enumeration areas with fewer than 25 households.

⁹For example, despite increases in high school graduation and achievement of students on reserves, reserve schools still struggle with underfunding, geographic isolation, and additional social problems compared to off-reserve schools (Barsh, 1994; Richards, 2008; Anderson and Richards, 2016). Taken together, these factors would suggest that our estimates underestimate the test gap.

¹⁰Originally this question asked, “*How would you best describe your race or colour?*”, where students could respond with “Inuit/Eskimo”, “Métis”, or “North American Indian”, which are the three Indigenous groups in Canada. Students could also respond with “Black”, “Chinese”, “South Asian”, as well as a set of other ethnicities, which we combine in a category called “Other”. By 2004-2005, another question was added that read, “*Are you an Aboriginal person, that is, North American Indian, Métis or Inuit?*”. No analogous question was added for students of other

we exploit the longitudinal nature of the data and classify students' ethnicity based on whether they indicate that they identify as an ethnic minority in any of the later survey waves.^{11,12}

The test score used in this analysis is a mathematics test that is administered to all students in the NLSCY in grade 2 or above. The test was a shortened version of the Mathematics Computation Test from the standardized Canadian Achievement Test. It was designed to test a student's understanding of the operations of addition, subtraction, multiplication, and division. The test was 10 questions long for students in grade 2 and 3, and 15 questions long for higher grades. The NLSCY provides the raw score of correct answers, as well as a standard score. Standard scores were developed based on a sample of Canadian children selected by the Canadian Testing Centre across all ten provinces, referred to as the norm sample. Since we examine both how the test gap has evolved over age and over time, we pool all waves of the NLSCY and standardize the standard scores to have a mean of 0 and variance of 1 for each age.¹³ We restrict our sample in several ways to circumvent issues that arise from the nature of the data and questionnaire. First, we only consider students who would have been within the age range of the math test, that is, between 7 and 15. Of these students, we drop those who are 7 because of a low sample size, and we drop the first cycle (1994-1995) because a number of students in this wave received perfect scores on the math test, leading to low variation in this subsample.

Perhaps the most important sample selection issue for our analysis arises because the math test is only administered to students if parents consent to its administration. This resulted in a surprisingly large number of students without any math score at all, and since the math score is our outcome of interest we must drop students who do not have this score.¹⁴

ethnicities.

¹¹One should note that willingness to identify as Indigenous has changed over time in Canada (Guimond, 1999, 2009; Caron-Malenfant, Coulombe, Guimond, Grondin, and Lebel, 2014). This would be a problem in our analysis if unobserved characteristics correlated with willingness to identify as Indigenous were also correlated with test scores. Since we classify students as having Indigenous identity based on whether they identify as Indigenous in any year of the survey, we do not expect this to affect our results.

¹²An alternative measure of ethnicity could be to classify students based on their reported ancestry. Each survey wave included a question that read, "*To which ethnic or cultural group(s) did your ancestors belong?*" We have also used this question to classify students as having ancestry belonging to an ethnic minority. The question relating to Indigenous ancestry was repeated each year. In a small number of cases some students reported having Indigenous ancestry in one year and not in a subsequent year. To deal with these cases, we recoded all students as having Indigenous ancestry in all years if they indicate ancestry in one year. Results using this classification are qualitatively similar to those presented in the empirical section, but smaller in magnitude.

¹³The main analysis is invariant to using the original standardized scores included in the raw data and to standardizing by age-cycle. These results are unreported, but available upon request.

¹⁴In the empirical section, we use a Heckman correction model (Heckman, 1976, 1979) to evaluate the extent to

Finally, while the NLSCY does not directly report household income, they do include a related measure called the low-income cut-off (LICO) ratio. This variable is a ratio of the household's income to the LICO for their economic family.¹⁵ Unfortunately, the LICO ratio is missing for a small number of households. Since income has high explanatory power for test scores, and is also highly correlated with ethnicity, we posit that the bias that would result from excluding income is larger than the bias induced by excluding those who do not report income. Table 10 of the Online Appendix confirms this proposition, so we restrict our sample to those with a LICO ratio for our empirical analysis.

Table 1 presents a set of descriptive statistics for students who identify as Indigenous and those who do not. The average standardized math score was 0.677 standard deviations lower for students who identify as Indigenous compared to those who do not. It is also apparent that the characteristics of Indigenous students differ substantially from non-Indigenous students.¹⁶

On average, the sample of Indigenous students is more likely to be male and slightly younger. Indigenous students tend to have parents with lower levels of education. For instance, the probability that an Indigenous student has a primary care giver (PCG) with a high school degree or a post-secondary degree is nearly 11 and 15 percentage points lower, respectively, than for students who do not identify as Indigenous. The likelihood that the spouse of the primary care giver is educated is also lower for students who identify as Indigenous. The portion of Indigenous students born to teenage mothers and who live in single parent households is higher; however, the share of Indigenous students born at a low birth weight is lower. Importantly, the LICO ratio is 0.670 points lower for students who identify as Indigenous. Compared to a mean LICO ratio of 2.265 for the non-Indigenous population, this difference is both large in magnitude and statistically significant. Not surprisingly, a higher share of Indigenous students report having Indigenous ancestry, and a lower share also identify as white. It should be noted that a small portion (3.8 percent) of

which this selection issue may bias our results.

¹⁵The LICO is essentially a stand-in for the concept of a poverty line in Canada. It identifies a family as low income based on the portion of their income spent on essentials, signifying a greater likelihood of being economically strained. The LICO is determined based on the size of the city in which the individual resides, as well as their family size. Its construction therefore reflects consumption patterns that may differ across city and family sizes (Murphy et al., 2012).

¹⁶In the empirical specifications, we also control for whether the individual is an immigrant, or if they identify as black, South Asian, or Chinese identity; however, the portion of the sample reporting these ethnicities or who are immigrants is so small that we are unable to display summary statistics due to Statistics Canada reporting rules.

Table 1: Descriptive Statistics by Indigenous Identity

	Indigenous	Non-Indigenous	Difference
Math Score	-0.619 (0.057)	0.058 (0.009)	-0.677*** (0.058)
Male	0.559 (0.031)	0.488 (0.004)	0.071** (0.031)
Child's Age	11.513 (0.151)	11.833 (0.017)	-0.321*** (0.152)
PCG High School	0.747 (0.024)	0.855 (0.003)	-0.108*** (0.025)
PCG Post Secondary	0.287 (0.026)	0.432 (0.004)	-0.145*** (0.027)
Spouse High School	0.371 (0.032)	0.666 (0.004)	-0.295*** (0.032)
Spouse Post Secondary	0.213 (0.028)	0.364 (0.004)	-0.151*** (0.029)
Low Birth Weight	0.028 (0.008)	0.060 (0.002)	-0.031*** (0.008)
Teenage Mother	0.065 (0.010)	0.033 (0.001)	0.032*** (0.010)
Number of Siblings	1.502 (0.072)	1.389 (0.008)	0.113 (0.072)
LICO Ratio	1.596 (0.047)	2.265 (0.013)	-0.670*** (0.049)
Public School	0.645 (0.031)	0.604 (0.004)	0.041 (0.031)
Catholic School	0.123 (0.017)	0.137 (0.003)	-0.014 (0.017)
Other School	0.036 (0.010)	0.008 (0.001)	0.029*** (0.010)
Missing School	0.178 (0.027)	0.202 (0.003)	-0.025 (0.027)
Single Parent	0.442 (0.033)	0.193 (0.003)	0.249*** (0.033)
Identifies Black	0.009 (0.002)	0.011 (0.001)	-0.002 (0.002)
Identifies White	0.415 (0.031)	0.602 (0.004)	-0.187*** (0.032)
Indigenous Ancestry	0.717 (0.034)	0.038 (0.001)	0.680*** (0.034)

Notes: All descriptive statistics are calculated using the survey weights provided in the NLSCY. The last column presents difference-in-means tests where * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

students who do not self-identify as Indigenous also report having Indigenous ancestry.

3 Empirical Framework and Results

3.1 Average Effects

3.1.1 OLS Framework and Results

Our first empirical specification attempts to quantify the overall test gap in math scores between Indigenous and non-Indigenous youth. We assume that the standardized math score Y of individual i from region r of age a sampled in survey s , is a function of observable and unobservable inputs:

$$Y_{i,r,a,s} = \alpha + \sum_k \beta_k \mathbb{1}(\text{ethnicity}_i = k) + \mathbf{X}_{i,a,s} \boldsymbol{\theta} + \delta_a + \gamma_s + \zeta_r + \epsilon_{i,r,a,s}, \quad (1)$$

The set of indicators $\mathbb{1}(\text{ethnicity}_i = k)$, $k \in \{\text{Indigenous, black, Chinese, South Asian, other}\}$, equals 1 if individual i is of the corresponding ethnicity. The omitted ethnicity category is “white”, so that each coefficient estimate $\hat{\beta}_k$ measures the conditional difference in test scores between group k and white students. We include a matrix of controls in $\mathbf{X}_{i,a,s}$.¹⁷ We include a set of dummy variables for student age, δ_a , and for the survey wave in which the respondent was tested, γ_s . Finally, to account for unobservable differences across regions, all of our specifications include census metropolitan area (CMA) interacted with province dummies, ζ_r . The reason for this interaction is that the CMA variable includes a unique identifier for each CMA in Canada, but those living outside of CMAs are coded according to four degrees of rurality. Thus, someone living in a rural area of British Columbia would be coded as having the same fixed effect as someone living in a rural area of Newfoundland. For this reason, we interact CMA with province, which more accurately assigns clusters to rural regions within provinces, but does not change the clusters of those living in a CMA. Standard errors are clustered at the provincial level, as this is the jurisdictional level at which education policy is determined. Since only 10 provinces are included in the NLSCY, inference based on the assumption that the number of clusters is large will be invalid. We therefore also report significance levels on our identity variables that are constructed using the wild cluster bootstrap in the `boottest` Stata package of Nielsen, MacKinnon, Roodman, and Webb

¹⁷Specifically, we include indicators for whether the child is an immigrant, their gender, if they come from a single parent family, if they were born low birth weight, if they were born to a teenage mother, if they attend a public, private, catholic, or other type of school, education levels of their primary care giver and the primary care giver’s spouse, and we include continuous variables that control for the LICO ratio, and the number of siblings.

(2019) with 999 replications. All regressions are weighted by the survey weights included in the NLSCY. Due to the reporting restrictions associated with the NLSCY, observations in regression tables have been rounded to the nearest 100.

The results from estimating equation 1 can be found in Table 2.¹⁸ Each column adds a new covariate (or group of covariates) to see how the coefficient on each of the minority indicators varies with the addition of new controls. In our most restrictive specification, column (6), students who identify as Indigenous have test scores that are up to 0.310 standard deviations lower compared to white students. This finding is not dissimilar from Clotfelter et al. (2009) who estimate the math test gap between Native American and white students in the state of North Carolina to be between -0.237 (grade 8) and -0.299 (grade 3). It is also similar to the findings in Leigh and Gong (2009) who show that the Indigenous-non-Indigenous test gap in Australia is approximately 0.3-0.4 standard deviations for five year olds.¹⁹

Comparing the movement in the coefficient estimate on the Indigenous indicator reveals that the largest movements arise when we add controls that proxy for income (the LICO ratio) and parental education (primary care giver and spouse). The covariates that do not appear to predict test scores in a statistically meaningful way are the gender of the child, whether the child is an immigrant, and whether the child attends a public or catholic school. In general the coefficient estimates enter the equations with the expected sign: higher income correlates with higher test scores, students whose parents have a higher level of education perform better on the test, and students who are born at a low birth weight perform worse on the test. The coefficient on the gender indicator is negative in each specification, although it is not statistically different from 0. The direction of the gender coefficient is consistent with previous work showing that girls typically outperform boys in school (Bertrand and Pan, 2013; Murnane, 2013; Autor et al., 2016b,a).²⁰ Perhaps somewhat surprising is that once we condition on parental education, we find that students who live in a single parent family have higher test scores. A closer consideration of this finding reveals that there is an intuitive

¹⁸Results using inverse propensity score weighting do not differ substantially and are found in Table 11 of the Online Appendix.

¹⁹Leigh and Gong (2009) use the Peabody Picture Vocabulary Test to examine the test gap, so their results are not directly comparable to results using math tests.

²⁰There are many explanations for why boys are currently outperformed by girls in the classroom. For instance, Autor et al. (2016a) suggest that boys are disproportionately affected by family disadvantage, while Bertrand and Pan (2013) find that boys raised in single-parent households are more likely to display behavioural or disciplinary issues which may contribute to worse performance in school.

Table 2: Difference in Test Scores between Indigenous and Non-Indigenous Youth

	(1)	(2)	(3)	(4)	(5)	(6)
Identifies Indigenous	-0.444*** (0.057)++	-0.361*** (0.046)++	-0.352*** (0.052)+++	-0.308*** (0.058)+++	-0.310*** (0.055)+++	-0.310*** (0.053)+++
Identifies Black	-0.452*** (0.071)	-0.375*** (0.077)	-0.372*** (0.077)	-0.326*** (0.080)	-0.341*** (0.078)	-0.335*** (0.076)
Identifies South Asian	0.112** (0.038)+	0.178*** (0.037)	0.152*** (0.045)+	0.163*** (0.031)+	0.182*** (0.035)+	0.183*** (0.033)+
Identifies Chinese	0.487*** (0.043)++	0.477*** (0.059)+	0.474*** (0.060)+	0.497*** (0.060)+	0.500*** (0.061)+	0.487*** (0.077)++
Identifies Other	-0.0217 (0.041)	0.00919 (0.043)	0.00494 (0.046)	0.0105 (0.045)	0.0146 (0.044)	0.0139 (0.044)
Immigrant	0.278 (0.185)	0.329 (0.196)	0.320 (0.179)	0.280* (0.150)	0.277 (0.152)	0.254 (0.143)
Male	-0.0377 (0.048)	-0.0318 (0.049)	-0.0335 (0.049)	-0.0330 (0.046)	-0.0317 (0.046)	-0.0343 (0.044)
LICO Ratio		0.0921*** (0.008)	0.0932*** (0.009)	0.0637*** (0.002)	0.0619*** (0.003)	0.0577*** (0.003)
Single Parent			-0.0362* (0.019)	0.103*** (0.014)	0.103*** (0.014)	0.0994*** (0.016)
Number of Siblings			0.0565*** (0.012)	0.0464*** (0.012)	0.0444*** (0.012)	0.0412** (0.013)
Primary Care High School				0.184*** (0.019)	0.173*** (0.019)	0.174*** (0.019)
Primary Care Post-Secondary				0.132*** (0.038)	0.128*** (0.038)	0.120*** (0.034)
Spouse High School				0.0986*** (0.020)	0.0952*** (0.020)	0.0984*** (0.019)
Spouse Post-Secondary				0.158*** (0.022)	0.156*** (0.022)	0.148*** (0.019)
Low Birth Weight					-0.165*** (0.008)	-0.163*** (0.008)
Teenage Mother					-0.198*** (0.024)	-0.195*** (0.026)
Public School						0.0196 (0.101)
Catholic School						-0.0268 (0.126)
Private School						0.302** (0.123)
Other School						0.284* (0.148)
Constant	0.465*** (0.039)	0.273*** (0.045)	0.204*** (0.060)	-0.0881 (0.065)	-0.0550 (0.065)	-0.0575 (0.149)
Observations	34500	34500	34500	34500	34500	34500
Adjusted- R^2	0.120	0.140	0.143	0.164	0.167	0.171

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. The omitted ethnicity is white, so that all coefficient estimates represent the test score gap relative to white students. Omitted schooling category is whether the school information is missing, and the omitted category for parental education is no formal schooling. All columns include CMA-province, age, and survey wave fixed effects. Standard errors clustered by province are in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Given the small number of clusters, we also report significance levels on all “Identity” coefficients from the wild cluster bootstrap constructed using the `boottest` stata package of Nielsen et al. (2019) with 999 replications: + $p < 0.1$, ++ $p < 0.05$, +++ $p < 0.01$.

explanation. First, single parent families often have lower educational attainment, on average, and since parental education is positively correlated with children’s educational achievement, this should lead to a downward bias in the coefficient estimate for single parent families in column (2). Second, since the variable “Spouse High School” (and similarly “Spouse Post-Secondary”) equals 1 if the primary care giver’s spouse has a high school degree and equals 0 if the primary care giver’s spouse either does not have a high school degree or the primary care giver does not have a spouse, then mechanically this variable will be negatively correlated with the single parent variable. For the same reason as before—namely, that spousal education should be positively correlated with children’s test scores—this should also result in a downward bias in the coefficient estimate of single parent households in column (2).

While the purpose of this paper is to emphasize the determinants of the Indigenous-white test gap, we also find a large gap in test scores for students who identify as black compared to white students. Under the full set of controls, the black-white test gap is 0.335 standard deviations, slightly larger than the Indigenous-white test gap, and in line with studies that have focussed on the test gap in the United States.²¹ We also find that the test scores of Chinese and South Asian students are 0.487 and 0.183 standard deviations higher than white students, respectively. This result is also consistent with Clotfelter et al. (2009) who find that Asian students outperform whites.

To further our understanding of some of the barriers faced by minority groups in Canada, Table 3 reports the results of a subsample analysis, where each column presents estimates of equation 1 for the subsample described in the column header. Column (1) replicates the results for the full sample for comparison. The second two columns focus on the test gap across genders. For the Indigenous-white test gap, there does not appear to be differences across genders. For black students, the test gap is completely driven by male students. The female black-white test gap, while negative, is not statistically different from 0. This result is similar to the findings of Fryer and Levitt (2004) and Fryer and Levitt (2006) in that the black-white test gap in the United States is larger for males than females; however, they still find a sizeable difference between the

²¹Fryer and Levitt (2004) find a regression adjusted black-white test gap of 0.094 standard deviations in the fall of kindergarten, which increases to 0.201 by the spring of first grade. By third grade the test gap increases to 0.382 (Fryer and Levitt, 2006). Clotfelter et al. (2009) find a slightly larger estimate of the black-white test gap, but focus on cohorts in grades 3 through 8.

Table 3: Difference in Test Scores between Indigenous and Non-Indigenous Youth by Subsample

	Full Sample	Gender		LICO		Low Birth Weight		Single Parent		Region		
		Female	Male	Below Med	Above Med	Yes	No	Yes	No	Western	Central	Atlantic
Identifies Indigenous	-0.310*** (0.053)+++	-0.318** (0.104) ⁺	-0.326*** (0.092)+++	-0.379*** (0.036) ⁺⁺	-0.0780 (0.069)	-0.902*** (0.076)	-0.291*** (0.052) ⁺⁺	-0.491*** (0.090) ⁺⁺	-0.140 (0.079) ⁺⁺	-0.389*** (0.047)+++	-0.226* (0.031)+++	0.130 (0.093)
Identifies Black	-0.335*** (0.076)	-0.154 (0.175)	-0.426*** (0.035) ⁺⁺	-0.314** (0.105)	-0.231*** (0.045)	-0.852*** (0.194)	-0.333*** (0.082)	-0.362* (0.194)	-0.346* (0.163)	0.0615 (0.162)	-0.414* (0.033)+++	0.0592 (0.148)
Identifies South Asian	0.183*** (0.033) ⁺	0.242*** (0.016) ⁺⁺	0.107 (0.065)	0.193*** (0.023)	0.314*** (0.071) ⁺	0.414 (0.276)	0.132** (0.057)	0.0413 (0.236)	0.199*** (0.045)	0.195** (0.057)+++	0.180* (0.021)	. .
Identifies Chinese	0.487*** (0.077) ⁺⁺	0.588*** (0.089) ⁺⁺	0.388*** (0.074) ⁺	0.720*** (0.140) ⁺⁺	0.228*** (0.054)	0.358 (0.431)	0.488*** (0.067)	0.432** (0.162)	0.509*** (0.101) ⁺⁺	0.437** (0.080)+++	0.600 (0.131) ⁺⁺	. .
Identifies Other	0.0139 (0.044)	-0.0474 (0.040)	0.0694 (0.065)	0.0266 (0.063)	0.00634 (0.030)	-0.187** (0.071) ⁺	0.0199 (0.042)	0.0785 (0.079)	-0.00419 (0.037)	-0.000947 (0.077)	0.00841 (0.077)	0.0319 (0.036)
Constant	-0.0575 (0.149)	0.0732 (0.204)	-0.236* (0.114)	-0.0254 (0.229)	-0.322* (0.174)	0.156 (0.208)	-0.0701 (0.147)	0.172 (0.104)	-0.232 (0.226)	-0.378* (0.159)	0.126 (0.191)	-0.169 (0.128)
Observations	34500	17300	17200	16800	17700	2000	32500	6000	28500	11600	15200	7600
Adj. R^2	0.171	0.188	0.168	0.173	0.168	0.161	0.172	0.189	0.168	0.182	0.168	0.169

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. The omitted ethnicity is white, so that all coefficient estimates represent the test score gap relative to white students. All regressions include the full set of controls and fixed effects. Western provinces include British Columbia, Alberta, Saskatchewan and Manitoba; Central provinces include Ontario and Quebec; Atlantic provinces include Prince Edward Island, Nova Scotia, New Brunswick and Newfoundland and Labrador. Standard errors clustered by province are in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Given the small number of clusters, we also report significance levels on all “Identity” coefficients from the wild cluster bootstrap constructed using the `boottest` stata package of Nielsen et al. (2019) with 999 replications: ⁺ $p < 0.1$, ⁺⁺ $p < 0.05$, ⁺⁺⁺ $p < 0.01$.

test scores of black and white girls. In our sample, girls appear to be driving the higher test scores among South Asian and Chinese students.

There are large differences in the Indigenous-white test gap for students in low-income families. Indigenous students whose LICO ratio is below the median score an average of nearly 0.4 standard deviations lower than non-Indigenous students who are also below the median LICO ratio. The gap among students who live in households with above median LICO ratios is 0.078 and not statistically significant. These findings are important and worth highlighting, as they suggest that Indigenous students who face specific adversities—those who come from lower income families, who live in single parent families, and who are born at a low birth weight—have potentially less support for overcoming the ensuing academic challenges in comparison to non-Indigenous students.

Indigenous students who come from single parent families also seem to be driving the test gap, as well as those who are born at a low birth weight, and the Indigenous-white test gap is larger among students living in Western and Central, rather than Atlantic, provinces. The black-white test gap, on the other hand, does not differ as markedly between high- and low-income students or between single parent and two parent families, although the relation between the test gap and birth weight is similar to that of the Indigenous-white test gap. The black-white test gap is entirely driven by students living in central Canada. There are no other clear patterns in the difference between the test scores of white and Chinese or South Asian students.

As discussed in the data section, parents had to consent to the test in the household section of the interview prior to its administration. Unfortunately, in some waves the response rate for the math component was actually quite low. For instance, approximately 50% of students in grade 2 and above completed the test in cycle 1. To account for the possibility that households who did not consent to the test were in some way different from those who did consent, we run additional estimations using a Heckman selection model (Heckman, 1979). These results are found in Table 4 and suggest that selection into writing the math test leads to a bias in the coefficient estimates of the test score gap. Although the largest change in coefficient estimates occurs in the estimate of the black-white test gap, the Indigenous-white test gap also decreases in magnitude from -0.313 to -0.220 after applying this correction.

To understand whether minority children begin school at a disadvantage, or if the test gap

Table 4: Heckman Corrected Estimates of Test Score Gap

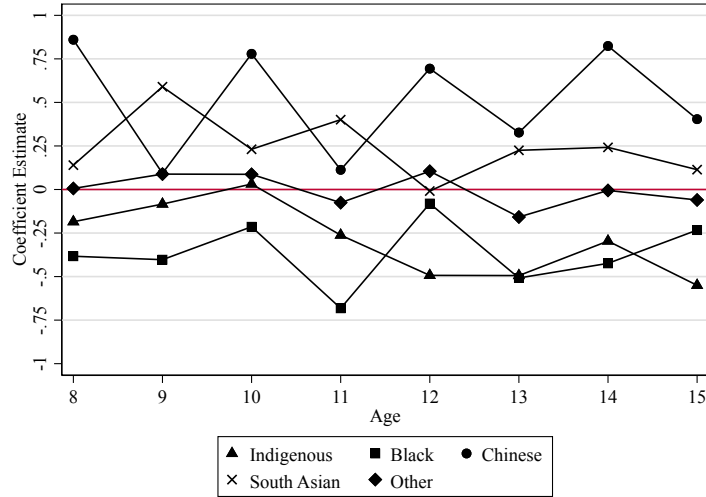
(1)	
Heckman Adjusted Estimates	
Identifies Indigenous	-0.220*** (0.045)
Identifies Black	-0.140** (0.065)
Identifies South Asian	0.231*** (0.064)
Identifies Chinese	0.489*** (0.060)
Identifies Other	-0.0194 (0.016)
Constant	0.165** (0.079)
Pr(Wrote Test)	
Identifies Indigenous	-0.284*** (0.057)
Identifies Black	0.0868 (0.100)
Identifies South Asian	-0.264*** (0.083)
Identifies Chinese	0.235** (0.099)
Identifies Other	-0.249*** (0.020)
Constant	0.226*** (0.078)
atanh ρ	-0.00211 (0.038)
ln(σ)	-0.0959*** (0.004)

Notes: The dependent variable in the top panel is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. The dependent variable in the second panel is whether the individual wrote the math test. Standard errors clustered by province are in parentheses. All columns include the full set of controls, as well as province, age, and survey wave fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

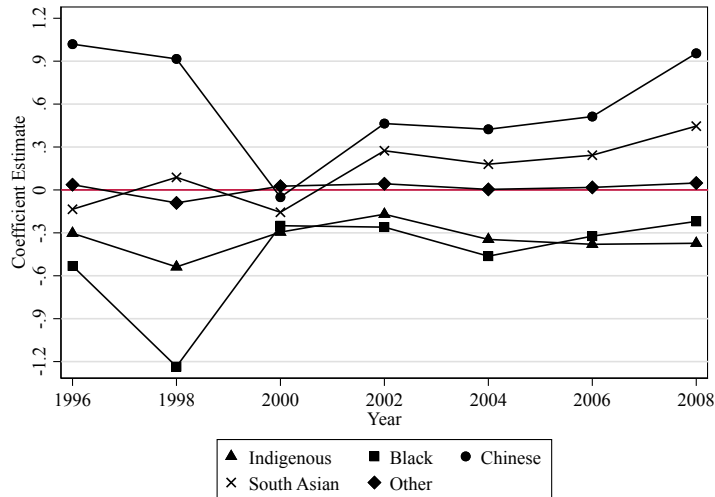
develops while the child is in school, we present results by age in Figure 1(a).²² Each cross-

²²Specifically, we run the following specification for each age, controlling for survey wave fixed effects, γ_s :

$$Y_{i,r,s} = \alpha + \sum_k \beta_k \mathbb{1}(\text{ethnicity}_i = k) + \mathbf{X}_{i,\alpha} \boldsymbol{\theta} + \gamma_s + \zeta_r + \epsilon_{i,r,s} \quad (2)$$



(a) Test Gap by Age



(b) Test Gap by Cycle

Figure 1: Regression adjusted test gap between minority and white students by age and cycle. Each cross section includes the coefficient estimates from a separate regression that includes the full set of controls and fixed effects.

section represents the coefficient estimates on each of the minority indicators for a separate age. The regression includes the full set of controls, analogous to column (7) in Table 2. Table 12 of the Online Appendix displays the full results, including standard errors and coefficients on controls. Figure 1(a) reveals that the test gap between Indigenous and white students is larger for older students and Table 12 shows that the estimate of this test gap for early ages is not

statistically different from 0. The Indigenous-white test gap peaks when students are 15 years of age at approximately -0.55 standard deviations. This finding is comparable to the work of Fryer and Levitt (2004) on the black-white test gap in the United States, which demonstrates that, after controlling for observables, the black-white gap in test scores among kindergarten students disappears, but that a persistent gap in test scores develops during the schooling years. Our results suggest that, at least for Indigenous students, the gap may become even larger in later years. Our findings differ slightly from Clotfelter et al. (2009), in that their estimates reveal a fairly consistent test gap for students in grades 3 through 8 for both Indigenous as well as black students. Figure 1(a) does not depict a downward trend in the black-white test gap, as in Fryer and Levitt (2004). We do see a slight upward trend in the test gap between Chinese and white students, which is similar to Clotfelter et al. (2009).

Since the NLSCY was conducted over a time period when discussion of the socioeconomic disparities faced by Indigenous peoples was gaining traction in policy and government circles, Figure 1(b) examines whether the test gap improved over time. Each cross section represents the regression-adjusted test gap for each of the minority groups (relative to white students) for a separate survey wave.²³ Table 13 of the Online Appendix displays the coefficient estimates for the full results, including all coefficients on controls. The first wave of the NLSCY was administered in 1994 and the last wave in 2008, however, we exclude 1994 because the test administered in this year underestimated students' abilities and resulted in nearly everyone receiving a perfect score.

The first thing to notice in this graph is that 1998 appears to be an outlier, in that the scores of Chinese and black students were far outside the trend for each of these groups. The second point is that there does not appear to be a noticeable trend in the test gap for any race. Indigenous and black students have lower test scores than white students over the entire time period, while Chinese and South Asian students have higher test scores compared to their white counterparts. For Indigenous students, the test score gap stays relatively close to 0.3 for the entire time period. These coefficient estimates are not statistically significant in 1996 and 1998, and are significant at the 10% level for the remainder of the cycles.

²³The coefficient estimates are obtained from the following specification for each survey wave, controlling for age fixed effects, δ_a :

$$Y_{i,r,a} = \alpha + \sum_k \beta_k \mathbb{1}(\text{ethnicity}_i = k) + \mathbf{X}_{i,s} \boldsymbol{\theta} + \delta_a + \zeta_r + \epsilon_{i,r,a} \quad (3)$$

The lack of any measurable improvement in the test gap between Indigenous and white students over this time period is troubling. In 1996, the Royal Commission on Aboriginal Peoples released a report documenting 5 years of public hearings, visitations to Indigenous communities, consultations with experts, research studies, and examinations into past inquiries and reports, to understand how to create a better relationship between Indigenous and non-Indigenous peoples in Canada. Among their findings, they highlighted the failure of provincial education systems to adequately reach the Indigenous population, with the majority of Indigenous youth leaving high school before graduation. The reasons, they suggested, were a combination of underfunding, a lack of mechanisms of accountability to Indigenous peoples, lack of cultural compatibility, and a focus on childhood education that left out important cultural components. Since our analysis does not include the on-reserve population, who may have been more likely to be affected by policies addressing these shortcomings, we cannot tell whether the report was followed by improvements among the on-reserve population. However, we do not see any improvement in the test gap for the off-reserve population during this time period, suggesting that even if these policies were effective at addressing educational disparities on reserves, there was no change in the test gap for the nearly 58% of Indigenous peoples who live outside reserves.

3.1.2 Oaxaca-Blinder Decomposition

To further understand the determinants of the average test score gap we estimate Oaxaca-Blinder (OB) decompositions. This methodology allows us to examine the relative importance of differences in the levels of observable characteristics, like income, have contributed to the mean test gap between minority and white students, and the importance of the returns to these characteristics. In what follows, we use the notation from Fortin, Lemieux, and Firpo (2011). The main purpose of the OB methodology in our context is to decompose the overall difference in average test scores between minority (M) and white (W) students:

$$\hat{\Delta}_Y^\mu = \bar{Y}_M - \bar{Y}_W$$

into an unexplained component that captures differences in the returns to the covariates and an explained component that captures differences in the levels of the covariates:

$$\hat{\Delta}_Y^\mu = \underbrace{(\hat{\alpha}_M - \hat{\alpha}_W) + \sum_{k=1}^K \bar{X}_{Mk}(\hat{\theta}_{Mk} - \hat{\theta}_{Wk})}_{\text{Unexplained}} + \underbrace{\sum_{k=1}^K (\bar{X}_{Mk} - \bar{X}_{Wk})\hat{\theta}_{Wk}}_{\text{Explained}},$$

where $\hat{\alpha}_{g0}$ and $\hat{\theta}_{gk}$, $k \in \{1, \dots, K\}$ and $g \in \{I, N\}$ are the estimated intercept and slope coefficients of equation 1 for minority and white students.

The intuition of the OB decomposition is that minority and white households may differ across several observable characteristics that are important determinants of children’s test scores. For example, as we show in Table 1 the average Indigenous household is poorer and parents have lower levels of education than the average non-Indigenous household. These level differences certainly play a role in explaining the test gap observed between Indigenous and non-Indigenous children, which would be captured by the explained portion of the OB decomposition. On the other hand, if Indigenous and non-Indigenous groups face differences in the returns to these observables, perhaps due in part to discrimination, this would be reflected by the unexplained portion of the decomposition. It is important to note that, while the OB decomposition is a useful tool to examine the structure of the test gap between minority and white students, it does not provide causal estimates of the various factors that contribute to the test gap. Rather, the OB decomposition suggests possible reasons why we observe differences in test scores.

Table 5 displays the results from the overall decomposition. Each column presents the results of the OB decomposition for a separate minority group with respect to white students. The first two rows reiterate what the descriptive statistics show—the average standardized test score for white students is -.0542. For those who identify as Indigenous, the average standardized test score is approximately -0.619 compared to -.473 for those who identify as black, .559 for those who identify as Chinese, and .135 for those who identify as being South Asian. The third row summarizes the difference between test scores of non-minority students and each respective ethnic group—e.g., $\bar{Y}_M - \bar{Y}_W = 0.565$ for Indigenous students. The last two rows of panel A decompose this gap into a portion that is attributable to differences in observable characteristics (the explained part) and a portion that is attributable to difference in returns (the unexplained part). What is immediately

Table 5: Oaxaca-Blinder Decomposition of Test Score Gap

	Indigenous	Black	Chinese	South Asian
Identifies White	-0.0542 (0.061)	-0.0542 (0.061)	-0.0542 (0.061)	-0.0542 (0.061)
Identifies Minority	-0.619*** (0.070)	-0.473*** (0.063)	0.559*** (0.029)	0.135*** (0.016)
Difference	0.565*** (0.099)	0.419*** (0.068)	-0.613*** (0.064)	-0.190*** (0.055)
Explained	0.207*** (0.048)	0.194*** (0.021)	0.00591 (0.067)	0.0958* (0.057)
Unexplained	0.358*** (0.074)	0.225*** (0.075)	-0.619*** (0.067)	-0.285*** (0.017)
Observations	24000	24000	24000	24000

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. Standard errors clustered by province are in parentheses. All columns include CMA-province and age fixed effects and the decomposition is computed using the full set of controls. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

clear is that differences in returns account for more of the gap than differences in observables for all groups.

Approximately 37% of the difference in test scores between Indigenous and white students is attributable to the fact that these groups have different levels of covariates. Similarly, 46% of the black-white test gap is attributable to this explained portion. What is troubling, is that this means that 63% of the difference in test scores between Indigenous and white students is attributable to the unexplained component, which represents differences in the returns to covariates. While it may not necessarily be easy to reduce differences in observable characteristics, the interpretation of these differences lends itself to somewhat clearer policy interpretations. For example, if Indigenous families have, on average, lower incomes than white families, and if this contributed in a meaningful way to the test gap, then the policy implication would be to provide income assistance to these families. However, a difference in the return to covariates has a less straightforward interpretation and consequently, a less straightforward set of policy implications.

The full characteristic level decomposition can be found in Table 14. This exercise suggests that household income, parental education, teenage pregnancies, and single parent households are driving the majority of the explained gap in test scores. Policies targeted accordingly may assist in reducing the test score gap.

For the Indigenous-white test gap, the only variables that are statistically different from zero are the number of siblings, whether the primary care giver has a high school degree, whether the spouse has a post-secondary degree, and whether the student attends a public school, catholic school, or is missing the school category. The interpretation of the coefficient on the number of siblings is for each additional sibling, non-Indigenous students test score increases by 0.249 standard deviations more than for Indigenous students. Thus the *return* to having siblings is higher for non-Indigenous students than for Indigenous students.

Similarly, the interpretation of the coefficient on whether the primary care giver has high school is that the average test score of a non-Indigenous student whose primary care giver has a high school degree increases by 0.430 standard deviations more than the increase in test scores achieved when an Indigenous student's primary care giver has a high school degree. One finding of particular interest is that there are no differences in the returns to attending public schools between Indigenous and white youth. This is in contrast to the results in Fryer and Levitt (2004) who suggest that the leading explanation for the worsening of the black-white test gap across ages is that blacks attend worse quality schools than whites. Our findings are perhaps promising along this front, as they suggest that the provincial public school systems in Canada may have the potential to equalize the playing field for traditionally marginalized groups. Of course this result comes with the caveat that we cannot speak to school quality on reserves.

For black students, the return to having more siblings is also lower than for white students, and so is the return to income, and the primary care giver's education. Interestingly, white children in single parent families perform 0.342 standard deviations lower than black children from single parent families. This result differs markedly from the Indigenous-white test gap. Although it is difficult to target the unexplained portion of the test gap, the results of the OB decomposition suggest that universal policies may not be the answer to addressing the test score gap across all minority groups.

3.2 Distributional Analysis: Going Beyond the Mean

The previous set of results are useful for quantifying the test gap over time, across ages, and more generally, but they do not reveal whether there are inequalities at different parts of the distribution of test scores. In this section we address this question in part by using quantile

regressions to examine the gap across the full distribution of test scores. We also implement the decomposition methodology of Lemieux (2002) to evaluate the contribution of the levels of and returns to covariates at explaining the test gap at different points in the distribution of test scores. This decomposition method generates three counterfactual distributions: the first counterfactual distribution applies the coefficients from non-Indigenous students to Indigenous students; the second counterfactual re-weights the distribution of Indigenous test scores to have the same distribution of covariates as the non-Indigenous group; the final counterfactual applies both the counterfactual coefficients and covariates to the Indigenous test score distribution.

3.2.1 Quantile Regression

To get a sense of the magnitude of the test gap at different points in the distribution, Table 6 displays both the uncontrolled and controlled results of estimating equation 1 by quantile regressions. Although the Indigenous-white test gap seems to be slightly larger in higher quantiles, a discernible pattern does not exist in either the uncontrolled, nor the controlled quantile regressions. The black-white test gap is also generally larger for higher quantiles and the test gap for South Asian and Chinese students is fairly constant throughout the distribution.

Table 6: Quantile Regression Estimates of Differences in Test Scores

	OLS		p(10)		p(25)		p(50)		p(75)		p(90)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Identifies Indigenous	-0.490*** (0.061)	-0.347*** (0.060)	-0.399*** (0.029)	-0.284*** (0.046)	-0.420*** (0.053)	-0.243*** (0.048)	-0.591*** (0.031)	-0.410*** (0.039)	-0.483*** (0.156)	-0.254** (0.101)	-0.445*** (0.128)	-0.360*** (0.112)
Identifies Black	-0.378*** (0.096)	-0.271*** (0.095)	-0.0964 (0.359)	-0.0901* (0.055)	-0.244*** (0.064)	-0.294 (0.210)	-0.456*** (0.052)	-0.284* (0.165)	-0.461*** (0.053)	-0.448*** (0.071)	-0.386*** (0.133)	-0.313*** (0.121)
Identifies South Asian	0.197** (0.091)	0.258*** (0.090)	0.287*** (0.041)	0.419*** (0.037)	0.146 (0.154)	0.178 (0.141)	0.154* (0.086)	0.247*** (0.044)	0.228*** (0.028)	0.367*** (0.134)	0.225*** (0.062)	0.242*** (0.040)
Identifies Chinese	0.595*** (0.087)	0.575*** (0.091)	0.323*** (0.111)	0.477*** (0.066)	0.698*** (0.129)	0.602*** (0.060)	0.686*** (0.128)	0.629*** (0.094)	0.624*** (0.053)	0.605*** (0.045)	0.558*** (0.068)	0.611*** (0.033)
Identifies Other	-0.0114 (0.027)	0.0201 (0.026)	0.0189 (0.042)	0.0549* (0.030)	0.0470* (0.027)	0.0376 (0.028)	0.00152 (0.030)	0.0161 (0.027)	-0.0529* (0.029)	0.00622 (0.030)	-0.0763* (0.045)	-0.0395 (0.035)
Controls	X		X		X		X		X		X	
Observations	34500	34500	34500	34500	34500	34500	34500	34500	34500	34500	34500	34500

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. The omitted ethnicity is white, so that all coefficient estimates represent the test score gap relative to white students. All regressions include the full set of controls. All columns include CMA-province, age, and survey wave fixed effects. The omitted ethnicity is white, so that all coefficient estimates represent the test score gap relative to white students. Robust standard errors are in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

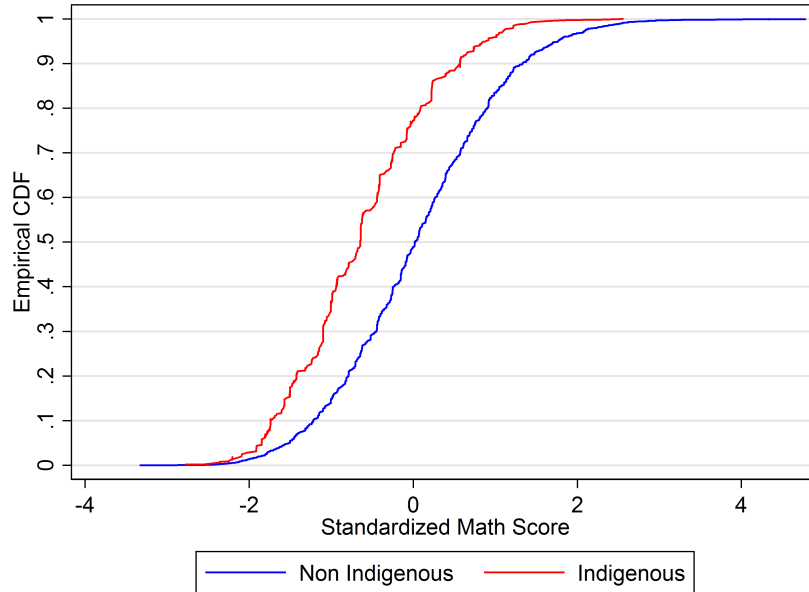


Figure 2: Empirical CDF of Indigenous and non-Indigenous test scores

One of the limitations of previous analysis is that it has the potential to be affected by scaling decisions (Ho and Haertel, 2006; Reardon, 2008; Bond and Lang, 2013; Penney, 2017; Bond and Lang, 2018). One way to analyze the test-gap in a metric free way cross-references percentiles of the Indigenous test score distribution with the percentiles of the non-Indigenous distribution. This exercise is invariant to monotone transformations of the test score distributions and therefore does not depend on the scaling choice of the test scores. Figure 2 displays the empirical cumulative density functions for Indigenous and non-Indigenous students. If there were no differences in the distribution of test scores, 10% of Indigenous students would score below the 10th percentile of the non-Indigenous distribution, 25% of Indigenous students below the 25th percentile of the non-Indigenous distribution, and so forth. The reality is much different. Table 7 summarizes the differences between the empirical distributions by analyzing the percent of Indigenous students scoring below specific percentiles of the non-Indigenous distribution. At all percentiles, Indigenous students score lower on their mathematics test. Nearly 25 percent of Indigenous students score below the 10th percentile of non-Indigenous students, 50 percent below the 25th percentile, and nearly 95 percent below the 75th percentile of non-Indigenous students.

Table 7: Percent of Indigenous Students Below Percentiles of Non-Indigenous Distribution

	p1	p10	p25	p50	p75	p90	p99
Identifies Indigenous	.025	0.24	0.50	0.78	0.94	0.99	1.00

3.2.2 Lemieux (2002) Methodology

To further our understanding of differences across the distribution of test scores, we use the decomposition method of Lemieux (2002). This allows us to evaluate the relative importance of differences in the levels of observable characteristics—like income or parental education—versus difference in the returns to the same characteristics in explaining differences in the distribution of test scores between minority and non-minority youth. In some sense, this technique is a generalization of the popular Oaxaca (1973) and Blinder (1973) decomposition of means to a decomposition of the full distribution of test scores. We focus here on the test score gap between Indigenous and non-Indigenous students, due to the fact that the larger number of Indigenous (compared to, e.g., black) students in the dataset permits this type of distributional analysis.

The first step is to estimate the relationship between the test gap and the set of observables separately for the Indigenous (I) and non-Indigenous (N) samples:

$$Y_{i,r,a,s} = \alpha + \mathbf{X}_{i,a,s}\boldsymbol{\theta} + \delta_a + \gamma_s + \zeta_r + \epsilon_{i,r,a,s}, \quad (4)$$

from which we can obtain the predicted values from the non-Indigenous regression, $\mathbf{X}_{i,a,s}\widehat{\boldsymbol{\theta}}^N$, and the residuals from the Indigenous regression, $\hat{\epsilon}_{i,r,a,s}^I$. Together, these two vectors give the distribution of test scores under the counterfactual scenario where Indigenous students are assigned the same returns to characteristics as their non-Indigenous counterparts:

$$Y_{i,r,a,s}^{coef} = \mathbf{X}_{i,a,s}\widehat{\boldsymbol{\theta}}^N + \hat{\epsilon}_{i,r,a,s}^I \quad (5)$$

To account for differences in the distribution of covariates, we run a probit model of the Indigenous indicator on our full set of covariates:

$$\text{Indigenous}_{i,r,a,s} = \mathbf{X}_{i,a,s}\boldsymbol{\Psi} + \delta_a + \gamma_s + \zeta_r + u_{i,r,a,s}, \quad (6)$$

Table 8: Procedure for Generating Counterfactual Distributions

Variable	Weight	Distribution
$Y_{i,r,a,s}$	$\omega_{i,r,a,s}$	Raw distribution
$Y_{i,r,a,s}^{coef}$	$\omega_{i,r,a,s}$	Coefficients
$Y_{i,r,a,s}$	$\omega_{i,r,a,s}^{covar}$	Covariates
$Y_{i,r,a,s}^{coef}$	$\omega_{i,r,a,s}^{covar}$	Coefficients and covariates

from which we obtain $P_{i,r,a,s}^I$, the predicted probability that individual i from region r of age a in survey s with covariates $\mathbf{X}_{i,a,s}$ is Indigenous:

$$P_{i,r,a,s}^I = \Pr(\text{Indigenous} = 1 | \mathbf{X}_{i,a,s})$$

We use this probability, combined with the unconditional probability of being Indigenous, P^I , to create a weighting factor that re-weights the covariates of the Indigenous population to have the distribution of covariates of the non-Indigenous population:

$$\omega_{i,r,a,s}^{covar} = \frac{1 - P_{i,r,a,s}^I}{P_{i,r,a,s}^I} * \frac{P^I}{1 - P^I} * \omega_{i,r,a,s},$$

where $\omega_{i,r,a,s}$ are the original sampling weights included in the survey. This re-weighting exercise produces the distribution of test scores under the counterfactual scenario where Indigenous students are assigned the same level of covariates as the non-Indigenous population.

We use the two counterfactuals, $Y_{i,r,a,s}^{coef}$ and $\omega_{i,r,a,s}^{covar}$ to estimate three counterfactual test score distributions for those who identify as Indigenous: one in which Indigenous students have the same level of covariates as the non-Indigenous population, one in which the returns to covariates are the same, and one in which both the level and returns are the same. Once each of these components is account for, the remaining difference in test scores is attributed to unobserved differences between Indigenous and non-Indigenous students.²⁴ Table 8 summarizes how we combine each of the counterfactuals to produce a distribution.

Like the OB decomposition, the Lemieux (2002) decomposition does not produce causal estimates of the various factors that contribute to the test gap. It simply suggests possible reasons

²⁴The unobserved component may be either level differences or differences in the returns to unobservables. We do not differentiate between the two here.

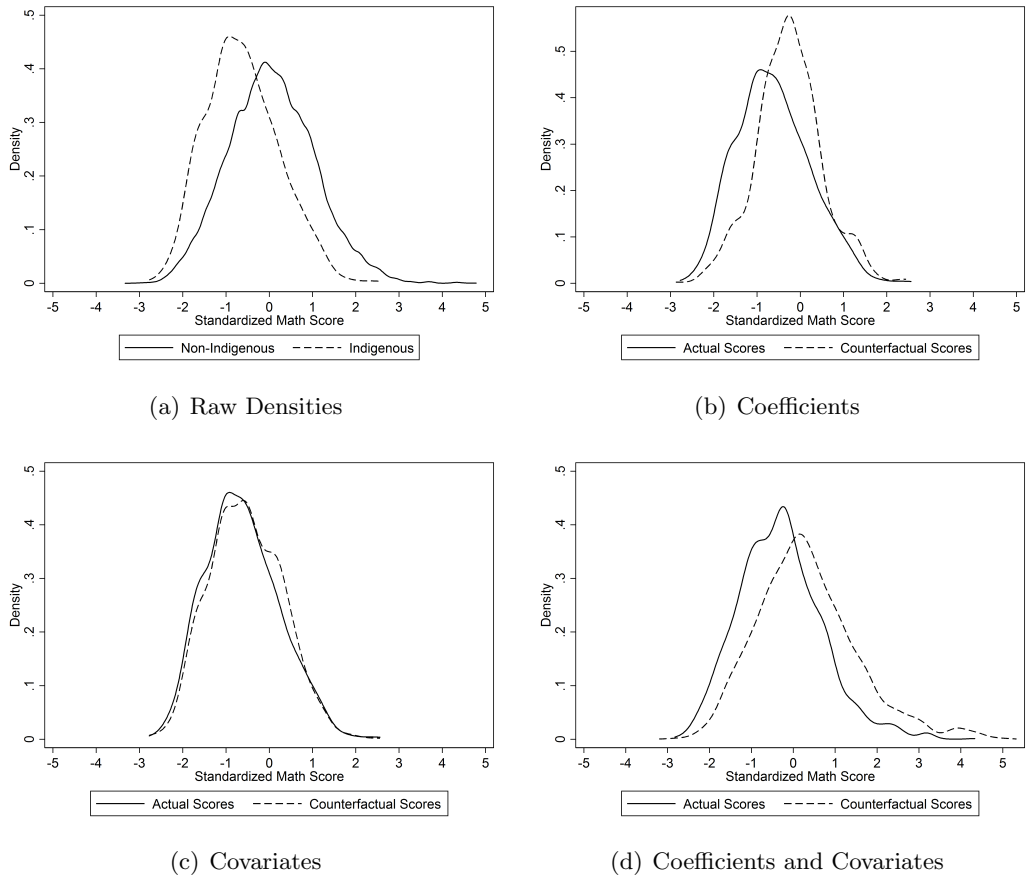


Figure 3: Comparison between the true distribution of Indigenous and non-Indigenous test scores and the counterfactual distributions computed using the methodology of Lemieux (2002).

why we observe differences in test scores between Indigenous and non-Indigenous students at different points in the test score distributions, which may assist in better targeting policies to assist in reducing the test score gap. Figure 3 compares the three counterfactual distributions to the true distribution of test scores. Table 9 summarizes the differences between each counterfactual Indigenous distribution and actual non-Indigenous distributions using OLS and quantile regression.

Figure 3(a) begins by displaying the true distributions of standardized test scores for the Indigenous and non-Indigenous populations to demonstrate how they differ initially. These distributions are the raw standardized test score distributions weighted by the survey weights included in the NLSCY. They do not correct for age, cycle, or CMA-province fixed effects. As was evident from the previous set of OLS regressions, we can see that there are large differences in the means of the two groups, but there are also other differences that would not be apparent from simply analyzing

the mean. Most notably, the non-Indigenous distribution has a much longer right tail. The first panel in Table 9 summarizes the results of Figure 3(a), where we see that the test score gap in the 90th percentile is over 0.2 standard deviations larger than at the 10th percentile.

Figure 3(b) shows how the non-Indigenous distribution compares to the counterfactual Indigenous distribution constructed from assigning the non-Indigenous coefficients ($Y_{i,r,a,s}^{coef}$) to the Indigenous group. This exercise eliminates most of the gap in the left tail of the distribution, but does little to bring the right tail of the Indigenous distribution closer to that of the non-Indigenous distribution. In panel B of Table 9, we see that the test gap is almost 0 and statistically insignificant at the 10th percentile, but only reduces in magnitude from -0.759 to -0.649 at the 90th percentile.

Re-weighting the characteristics of Indigenous students to be comparable to those of non-Indigenous students ($\omega_{i,r,a,s}^{covar}$) appears to be less effective at reducing the test gap. This counterfactual distribution is displayed in Figure 3(c) and is slightly to the right of the original Indigenous distribution, indicating that level differences do play a small role in reducing the test gap. Panel C of Table 9 tests whether the differences between the counterfactual covariate distribution and the actual distribution are statistically significant. The test score gap is reduced in the 10th percentile, from approximately -0.523 to -0.352, and this difference is statistically significant. At the 90th percentile there is no change in the test score gap.

The final counterfactual distribution computes the combined effect of both counterfactual coefficients and covariates. Figure 3(d) displays this counterfactual. The counterfactual distribution from the combined coefficient and covariate effects lies to the right of the original distribution for the lower tail, suggesting that the combined effect more than eliminates the test score gap for the lowest quantiles. Panel D of Table 9 reveals, however, that this difference is not statistically different from 0. The test gap is reduced from -0.664 to -0.270 at the median, from -0.792 to -0.493 at the 75th percentile, but remains unchanged at the 90th percentile.

Interestingly, the results of this decomposition exercise suggest that once we have accounted for differences in the returns and levels of the observable characteristics, unobservable factors no longer help to explain the test gap for the lower end of the distribution, but continue to play a sizeable role at the upper end. The covariate analysis provides a starting point for policy makers to target specific characteristics—parental education, income, teenage pregnancies, low birth weight,

Table 9: Differences between Counterfactual Indigenous and non-Indigenous Distributions

	OLS	p(10)	p(25)	p(50)	p(75)	p(90)
Panel A: Actual Distribution						
Identifies Indigenous	-0.677*** (0.058)	-0.523*** (0.061)	-0.514*** (0.105)	-0.664*** (0.041)	-0.792*** (0.063)	-0.759*** (0.061)
Panel B: Counterfactual Coefficients						
Identifies Indigenous	-0.325*** (0.047)	-0.00120 (0.243)	-0.126* (0.075)	-0.282*** (0.021)	-0.493*** (0.075)	-0.649** (0.272)
Panel C: Counterfactual Covariates						
Identifies Indigenous	-0.581*** (0.073)	-0.353* (0.211)	-0.456*** [†] (0.017)	-0.640*** (0.194)	-0.625*** (0.208)	-0.759*** (0.074)
Panel D: Counterfactual Covariates and Coefficients						
Identifies Indigenous	-0.260*** (0.051)	0.216 (0.218)	0.0518 (0.095)	-0.270*** (0.035)	-0.493*** (0.095)	-0.754*** (0.059)
N. Obs	34500	34500	34500	34500	34500	34500

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. Standard errors clustered by province are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. [†] indicates that this estimate was computed at the 25.3th percentile, as there was not enough variation at the 25th percentile to compute the estimate.

etcetera—that can assist in reducing the gap amongst the lowest performing students. Furthermore, our results suggest that program evaluation of such policies may want to consider the potential indirect effects on children’s outcomes, like test scores.²⁵

The findings of the counterfactual coefficient analysis, although useful in that they explain a large portion of the gap in test scores at the lowest end of the distribution, do not lend themselves to straightforward policy interpretations, as differences in the returns to covariates may represent a number of phenomena. For instance, they may represent differences in school quality, discrimination, or differences in parents’ willingness to invest in children given their level of education. Understanding which, if any, of these factors can explain why Indigenous students face different returns to their characteristics compared to non-Indigenous students will require additional work to examine each of these channels.

²⁵For example, studies of income support (Levine et al., 2005; Forget, 2011), or post-secondary funding programs (Jones, 2017) may not fully capture the potential benefits of such programs that extend to future generations.

Finally, accounting for differences in the returns to and levels of covariates does not eliminate the test gap in the highest end of the distribution. This suggests that unobservable characteristics—either the levels of unobservables or the returns to unobservables—explain a large portion of the gap in the upper tail. This result is perhaps the most worrying of all our findings, since it suggests that without more research on the determinants of the test score gap, there is little that can be done in terms of implementing policies that will assist in reducing the gap in the highest end of the distribution.

4 Conclusion

This paper quantifies the gap in math test scores between minority and non-minority youth in Canada with a particular focus on the Indigenous-white test gap. We use data from the National Longitudinal Survey of Children and Youth (NLSCY) to document that, conditional on a rich set of controls, including household income, parental education, and other background characteristics, test scores are approximately 0.310 standard deviations lower for Indigenous students in comparison to white students. While not quite as large as the estimates of the test gap between American Indian and white students found by Clotfelter et al. (2009), our results may over or underestimate the test gap since our sample does not include the on-reserve population, where education falls under federal jurisdiction, in contrast to off-reserve schools, which are under provincial control. This data limitation excludes just under half of Indigenous students in Canada.

Studies of the test gap have traditionally focussed on the United States and on differences between black and white students. We find a black-white test gap of approximately 0.335 standard deviations, which is slightly smaller than the estimates uncovered in the American studies (Fryer and Levitt, 2004, 2006; Clotfelter et al., 2009), and consistent with other work we find that Asian students outperform their white counterparts. The results of an Oaxaca-Blinder decomposition suggest that a potential difference between the Canadian and American contexts lies in the differences in the quality of public schools between the two countries. We find that the type of school students attend does not contribute substantially to the test gap for either Indigenous or black students, whereas Fryer and Levitt (2004) show that the black-white test gap appears among children in public schools, but not in private schools. The difference in findings between our study and Fryer and Levitt (2004) suggests that the provincial public schools in Canada may be better

equipped to equalize outcomes between minority and non-minority groups.

We show that the Indigenous-white test gap emerges after 11 years of age, but in contrast to studies of the black-white test gap in the United States, there are no distinct patterns in the test score gap for other minority groups. We do not find any discernible patterns in the Indigenous-white test gap across time, despite the fact that the year after the first wave of the NLSCY was administered, the results from a major public inquiry into the status of Indigenous peoples in Canada were released. This report, known as the Royal Commission on Aboriginal Peoples, lay out a twenty-year agenda to ameliorate the conditions of Indigenous peoples in Canada, which included a call to reform provincial public education systems in a way that would address specific barriers faced by Indigenous students. Our findings should therefore act as a cautionary note, since reducing educational disparities was a key component put forth more recently by the 2015 Truth and Reconciliation Commission’s 94 Calls to Action, as well as the 2019 Missing and Murdered Indigenous Women and Girls’ 231 Calls for Justice.

Looking beyond mean differences in test scores, we use the decomposition method of Lemieux (2002) to show that if Indigenous peoples had the same level of covariates and returns to characteristics as their non-Indigenous counterparts, the test gap between Indigenous and non-Indigenous students in the lowest quantiles of the test score distribution would be eliminated, but that the test gap at the top of the distribution would be unaffected. The results of this exercise suggest that it is an important tool for understanding between-group differences across the whole distribution of achievement outcomes.

While our results are in line with previous studies that show socioeconomic factors contribute considerably to the difference in test scores between minority and non-minority students, the findings of both the Oaxaca Blinder and the Lemieux (2002) methodology point to some difficulties in determining the most appropriate policies for reducing the test gap. First, general differences across ethnicities suggest that there is no “one size fits all” policy to reduce educational disparities among disadvantaged groups more broadly. Second, the majority of the test gap is attributable to differences in the returns to characteristics, rather than in the level of these characteristics. Since returns may differ for a variety of reasons, uncovering the *why* behind such variation in the return to observables is an important subsequent step in understanding which policies may best address

the test gap. We leave this for future work.

References

- Anderson, B. and J. Richards (2016). Students in jeopardy: An agenda for improving results in band-operated schools. *C.D. Howe Commentary No. 444*.
- Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2016a). Family disadvantage and the gender gap in behavioral and educational outcomes. *NBER Working Paper No. 22267*.
- Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman (2016b). School quality and the gender gap in educational attainment. *American Economic Review: Papers and Proceedings* 106(5), 289–295.
- Barsh, R. (1994). Canada’s Aboriginal peoples: Social integration or disintegration. *Canadian Journal of Native Studies* 14(1), 1–46.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Educatiaon* (Third ed.). The University of Chicago Press.
- Bertrand, M. and J. Pan (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics* 5(1), 32–64.
- Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources* 8(4), 436–455.
- Bond, T. N. and K. Lang (2013). The evolution of the black-white test score gap in grades K-3: The fragility of results. *Working Paper*.
- Bond, T. N. and K. Lang (2018). The black-white education-scaled test-score gap in grades k-7. *Journal of Human Resources* 53(Fall 2018), 891–917.
- Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics* 3, 1801–1863.
- Caron-Malenfant, Éric., S. Coulombe, E. Guimond, C. Grondin, and A. Lebel (2014). Ethnic mobility of aboriginal peoples in canada between 2001 and 2006 censuses. *Population* 69(1), 29–54.
- Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2009). The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics* 91(2), 398–419.
- Ding, W. and S. F. Lehrer (2014). Understanding the role of time-varying unobserved ability heterogeneity in education production. *Economics of Education Review* 40(C), 55–75.
- Feir, D. (2013). Size, structure, and change: Exploring the sources of Aboriginal earnings gaps in 1995 and 2005. *Canadian Public Policy* 32(2), 309–334.
- Forget, E. L. (2011). The town with no poverty: The health effects of a Canadian guaranteed annual income field experiment. *Canadian Public Policy* 37(3), 283–305.
- Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. *Handbook of Labor Economics* 4(1), 1–102.
- Friesen, J. and B. Krauth (2010). Sorting, peers, and achievement of aboriginal student in british columbia. *Canadian Journal of Economics* 43(3), 1273–1301.

- Fryer, R. G. and S. D. Levitt (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics* 86(2), 447–464.
- Fryer, R. G. and S. D. Levitt (2006). The black-white test score gap through third grade. *American Law and Economics Review* 8(2), 249–281.
- Gagné, R. (1968). Learning hierarchies. *Educational Psychologist* 6(1), 1–9.
- George, P. and P. Kuhn (1994). Size and structure of native-white wage differentials in Canada. *Canadian Journal of Economics* 27(1), 20–42.
- Guimond, E. (1999). Ethnic mobility and the demographic growth of Canada’s Aboriginal populations from 1986 to 1996. *Report on the Demographic Situation in Canada Statistics Canada Catalogue no. 91-209*, 187–200.
- Guimond, E. (2009). *L’explosion démographique des populations autochtones du Canada de 1986 à 2001*. Ph. D. thesis, University of Montréal.
- Hanushek, E. A. (2008). *Education Production Functions*. In: Palgrave Macmillan (eds) *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London.
- Heckman, J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5, 475–492.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), 153–161.
- Heckman, J. and G. Conti (2014). Understanding conscientiousness across the lifecycle: An economic perspective,. *Developmental Psychology* 50(5), 1451–1459.
- Heckman, J., P. Gertler, R. Pinto, A. Zanolini, C. Vermeerch, S. Walker, S. Chang, and S. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science* 344(6187), 998–1001.
- Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6), 2052–2086.
- Ho, A. D. and E. H. Haertel (2006). Metric-free measures of test score trends and gaps with policy-relevant examples. *CSE Report* 665.
- Jones, M. E. C. (2017). Student aid and the distribution of educational attainment. *QED Working Paper No. 1373*.
- Juhn, C., K. M. Murphy, and B. Pierce (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101(3), 410–442.
- Leigh, A. and X. Gong (2009). Estimating cognitive gaps between indigenous and non-indigenous australians. *Education Economics* 17(2), 239–261.
- Lemieux, T. (2002). Decomposing changes in wage distributions: a unified approach. *Canadian Journal of Economics/Revue canadienne d’économique* 35(4), 646–688.

- Levine, R. A., H. Watts, R. Hollister, W. Williams, A. O'Connor, and K. Widerquist (2005). *The Ethics and Economics of the Basic Income Guarantee*, Chapter A Retrospective on the Negative Income Tax Experiments: Looking Back at the Most Innovative Field Studies in Social Policy. Aldershot: Ashgate.
- Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour Economics* 12(4), 577–590.
- Murnane, R. J. (2013). U.S. high school graduation rates: Patterns and explanations. *Journal of Economic Literature* 51(2), 370–422.
- Murphy, B., X. Zhang, and C. Dionne (2012). Low income in Canada: a multi-line and multi-index perspective. *Statistics Canada Income Research Paper Series ISSN 1707-2840*.
- Nielsen, M. Ø., J. G. MacKinnon, D. Roodman, and M. D. Webb (2019). Fast and wild: Bootstrap inference in stata using boottest. *The State Journal* 19(1).
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review* 14(3), 693–709.
- Penney, J. (2017). Test score measurement and the black-white test score gap. *Review of Economics and Statistics*, forthcoming.
- Reardon, S. F. (2008). Thirteen ways of looking at the black-white test score gap. *Stanford University Institute for Research on Education Policy and Practice Working Paper 2008-08*.
- Richards, J. (2008). Closing the Aboriginal/non-Aboriginal education gaps. *G.D. Howe Institute Working Paper*.
- Richards, J., J. Howe, and K. Afolabi (2008). Understanding the aboriginal/non-aboriginal gap in student performance: Lessons from British Columbia. *C.D. Howe Institute Commentary*.
- Richards, J. G., A. R. Vining, and D. L. Weimer (2010). Aboriginal performance on standardized tests: Evidence and analysis from provincial schools in British Columbia. *Policy Studies Journal* 38(1), 47–67.
- Rose, H. and J. R. Betts (2004). The effect of high school courses on earnings. *The Review of Economics and Statistics* 86(2), 497–513.
- Sakellariou, C. (2008). Peer effects and the Indigenous/non-Indigenous early test-score gap in peru. *Education Economics* 16(4), 371–390.
- Sohn, K. (2012). The dynamics of the evolution of the black-white test score gap. *Education Economics* 20(2), 175–188.
- Todd, P. E. and K. I. Wolpin (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal* 113, F3–F33.

A Online Appendix Tables and Figures

Table 10: Assessment of Bias from Omitting LICO

	Full Sample	Has LICO	LICO Control
Identifies Indigenous	-0.450*** (0.057) ⁺⁺	-0.456*** (0.057) ⁺⁺	-0.380*** (0.047) ⁺⁺
LICO Ratio			0.0920*** (0.007)
Observations	34700	34500	34500
Adj. R^2	0.113	0.112	0.133

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. All columns include CMA-province, age, and survey wave fixed effects. Standard errors clustered by province are in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Given the small number of clusters, we also report significance levels on all “Identity” coefficients from the wild cluster bootstrap constructed using the `boottest` stata package of Nielsen et al. (2019) with 999 replications: ⁺ $p < 0.1$, ⁺⁺ $p < 0.05$, ⁺⁺⁺ $p < 0.01$.

Table 11: Difference in Test Scores between Indigenous and Non-Indigenous Youth (Inverse Propensity Weights)

	(1)	(2)	(3)	(4)	(5)	(6)
Identifies Indigenous	-0.401*** (0.034)+++	-0.340*** (0.030)+++	-0.318*** (0.032)+++	-0.309*** (0.027)+++	-0.318*** (0.025)+++	-0.314*** (0.022)+++
Identifies Black	-0.256** (0.109)	-0.226* (0.104)	-0.194* (0.096)	-0.252** (0.079)	-0.258** (0.082)	-0.245** (0.104)
Identifies South Asian	0.0671 (0.050)	0.130** (0.046)++	0.102* (0.050)++	0.123** (0.039)++	0.143*** (0.043)++	0.149*** (0.037)++
Identifies Chinese	0.442*** (0.081)+++	0.423*** (0.096)+++	0.424*** (0.094)+++	0.460*** (0.099)+++	0.462*** (0.100)+++	0.450*** (0.115)+++
Identifies Other	-0.0467 (0.035)	-0.00635 (0.039)	-0.00673 (0.040)	0.00696 (0.037)	0.0102 (0.036)	0.00655 (0.038)
Immigrant	0.267** (0.096)	0.310** (0.106)	0.306*** (0.091)	0.275*** (0.075)	0.274*** (0.079)	0.269*** (0.082)
Male	-0.0464 (0.053)	-0.0389 (0.050)	-0.0406 (0.049)	-0.0394 (0.053)	-0.0366 (0.052)	-0.0389 (0.052)
LICO Ratio		0.108*** (0.016)	0.106*** (0.018)	0.0720*** (0.010)	0.0702*** (0.011)	0.0658*** (0.009)
Single Parent			-0.0680 (0.043)	0.0723* (0.033)	0.0703* (0.033)	0.0629* (0.034)
Number of Siblings			0.0504*** (0.015)	0.0409** (0.015)	0.0386** (0.015)	0.0350* (0.017)
Primary Care High School				0.223*** (0.035)	0.213*** (0.034)	0.210*** (0.031)
Primary Care Post-Secondary				0.0883* (0.045)	0.0855* (0.045)	0.0753 (0.042)
Spouse High School				0.112*** (0.014)	0.108*** (0.015)	0.106*** (0.016)
Spouse Post-Secondary				0.140*** (0.023)	0.135*** (0.022)	0.132*** (0.021)
Low Birth Weight					-0.192*** (0.022)	-0.187*** (0.024)
Teenage Mother					-0.212*** (0.039)	-0.206*** (0.041)
Public School						-0.0101 (0.121)
Catholic School						-0.0181 (0.135)
Private School						0.314* (0.140)
Other School						0.822** (0.280)
Constant	0.534*** (0.040)	0.305*** (0.056)	0.253** (0.085)	-0.0479 (0.099)	-0.0128 (0.102)	0.0103 (0.206)
Observations	34500	34500	34500	34500	34500	34500
Adjusted- R^2	0.135	0.158	0.161	0.178	0.181	0.187

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. The omitted ethnicity is white, so that all coefficient estimates represent the test score gap relative to white students. Omitted schooling category is whether the school information is missing, and the omitted category for parental education is no formal schooling. All columns include CMA-province, age, and survey wave fixed effects. Standard errors clustered by province are in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Given the small number of clusters, we also report significance levels on all “Identity” coefficients from the wild cluster bootstrap constructed using the `boottest` stata package of Nielsen et al. (2019) with 999 replications: + $p < 0.1$, ++ $p < 0.05$, +++ $p < 0.01$.

Table 12: Difference in Test Scores between Indigenous and Non-Indigenous Youth by Age

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	8	9	10	11	12	13	14	15
Identifies Indigenous	-0.185 (0.172)	-0.0837 (0.185)	0.0316 (0.115)	-0.262 (0.196)	-0.493** (0.165) ⁺	-0.494*** (0.148) ⁺⁺	-0.296** (0.100) ⁺	-0.550** (0.176) ⁺
Identifies Black	-0.383 (0.213)	-0.403*** (0.048) ⁺⁺	-0.214 (0.151)	-0.682*** (0.103) ⁺	-0.0804 (0.061)	-0.508** (0.217)	-0.424** (0.187)	-0.233 (0.236)
Identifies South Asian	0.140 (0.127)	0.590*** (0.108)	0.231*** (0.023)	0.401** (0.134) ⁺	-0.00872 (0.058)	0.225** (0.077) ⁺	0.242* (0.117) ⁺	0.114 (0.188)
Identifies Chinese	0.860*** (0.229) ⁺⁺⁺	0.0928 (0.077)	0.779*** (0.061) ⁺⁺	0.113* (0.061)	0.694** (0.235) ⁺	0.327*** (0.081)	0.824** (0.289) ⁺⁺	0.404*** (0.081) ⁺
Identifies Other	0.00603 (0.054)	0.0885 (0.085)	0.0872* (0.045)	-0.0755 (0.050)	0.105 (0.100)	-0.158 (0.134)	-0.00530 (0.076)	-0.0601 (0.094)
Constant	0.420** (0.168)	-0.243 (0.428)	0.00803 (0.107)	0.292 (0.204)	0.465 (0.614)	-0.203 (0.244)	-0.711** (0.269)	-0.756** (0.283)
Observations	3700	3500	4500	4200	5100	4700	4500	4300
Adjusted- R^2	0.133	0.157	0.200	0.216	0.224	0.197	0.160	0.176

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. The omitted ethnicity is white, so that all coefficient estimates represent the test score gap relative to white students. Each column includes the full set of controls and fixed effects. Standard errors clustered by province are in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Given the small number of clusters, we also report significance levels on all “Identity” coefficients from the wild cluster bootstrap constructed using the `boottest` stata package of Nielsen et al. (2019) with 999 replications: ⁺ $p < 0.1$, ⁺⁺ $p < 0.05$, ⁺⁺⁺ $p < 0.01$.

Table 13: Difference in Test Scores between Indigenous and Non-Indigenous Youth by Year

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1996	1998	2000	2002	2004	2006	2008
Identifies Indigenous	-0.303 (0.210)	-0.538 (0.350)	-0.295** (0.121) ⁺	-0.170** (0.073) ⁺	-0.346** (0.121) ⁺	-0.380* (0.176) ⁺	-0.373* (0.178) ⁺
Identifies Black	-0.531*** (0.088)	-1.235*** (0.193)	-0.250* (0.131)	-0.260* (0.119)	-0.463*** (0.115)	-0.323 (0.286)	-0.220 (0.258)
Identifies South Asian	-0.135 (0.348)	0.0877 (0.076)	-0.156 (0.095)	0.274*** (0.073) ⁺	0.180*** (0.044)	0.243** (0.097)	0.446*** (0.095) ⁺
Identifies Chinese	1.019** (0.429) ⁺	0.916** (0.388)	-0.0507 (0.192)	0.464** (0.149) ⁺⁺	0.424** (0.137) ⁺⁺	0.513*** (0.131) ⁺⁺	0.955*** (0.220) ⁺⁺⁺
Identifies Other	0.0368 (0.074)	-0.0906 (0.089)	0.0263 (0.053)	0.0432 (0.044)	0.00372 (0.085)	0.0176 (0.074)	0.0484 (0.081)
Constant	0.0294 (0.338)	-0.505* (0.244)	-0.752*** (0.159)	-0.246 (0.158)	0.383*** (0.088)	-0.710 (0.392)	-0.712** (0.260)
Observations	4800	4500	6200	7200	5800	4000	2000
Adjusted- R^2	0.121	0.189	0.132	0.129	0.170	0.172	0.183

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. The omitted ethnicity is white, so that all coefficient estimates represent the test score gap relative to white students. Standard errors clustered by province are in parentheses. Each column includes the full set of controls. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Given the small number of clusters, we also report significance levels on all “Identity” coefficients from the wild cluster bootstrap constructed using the `boottest` stata package of Nielsen et al. (2019) with 999 replications: ⁺ $p < 0.1$, ⁺⁺ $p < 0.05$, ⁺⁺⁺ $p < 0.01$.

Table 14: Oaxaca-Blinder Decomposition of Test Score Gap

	Explained				Unexplained			
	Indigenous	Black	Chinese	South Asian	Indigenous	Black	Chinese	South Asian
LICO Ratio	0.0437*** (0.01)	0.0455*** (0.009)	-0.00746 (0.011)	0.0336*** (0.008)	-0.0218 (0.095)	0.0410*** (0.013)	0.284** (0.144)	0.0132 (0.065)
Male	0.00311 (0.004)	-0.00564 (0.008)	-0.00405 (0.006)	0.000316 (0.003)	-0.0888 (0.064)	-0.0605 (0.07)	0.0528 (0.043)	-0.152** (0.067)
Single Parent	-0.0304** (0.012)	-0.0431*** (0.009)	0.00215 (0.004)	0.0125*** (0.004)	0.0433 (0.034)	-0.342*** (0.127)	0.0504 (0.037)	0.0117 (0.030)
Number of Siblings	-0.00755 (0.011)	-0.00796** (0.003)	-0.00616 (0.011)	-0.0219*** (0.008)	0.249*** (0.052)	0.0815* (0.045)	-0.155 (0.119)	0.480*** (0.078)
PCG High School	0.0250** (0.011)	0.0249*** (0.004)	-0.00812** (0.004)	0.00157 (0.011)	0.430*** (0.08)	0.483*** (0.148)	0.292 (0.256)	0.109 (0.230)
PCG Post-Secondary	0.0325*** (0.009)	0.0142*** (0.005)	-0.00294 (0.006)	-0.0143*** (0.004)	-0.000226 (0.06)	0.0808* (0.047)	0.230*** (0.041)	0.0702 (0.108)
Spouse High School	0.0306*** (0.009)	0.0348*** (0.007)	-0.000041 (0.004)	-0.0105*** (0.003)	-0.0735** (0.036)	-0.383* (0.202)	0.134 (0.081)	-0.0438 (0.165)
Spouse Post-Secondary	0.0265** (0.011)	0.0274*** (0.007)	-0.0109** (0.005)	-0.0097 (0.011)	-0.0536 (0.052)	0.064 (0.094)	-0.160** (0.077)	-0.0112 (0.059)
Low Birth Weight	-0.00279 (0.002)	-0.00495*** (0.002)	0.00146 (0.003)	0.0118 (0.007)	0.0202 (0.014)	-0.00774** (0.004)	-0.00808 (0.019)	0.0535* (0.028)
Teenage Mother	0.00450* (0.003)	-0.00346** (0.001)	-0.000152 (0.004)	0.00022 (0.002)	0.0116** (0.005)	-0.00290** (0.001)	0.0396 (0.046)	0.00607 (0.008)
Public School	0.000195 (0.002)	-0.0034 (0.012)	0.000567 (0.004)	0.00271 (0.009)	-0.0585 (0.231)	0.0944 (0.234)	-0.0738 (0.066)	-0.350 (0.399)
Catholic School	0.000205 (0.004)	0.0114 (0.016)	-0.00504 (0.008)	-0.000226 (0.003)	0.0022 (0.071)	0.0797 (0.146)	-0.00493 (0.007)	-0.144 (0.103)
Private School	0.00879** (0.004)	0.00109 (0.005)	-0.017 (0.023)	0.000341 (0.000341)	0.0266 (0.022)	0.0245 (0.037)	-0.0681 (0.056)	-0.0408 (0.035)
Other School	0.000516 (0.001)	-0.000213 (0.000)	0.000437 (0.001)	-0.0000102 (0.000)	0.00459 (0.005)	-0.000342 (0.001)	0.011 (0.011)	-0.0124 (0.014)
Observations	24000	24000	24000	24000	24000	24000	24000	24000

Notes: The dependent variable in each column is the standardized math score. Due to the reporting restrictions associated with the NLSCY, observations have been rounded to the nearest 100. The omitted ethnicity is white, so that all coefficient estimates represent the test score gap relative to white students. Omitted schooling category is whether the school information is missing, and the omitted category for parental education is no formal schooling. All columns include CMA-province and age fixed effects Standard errors clustered by province are in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.