

# PROVINCIAL DIFFERENCES IN HIGH SCHOOL ACHIEVEMENT: FOR WHOM DO SCHOOLS MATTER?

Miles Corak<sup>1</sup>, Darren Lauzon<sup>2</sup>

## ABSTRACT

Studies of regional variation in educational outcomes have focused surprisingly little on the entire distribution of achievement. Estimated effects typically apply to the “average” student. We apply the conditional weighting method of DiNardo, Fortin and Lemieux (DFL, 1996) to assess the contribution of differences in family background and school characteristics to differences in achievement distributions between Canadian provinces. Family background and school characteristics contribute differently in different parts of the distribution and results vary widely by province and skill area. This is a short version of a larger paper available by request from the authors.

KEYWORDS: Provincial Differences, Test Scores, High School, Semiparametrics

## 1. INTRODUCTION

Variation in school outcomes across policy jurisdictions raises important questions about access to quality education and the effectiveness of policy alternatives. Regional differences in both school inputs and the composition of family background factors contribute to these differences and comparisons of average test scores without accounting for both of these contributing factors can lead to misleading inference about the effectiveness of schools (Hanushek and Taylor, 1990).

Differences between Canadian provinces in high school achievement have been observed in all recent, national assessments. Québec and Alberta have tended to outperform other provinces in mathematics. In the Programme for International Student Assessment (PISA), conducted in 2000, the central and western provinces did better than the eastern provinces. Though the comparability of results from different assessments is limited because of differences in content, measurement of achievement, target populations, and participating provinces, the east-west pattern observed in the PISA is generally consistent with other assessments dating back to the mid-1980s.

The relative impact of family background and schooling inputs on observed provincial differences is not well understood. Some emphasize differences in educational policies and practices. Others emphasize differences in family background. Some provinces have higher and more persistent unemployment; others have a greater concentration of immigrants whose mother tongue is neither English nor French. Observed differences in educational policy and practices arise in part because of these differences in student populations, so it is tempting for some to attribute most provincial variation in achievement to variation in student populations.

At first glance, there is some support in the literature for this conclusion. In a widely cited review, Hanushek (1986) shows that school inputs like class size, teacher experience and qualification, per-student funding etc. are consistently either small in magnitude relative to family background variables or statistically insignificant. On this basis, many conclude that “schools don’t matter”. This, however, is counterintuitive given the size of the investment in formal education in industrialized countries, and is in contrast to the large literature that shows schooling inputs tend to be positively related to labour market and other adult outcomes (Card and Krueger, 1992; Haveman and Wolfe, 1995). Some have challenged the accuracy of the Hanushek review (Heges, Laine and Greenwald, 1994). Card and Krueger (1996) suggest that the number of positive results appearing in the literature (despite high

---

<sup>1</sup> Statistics Canada, Family and Labour Studies Division, 24<sup>th</sup> Floor, R.H. Coats Building, Tunney’s Pasture, Ottawa Canada, K1A 0T6. email: miles.corak@statcan.ca

<sup>2</sup> Statistics Canada, Family and Labour Studies Division and Research Data Centres Program, 600-300 West Georgia Street, Vancouver Canada, V6R 6C7. email darren.lauzon@statcan.ca

standard errors) is too large to be due merely to chance. Loeb and Bound (1996, p. 661) point out that “studies finding positive effects of school inputs typically use aggregate data on cohorts educated before 1960, while studies finding no effect tend to use micro-level data on more recent cohorts”. Focusing on the causal impact of a particular educational program, Angrist and Lavy (2001) estimate the impact of a teacher training program in Jerusalem and find that training in secular schools led to higher test scores and that the program was cost-effective.

Most recent studies of school effectiveness, however, focus on the specification and estimation of the education production function. This is the standard empirical framework for evaluating the relative importance of family background and school inputs and the typical specification is a linear, parametric regression model in which family and school inputs are related to a measure of student achievement. The effectiveness of school inputs and policy variables can be evaluated by estimating their marginal effect on outcomes (Hanushek, 1986). Misspecification of the regression model, however, biases estimates of these marginal effects. Measurement error, omitted variables and data aggregation are among the most cited problems. Recent studies making use of better data have attempted to account for these problems, particularly the issue of omitted variables (Goldhaber and Brewer, 1997; Montmarquette and Mahseredjian, 1989). The results for school inputs continue to be weak. Studies in the sociology of education literature using higher-order error components models have shown more positive results for schooling inputs, (Raudenbush and Bryk, 1986; Willms and Raudenbush, 1989; Raudenbush and Willms, 1995) but these models embody strong assumptions about the correlation between omitted variables and those variables included in the model that are not always justified.

Surprisingly few studies have addressed the use of the linear parametric regression framework and its emphasis on average outcomes. School systems do not uniformly impact students. If the role of schools is to bring all students to a minimum standard of achievement, regardless of initial cognitive endowments, the expected impact of schooling inputs would be greater for the least skilled students. For example, smaller class sizes are widely thought to be good school policy. Where smaller sizes have had a positive effect, it has tended to be greatest for disadvantaged students (Lazear, 1999). Two recent papers have used the linear parametric specification in quantile regression (Eide and Showalter (1998); Levin (2001)) to allow for different impacts at different points in the distribution. Eide and Showalter found that school size and per-student funding had positive impacts at the 5<sup>th</sup> percentile (while OLS showed small and insignificant effects for these variables at the mean). Levin, however, found no evidence of an effect for class-size anywhere in the achievement distributions of a sample of Dutch 4<sup>th</sup> and 8<sup>th</sup> graders.

To assess the relative impact of family background and school differences on provincial achievement differences, we take advantage of a recently developed semiparametric approach that avoids the linearity and parametric assumptions made in previous work. We apply the decomposition method of DiNardo, Fortin and Lemieux (DFL, 1996) to differences between Canadian provinces in the entire achievement distribution using data on 15 year olds from the 2000 PISA. The DFL technique extends the familiar Oaxaca (1973) decomposition to decompose differences in entire distributions into components that can be attributable to different sources.<sup>3</sup> The PISA data have the advantage of being both recent and assessing student ability in the three skill areas of reading, mathematics and science. In the present study, we examine differences between Alberta and the rest of the Canadian provinces. Alberta is chosen as the reference case because it performed the best, on average, in all three subject areas. We sequentially decompose differences in the achievement distributions into a component attributable to differences in the distribution of family background characteristics, differences in the distribution of school characteristics and a residual.

Section 2 discusses the DFL decomposition in the context of our data. Section 3 describes the data. Section 4 discusses the results for the differences between New Brunswick and Alberta.<sup>4</sup> Section 5 concludes.

---

<sup>3</sup>For a simple description of the Oaxaca decomposition see Greene (2000, p. 251). See Oaxaca (1973) for the original application.

<sup>4</sup>Complete results for the other provinces compared to Alberta are provided in the full version of our paper, available from the authors by request.

## 2. THE DFL DECOMPOSITION

This section develops the DFL approach, in the context of achievement differences between provinces. Let  $(y, p, x, w)$  be a jointly distributed random vector of test scores, provinces, family background and school covariates respectively. The goal is to compare the difference in marginal test score densities  $f_1(y) - f_0(y)$  between two provinces (labeled 1 and 0) and to decompose this difference into a part attributable to differences between the provinces in the distribution of family background characteristics and a part attributable to differences between the provinces in the distribution of school characteristics. Therefore, we require counterfactual density functions that depict the distribution of test scores in province 1 if the family background or school characteristics were distributed as they are in province 0 and students are otherwise educated as they would be in province 1. The central insight of DFL is that the counterfactual densities are obtained simply by re-weighting the actual density.

With  $z = (x, w)$ , the actual marginal test score for province  $i$  is

$$\begin{aligned}
 f_i(y) &= \int_{z \in \Omega} dF(y, z \mid p_y = p_z = i) \\
 &= \int_{z \in \Omega} f(y \mid z, p_y = i) dF(z \mid p_z = i) \\
 &= f(y; p_y = p_z = i) \\
 &i = 1, 0
 \end{aligned} \tag{1}$$

where  $\Omega$  is the covariate support. This provides the notational convention for expressing for which province the distribution of test scores and covariates is being considered. This marginal density can be estimated by means of the kernel density estimator:

$$\hat{f}_1(y) = \frac{\theta_i}{h} \sum_{i \in p_y=1} K\left(\frac{y_0 - y_i}{h}\right) \equiv \kappa(K, h, \theta_i). \tag{2}$$

The kernel density estimator has been discussed in several recent papers (DiNardo, Fortin and Lemieux, 1996; Blundell and Duncan, 1998; Yatchew, 1998; DiNardo and Tobias (2001)). Here  $\theta_i$  is the sample weight. The function  $K$  is the kernel and gives decreasing weight to points of greater distance from  $y_0$ . The kernel estimator is a generalization of the familiar histogram which can be obtained from (2) with a suitably chosen kernel. Generally, estimates are robust to choices of  $K$  but not to different choices of  $h$ . The tradeoff is one of variance versus bias. If  $h$  is too large, the density will be over-smoothed relative to the true density (bias) and if  $h$  is too small, the true shape of the density will be estimated imprecisely. The choice of  $h$  remains an open subject of research. DFL use the ‘‘plug-in’’ method of Sheather and Jones (1991) as this has been shown to be a better selection in cases of complex, multi-modal densities (Park and Turloch, 1992). Since the underlying plausible value estimates used in this paper are drawn from symmetric models, this is less of a concern with our data. In this study, we use the ‘‘rule-of-thumb’’ estimator suggested by Silverman (1986)  $h = 0.9(\min\{\hat{\sigma}, IQR/1.34\})n^{-1/5}$  where  $\hat{\sigma}$ , is the sample standard deviation,  $IQR$  is the inter-quartile range (the difference between the 75<sup>th</sup> and 25<sup>th</sup> percentiles) and  $n$  is the sample size.

The weights used to obtain the counterfactual densities follow directly from the expression for the actual density. Controlling first for differences in family background characteristics, we can adjust the actual density as follows:

$$\begin{aligned}
 f_1(y; p_y = 1, p_x = 0, p_w = 1) &= \int \int_{x \ z} f(y \mid x, w) dF(x \mid w, p_x = 0) dF(w \mid p_z = 1) \\
 &= \int \int_{x \ z} f(y \mid x, w) \psi_{x|w}(x, w) dF(x \mid w, p_x = 1) dF(w \mid p_z = 1)
 \end{aligned} \tag{3}$$

where

$$\psi_{x|w}(x, w) \equiv \frac{dF(x | w, p_x = 0)}{dF(x | w, p_x = 1)} = \frac{pr(p_x = 0 | x, w)}{pr(p_x = 1 | x, w)} x \frac{pr(p_x = 1 | w)}{pr(p_x = 0 | w)}. \quad (4)$$

The last equality follows from Bayes rule, and the ratios are easily estimated by means of a logit model.<sup>5</sup> With an estimate of  $\widehat{\psi}_{x|w}(x, w)$  in hand, the counterfactual density can be estimated as  $\kappa(K, h, \theta'_i)$  with  $\theta'_i = \theta_i \widehat{\psi}_{x|w}^i(x, w)$ .

The counterfactual density for differences in school characteristics is similarly obtained.

$$\begin{aligned} f_1(y; p_y = 1, p_x = 0, p_w = 1) &= \int \int_x f(y | x, w) dF(x | w, p_x = 0) dF(w | p_z = 0) \\ &= \int \int_x f(y | x, w) \psi_{x|w}(x, w) dF(x | w, p_x = 1) \psi_w(w) dF(w | p_z = 1) \end{aligned} \quad (5)$$

where  $\psi_{x|w}(x, w)$  is defined as before and

$$\psi_w(w) = \frac{dF(w | p_w = 0)}{dF(w | p_w = 1)} = \frac{pr(p_w = 0 | w)}{pr(p_w = 1 | w)} x \frac{pr(p_w = 1)}{pr(p_w = 0)}. \quad (6)$$

Again, the last equality follows from Bayes rule and the ratios can be estimated with logit models. With an estimate  $\widehat{\psi}_w(w)$  we estimate the counterfactual for both family background and school differences as  $\kappa(K, h, \theta''_i)$  with  $\theta''_i = \theta_i \widehat{\psi}_w^i(w)$ .

Finally, simplifying the notation  $\widehat{f}(y; p_y = i, p_x = i, p_w = i)$  to  $\widehat{f}_{iii}$  we decompose the differences in densities as follows:

$$\widehat{f}_{111} - \widehat{f}_{000} = \widehat{f}_{111} - \widehat{f}_{101} + \widehat{f}_{101} - \widehat{f}_{100} + \widehat{f}_{100} - \widehat{f}_{000} \quad (7)$$

Each difference represents, in order, the actual difference, the amount attributable to family background differences, the amount attributable to school differences, and a residual. The results are sensitive to the order of decomposition so we also compute the decompositions in reverse order (considering school differences first, then student differences).

### 3. DATA

The PISA assessment was administered in participating OECD countries in the Spring of 2000. Pisa employed a two-stage design, with schools sampled in the first stage and students within schools in the second. The Canadian sample contained about 30,000 students, large enough for provincial-level estimates. Students wrote a two hour test containing reading, mathematics and science items. The students and principals of the participating schools completed questionnaires designed to collect information about the schools, the students' family background and other factors related to achievement.<sup>6</sup> The achievement data we use are the set of five plausible value "test scores".

<sup>5</sup> Though DFL used a probit, we used the logit for computational convenience and because the average of the predicted success probabilities is the sample mean proportion if the model contains a constant term.

<sup>6</sup> In Canada, PISA was also integrated with the Youth in Transition Survey (YITS) a new longitudinal survey. Students and their parents completed questionnaires for this survey. At the time of this study, only the results from the student surveys were available.

Under certain conditions plausible value scoring yields unbiased estimates of aggregate statistics that are functions of unknown ability (in this case, ability in mathematics, reading and science). They are particularly advantageous in tests of relatively small numbers of items.<sup>7</sup>

Though considerable information is collected from students in both the PISA and YITS instruments, we concentrate on background factors that are most easily treated as exogenous to the education system. For students, then, we consider the following groups of variables:

1. Foreign-born status of parents and students
2. Language (whether student speaks the language of testing at home)
3. Single parent status
4. Education level of both parents
5. Occupation of both parents
6. Labour force status of both parents at the time of the test.

For schools, we focus attention on variables that have been traditionally used in education productivity studies as well as those that show considerable variation across provinces and represent important areas of public debate. The school variable groupings are:

1. Class-size (approximated by the student-teacher ratio)
2. How student evaluation is used by the school (to group students according to ability, to benchmark against national or provincial results, or to monitor performance year to year.).
3. Indicators of teacher morale
4. Hours per year of instructional time.

The actual definition of variables used in the final decomposition estimates was limited to a large extent by the demands of identifiability in the logit models used to estimate the re-weighting functions.<sup>8</sup>

## 4. RESULTS

In this paper we present the results for New Brunswick and Alberta in the reading and mathematics assessments. Complete results are reported in the larger version of the paper. Figure 1 depicts the actual densities for each of the Canadian provinces relative to Alberta. Overlaid on each panel is an indicator function that takes a non-zero value if the difference in estimated densities at that point is significantly different from zero at the 95 percent level. The standard errors for the density estimates are obtained using balanced repeated replication weights provided on the data set and adjusted for imputation variance associated with the plausible values. The provinces are depicted in east-west order from left to right on the graph. Significant differences from Alberta are greatest in the eastern provinces. This is consistent with the results for mean achievement. In particular, New Brunswick and Alberta had the largest mean gap.

---

<sup>7</sup>See Miselevey (1991) for more on plausible values.

<sup>8</sup>We had more dummy variables defined to capture more of the shape of the distribution of family background and school characteristics. In many cases, however, logit parameter could not be accurately estimated because some these variables perfectly predicted being in a particular province. This was the case for several of the eastern provinces where variation in these variables was in some cases limited. The final list of variables is a set that could be estimated for all of the provinces.

Figure 1. Reading Achievement Densities with Respect to Alberta

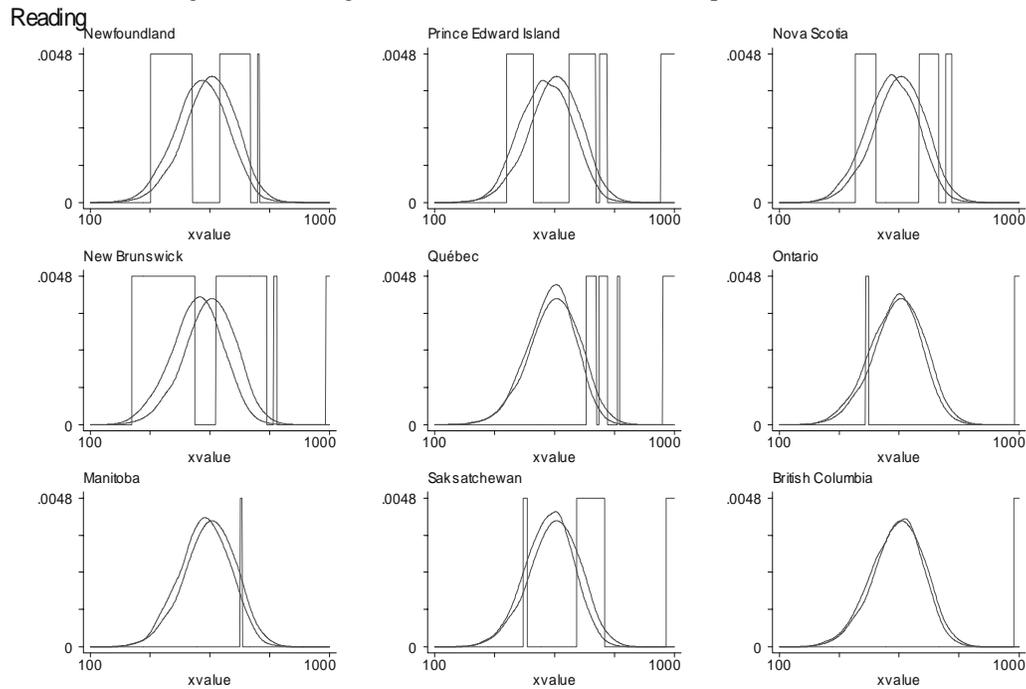
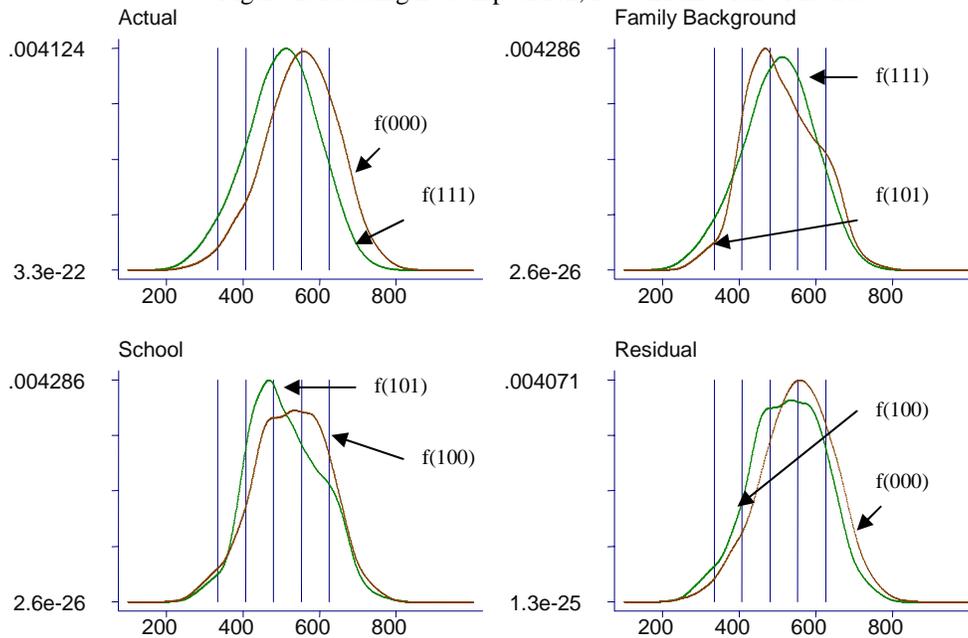


Figure 2 shows the primary order decomposition (family background first, then schools) for the reading assessment. Because reading was the focus subject in the 2001 PISA, proficiency cutoff values have been defined for the released data set that identify skill levels 1 through 5 with 5 and greater representing the highest performing students and below level 1 identifying the least proficient. These are indicated by vertical lines. The first panel depicts the actual densities for both provinces. Clearly the Alberta density is shifted to the right of New Brunswick, indicating that there are higher proportions of Alberta students performing at the higher proficiency levels.

The remaining panels show the results of the decomposition. The second panel shows the effect of controlling for family background. Family background differences contributed to achievement gaps among the lowest and highest performing students. Fewer students would perform below level 2 and more students would perform above level 5 in the absence of these differences. There would be no rightward shift in the distribution, though, and the extra density would be observed in the middle of the distribution. In other words differences in family background tend to decrease the achievement differences in roughly the 2<sup>nd</sup> quarter of the distribution.

When we further control for differences in the distribution of school characteristics, we observe a slight rightward shift of the New Brunswick test score distribution. Many fewer students would perform between levels 1 and 2 and between levels 2 and 3, meaning that after controlling for family background differences, school differences still contribute to gaps in these parts of the distribution. More students would perform between levels 4 and 5 and above level 5, meaning that even after controlling for family background, school differences contribute to achievement differences in these parts of the distribution also. The last panel shows the residual effect, that is, what is left to “explain” in the difference in distributions. Accounting for differences between Alberta and New Brunswick in the distribution of school and background student characteristics in the manner we have defined them does not eliminate the difference in test score distributions. Alberta’s distribution remains shifted to the right of the adjusted New Brunswick distribution with fewer students performing below level 3 and more students performing above level 4.

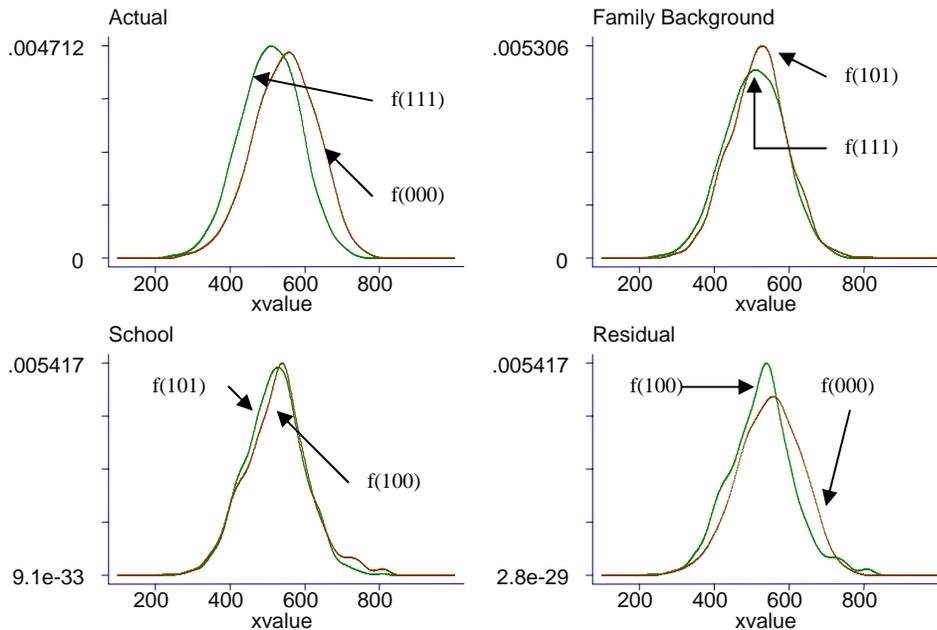
Figure 2. Reading Decomposition, New Brunswick - Alberta



The results differ in the case of mathematics (figure 3). Again, the Alberta density is shifted to the right of the New Brunswick density. Differences in family background explain a small part of the difference in the achievement of lower performing students but do not explain much of the difference in the upper portion of the distribution, in contrast to reading. The third panel indicates that further controlling for differences in school factors makes a small difference in the middle and upper tail of the test score density, implying school differences drive achievement differences in the upper and lower percentiles of the math distribution. The effects are less pronounced than for the reading assessment but are similar in where the impact is observed. Panel 4 indicates that except at the very upper percentiles of the distribution, Alberta's distribution remains to the right of New Brunswick's after controlling for differences in students and schools.

For the reading and math assessments, we can see that differences in student factors seemed to explain differences in the low end of the distribution with perhaps a small effect in the upper tail as well. Once controlling for student differences, controlling further for differences in school characteristics tended to have an equalizing effect, with their main impact concentrated in the middle parts of the test score distribution. One cannot easily argue, at least in the case of reading, that schools do not have an impact on the achievement distribution, even after controlling for differences in the student population.

Figure 3. Mathematics Decomposition



Tables 2 and 3 show the decompositions for a variety of location and shape statistics for reading and math. The first column of the table gives the actual value of the statistic for New Brunswick using New Brunswick’s estimated density. The second gives the actual difference using the estimated actual densities. The decompositions follow, in sequence, in columns 3, 4 and 5. Beneath each effect is the proportion of the gap accounted for. Referring back to equation (7), and denoting the estimated statistic from density  $\hat{f}_{iii}$  as  $\hat{\theta}_{iii}$ , the values in columns 3, 4 and 5 are respectively  $(\hat{\theta}_{111} - \hat{\theta}_{101})$ ,  $(\hat{\theta}_{101} - \hat{\theta}_{100})$  and  $(\hat{\theta}_{100} - \hat{\theta}_{000})$ .

Focusing first on the provincial mean test score, differences in family background contribute to the gap in reading and mathematics mean achievement (in their absence, the gap would be smaller). School differences also contribute to the difference in means and contribute more than family background to the reading gap, but less to the mathematics gap.

At the median results differ. Family background differences *reduce* the median reading gap (in their absence, the New Brunswick mean would be higher). School differences contribute considerably more to the median reading gap than the mean gap, much less than for mathematics. The residuals for the two tests are similar and comparable to the residual for the mean.

Focusing on the reading proficiency levels, family background explains over 80 percent of the gap in the proportion of students performing at or below level 1 proficiency. School differences partially reduce the gap, suggesting that New Brunswick schools have partially adjusted to the needs of the lowest achievers and that, controlling for student differences, Alberta’s school characteristics would not help the lowest reading achievers. Schools do contribute considerably to the gap in the proportions of students achieving between levels 1 and 2 and 2 and 3. Schools contribute greatly to the gap in the proportion of students performing between levels 4 and 5 and contribute to a lesser extent to the gap in the proportion of students performing above level 5.

For math we lack proficiency levels like those for reading. Instead, we examine the impact of family background and school differences on selected percentiles. Alberta percentiles are higher, given that Alberta’s math distribution lies to the right of New Brunswick’s. Consistent with figure 2, student differences contribute greatest to the differences in the lowest percentiles and the contribution falls almost monotonically with higher percentiles. Residuals for the percentile decompositions are high, indicating that less than half of the differences in the percentiles are accounted for by the variables used in the study.

The reverse order decompositions produce similar results. Graphs and tables are available in the complete paper.

TABLE 1. Decomposition Differences in Selected Statistics, Reading

	Value	Actual Difference	Student Factors	School Factors	Residual Factors
Mean	501.1505	-49.2325	-8.6927	-15.6718	-24.8679
		1	0.1766	0.3183	0.5051
St. Dev.	97.4912	-0.6386	2.5605	-3.0469	-0.1521
		1	-4.0094	4.7712	0.2382
C.V.	5.1405	-0.4683	-0.2302	0.0071	-0.2451
		1	0.4916	-0.0151	0.5234
Median	505.7878	-48.5754	4.6203	-27.3159	-25.8798
		1	-0.0951	0.5623	0.5328

**Proportions at Proficiency  
Level**

<= Level 1	5.2511	3.3279	2.7633	-0.7098	1.2745
		1	0.8303	-0.2133	0.383
1 and 2	11.7482	5.5905	1.0871	2.3024	2.201
		1	0.1945	0.4118	0.3937
2 and 3	22.8826	8.1339	-5.2425	7.1669	6.2095
		1	-0.6445	0.8811	0.7634
3 and 4	29.2961	2.7914	3.0564	-0.2815	0.0165
		1	1.0949	-0.1009	0.0059
4 and 5	21.0009	-6.7837	1.8994	-5.9667	-2.7163
		1	-0.28	0.8796	0.4004
> 5	9.8211	-13.06	-3.5637	-2.5112	-6.9851
		1	0.2729	0.1923	0.5349

TABLE 2: Decomposition of Differences in Selected Statistics, Mathematics

	Value	Actual Difference	Student Factors	School Factors	Residual Factors
Mean	506.1981	-40.7671	-11.936	-7.8101	-21.0211
		1	0.2928	0.1916	0.5156
St. Dev.	82.3808	-4.5545	2.4538	-8.7671	1.7588
		1	-0.5388	1.9249	-0.3862
C.V.	6.1446	-0.147	-0.338	0.5527	-0.3618
		1	2.2989	-3.7595	2.4607
90 <sup>th</sup>	610.9828	-47.9425	-9.8536	-12.7514	-25.3375
		1	0.2055	0.266	0.5285
80 <sup>th</sup>	576.192	-45.8409	-7.6417	-8.8287	-29.3705
		1	0.1667	0.1926	0.6407
75 <sup>th</sup>	565.0002	-43.3102	-5.9733	-7.5989	-29.7381
		1	0.1379	0.1755	0.6866
Median	508.3962	-42.5621	-12.5851	-6.1272	-23.8498
		1	0.2957	0.144	0.5604
25 <sup>th</sup>	451.7622	-37.7836	-13.3949	-4.6952	-19.6935
		1	0.3545	0.1243	0.5212
20 <sup>th</sup>	436.1152	-38.144	-15.5595	-2.2576	-20.3268
		1	0.4079	0.0592	0.5329
10 <sup>th</sup>	399.0307	-35.0127	-16.7098	1.8912	-20.1941
		1	0.4772	-0.054	0.5768

## 5. CONCLUSION

We examine the contribution of family background and school differences to differences in the achievement distributions of Canadian 15 year olds. Our approach differs from the standard empirical framework in that we directly estimate achievement distributions and counterfactual distributions that would be observed if differences in family background and school characteristics were eliminated. We find that even after controlling for family background, school differences matter in explaining differences between New Brunswick and Alberta, the widest achievement gap in the 200 PISA assessment. However, schools had different effects in different parts of the achievement distribution, and the patterns differed between the reading and mathematics assessments. These results might vary from conclusions based on a parametric regression. In the case of the mean math gap, such an analysis might conclude that family background factors contribute more to the New Brunswick - Alberta mathematics gap. Our results show that schools contributed more to the mathematics gap in the higher percentiles, though the difference is small.

DiNardo, Fortin and Lemieux did not discuss standard errors of the density estimates nor inference about the size of the effects. For our data, we were able to make standard error estimates using BRR weights provided on the data. Donald, Green and Paarsch (2000) provide an alternative estimator of distributions that is based on results from the hazard rate literature. Their approach admits large numbers of covariates and they provide a means of estimating standard errors.

There was no teacher survey in PISA, unlike other educational assessments, so good teacher data is lacking. We have not investigated the robustness of the decomposition results to changes in the mix of explanatory variables used. In our data, the method of re-weighting kernels has limitations in homogenous samples because the number of explanatory variables, and hence the detail with which student and school distributions can be measured, is lacking due to problems of over-identification. Notwithstanding these limitations, our results suggest that factors that contribute to differences in mean achievement do not always contribute equally to differences elsewhere in the achievement distribution for different populations and skill assessments. Thus studies that focus on differences in mean performance miss potentially important effects.

## REFERENCES

- Angrist, J. D. and V. Lavy (2001) "Does Teacher Training Effect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools" *Journal of Labour Economics* 19(2):343-369
- Blundell, R. and A. Duncan (1998) "Kernel Regression in Empirical Microeconomics" *The Journal of Human Resources* 33(1) 62-87
- Card, D. (1996) "School Resources and Student Outcomes: An Overview of the Literature and New Evidence From North and South Carolina", *Journal of Economic Perspectives*, 10(4):31-50
- Card, D. and A. B. Krueger (1992) "Does School Quality Matter? Returns to Education and the Characteristics of American Public Schools?" *The Journal of Political Economy* 100(1):1-40
- Donald, S. G., D. A. Green and H. J. Paarsch (2000) "Differences in Wage Distributions Between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates" *Review of Economic Studies* 67, 609-633
- DiNardo, J. and Tobias, J. L. (2001) "Nonparametric Density and Regression Estimation" *Journal of Economic Perspectives* 15(4):11-28
- DiNardo, J.; Fortin, Nicole M. and Lemieux, Thomas (1996) "Labour Market Institutions and the Distribution of Wages 1973-1992: A Semiparametric Approach" *Econometrica* 64(5):1001-1044
- Goldhaber, Dan D. and Brewer, Dominic J. (1997) "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity" *Journal of Human Resources* 32(3):505-523

- Greene, William H. (2000) *Econometric Analysis* 4<sup>th</sup> Edition, Prentice Hall, Upper Saddle River, N.J.
- Hanushek, Eric (1986) "The Economics of Schooling: Production and Efficiency in the Public Schools." *Journal of Economic Literature* 24(3):141-78
- Hanushek, Eric and Lori L. Taylor (1990) "Alternative Assessments of the Performance of Schools: Measurement of State Variation in Achievement: *Journal of Human Resources* 25(2):179-201
- Mislevy, R. J. (1991) "Randomization based inference about examinees in the estimation of item parameters" *Psychometrika* 56, 177-196
- Montmarquette, Claude and Mahseredjian, Sophie (1989) "Does School Matter for Achievement? A Two-Way Nested Error Components Analysis" *Journal of Applied Econometrics* 4(2):181-93
- Oaxaca, R. (1973) "Male Female Wage Differentials in Urban Labour Markets, *International Economics Review* 24:693-708
- Park, B. U and B. A. Turloch (1992) "Practical Performance of Several Data Driven Bandwidth Selectors" *Computational Statistics* 7:251-270
- Raudenbush, Stephen W. and Anthony S. Bryk (1986) "A Hierarchical Model For Studying School Effects" *Sociology of Education* 59(1):1-17
- Raudenbush, Stephen W. and Willms, J. Douglas (1995) "The Estimation of School Effects" *Journal of Educational and Behavioural Statistics* 20(4):307-35
- Sheather, S. and M. Jones (1991) "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation" *Journal of the Royal Statistical Society, B*, 53:683-690
- Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, London, Chapman and Hall.
- Yatchew, Adonis (1998) "Nonparametric Regression Techniques in Economics" *Journal of Economic Literature* 36 669-721
- Willms, J. Douglas (1996) "Indicators of Mathematics Achievement in Canadian Elementary Schools" in *Growing Up in Canada*, Minister of Industry, Ottawa
- Willms, J. Douglas and Stephen Raudenbusch (1989) "A Longitudinal Hierarchical Linear Model for Estimating School Effects and Their Stability" *Journal of Educational Measurement* 26(3):209-32