Sorting and Inequality in Canadian Schools*

Jane Friesen
Simon Fraser University
Brian Krauth
Simon Fraser University

March 5, 2007

Abstract

A student's peers are often thought to influence his or her educational outcomes. If so, an unequal distribution of advantaged and disadvantaged students across schools ("sorting") in a community will amplify existing inequalities. This paper explores the relationship between the degree of sorting across schools within a community and educational inequality as measured by the variance of standardized high school exam scores within the community. Cross-sectional OLS estimates suggest that the variance of test scores is related to sorting by ethnicity, but not to sorting by income or parental education. We then implement two strategies for addressing endogeneity in the degree of sorting: a standard unobserved effects (first-difference) approach, and a first-difference/instrumental variables approach in which the structure of school choice (number and relative size of schools) is used to construct instruments for the degree of sorting. The results from both approaches indicate that the variance of test scores is related to sorting by home language and parental education, but not to sorting by income. Our results also suggest that reducing sorting would have little effect on inequality of outcomes in the typical Alberta community, but would have substantial effects in the larger cities.

JEL codes: H75, I20
Keywords: Peer effects, school choice, education

1 Introduction

Since the 1966 publication of the "Coleman Report" (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld and York 1966), social scientists and policymakers have been concerned with the interaction of peer effects and segregation/sorting in maintaining or worsening economic inequality. Several recent studies (Boozer and Cacciola 2001, Ding and Lehrer 2007, Hanushek, Kain, Markman and Rivkin 2003, Hoxby 2000) provide evidence that student achievement as measured by performance on standardized tests is positively affected by the achievement level of classmates and/or schoolmates. This

*The authors are grateful for helpful comments from Stephen Jones, David Johnson, Susan Mowbray and seminar audiences at the CERF Conference on Education, Schooling and the Labour Market, Simon Fraser University and the TARGET Labour Workshop at the University of British Columbia, as well as two anonymous referees. Hayley Laker provided excellent research assistance, and funding was provided by the SFU/SSHRC Small Research Grant program. We also thank Peter Cowley for providing the data and for technical support. All errors are ours.

finding implies that an unequal distribution of high-resource (e.g., high income, high parental education, etc.) students across schools leaves most low-resource students twice disadvantaged: such students will have both lower private resource levels and a lower quality peer group. As this difference in human capital leads to differences in income, initial disparities in resources may be maintained and even amplified over generations despite the best efforts of policymakers to provide equal opportunity (Durlauf 1996, Fernandez and Rogerson 1996). Although empirical researchers have made some progress on establishing the existence and magnitude of peer effects in education, much remains to be learned. In particular, little is known about the overall effect of school sorting on inequality in educational outcomes.

This paper uses data from the Canadian province of Alberta to evaluate the hypothesis that sorting exacerbates inequality in student outcomes. Our research design uses a simple regression framework to determine the empirical relationship between community-level measures of sorting across schools by income, ethnicity, and other characteristics and the community's variance in standardized test scores at the grade 9 level. The use of aggregate data avoids a number of the identification problems that arise in using individual-level data to estimate peer effects (Manski 1993). In particular, our research design does not require the assumption of exogenous selection into schools, nor does it require the absence of idiosyncratic aggregate factors. Furthermore, by focusing directly on the relationship between sorting and inequality, we are able to directly quantify the portion of inequality within a given community that can be explained by sorting.

While our research design may also be applicable to U.S. states and other Canadian provinces, Alberta has a key advantage over many of these jurisdictions in that standardized testing is nearly universal. Almost all private schools in Alberta, including religious schools, receive direct per-pupil funding from the provincial government, and are subject to provincial testing and reporting requirements. In contrast, statewide standardized testing in the U.S. is generally limited to public school students, due in part to constitutional restrictions on both state authority over private education and state funding of religious schools. At the same time, a large portion of school sorting in the U.S. takes place between the private and public sectors rather than within the public sector. As a result, much of the variation in school composition and educational outcomes is censored in the data. For example, suppose high-income students tend to stay in homogeneous high-income public schools but leave more mixed-income public schools for private schools. Then censoring can produce a positive statistical relationship between measured sorting and measured inequality even in the absence of such a relationship between actual sorting and actual inequality. By considering a jurisdiction in which both public and private school students are tested, we avoid this difficult issue.

Our research makes a number of unique and constructive contributions to the literature on peer effects, sorting and inequality. First, we explore sorting on a wider variety of dimensions than previous studies, including ethnicity, family income, primary parent's education, and the language spoken at home. Second, we are able to link school-level exam results in different grades and years, and thus to construct a first-difference estimator that controls for community-level unobserved effects. For example, if greater unobserved heterogeneity among students within a community *causes* sorting, simple cross-sectional estimators will be biased in favor of finding a large effect of sorting on inequality. Third, we construct instrumental variables for sorting based on cross-community variation in the ability of families to sort across schools due to variation in the structure of school choice. These instruments are incorporated into the first-difference model to further weaken the necessary exogeneity assumptions.

Our cross-sectional results suggest that sorting by ethnicity is associated with increased variance in test scores, even after controlling for sorting by several other household characteristics. However, the first-difference results show that this pattern is not robust to controlling for unobserved heterogeneity within communities. Instead, the first-difference results suggest that sorting by parental education and language spoken in the home increases the variance of test scores. Neither the cross-section nor first-difference estimators produce evidence of a sorting effect arising from family wage and salary income. Finally, the first-difference estimates using instrumental variables are consistent with the findings of the basic first-difference analysis.

Finally, we use our estimates to explore the magnitudes of potential effects from different types of desegregation policies. The results imply that fully integrating the schools of the typical multiple school community will have little effect on community inequality, but integrating the schools in Alberta's major metropolitan areas may have a large effect.

1.1 Related literature

This paper shares both motivation and many econometric issues with the literature on peer effects and other social interaction effects. The key econometric concern in this literature is to address at least some of the identification problems noted by Manski (1993). Most early research on peer effects simply included the average of either behavior or background characteristics among the survey respondent's peer group as a supposedly exogenous explanatory variable in a simple regression for the respondent's behavior. The coefficient on the average peer variable would then be interpreted as a measure of the peer effect. Manski notes four major problems with this methodology. The first problem ("sorting") is that self-selection into groups may lead to positive within-group correlation in many characteristics that are relevant to the outcome, and some of these relevant characteristics may be unobserved. For example, low-income students who attend private schools or public schools in high income neighborhoods may have parents who particularly value education. The second problem ("common shocks") is that there may be unobserved aggregate influences on a particular group; for example one school may have a particularly effective or ineffective principal, which will tend to raise or lower all scores in a given school. The third problem ("simultaneity") is that the respondent influences his or her peers in the same way the peers influence the respondent. Each of these three problems implies that average peer behavior and/or characteristics will fail to be exogenous, and leads the reduced form coefficient to be an upwardly biased estimate of the true peer effect. The fourth problem in this approach is that even in the absence of the first three problems, there is no obvious way to distinguish empirically between the influence of peer characteristics ("contextual peer effects") and the influence of peer behavior ("endogenous peer effects") without imposing strong functional form assumptions on the econometric model. A sizeable empirical literature has pursued various strategies for identifying social effects while addressing the issues raised by Manski. Although a thorough survey of this literature is beyond the scope of this paper (see Moffitt 2001 for such a discussion), several recent studies of peer effects in education are particularly relevant to the approach developed here.

One class of studies is characterized by the use of natural experiments and other sources of credibly exogenous variation in peer group composition. This exogenous variation can then be used to instrument for peer group composition in a standard individual-level regression. Hoxby (2000) uses idiosyncratic variation in gender and racial composition between different cohorts at the same school to identify peer effects in test scores for Texas grade 3-6 students. Hanushek, et al. (2003) use a matched panel of Texas grade 3-8 students and schools to estimate peer effects while allowing for student and school-grade fixed effects. Boozer and Cacciola (2001) exploit exogenous variation in grades 1-3 classroom composition due to Tennessee's "Project STAR" class size experiment. Ding and Lehrer (2007) use data from a province in China that assigns students to schools on the basis of (observable) test scores. All four of these studies find evidence of nontrivial peer effects.

Another class of studies is characterized by the use of aggregate data to look for evidence of what are known as social multipliers. This approach, also known as the "excess variance" approach, starts from the insight that social interaction effects will tend to amplify the relationship between group composition and group outcomes. Glaeser, Sacerdote, and Scheinkman (2003) estimate individual-level and aggregate-level regressions, and measure the social multiplier (a monotonic but nonlinear function of the peer effect) by the ratio of aggregate-level coefficients to individual-level coefficients. Glaeser, Sacerdote, and Scheinkman (1996) show that random group assignment implies a specific relationship between group size and (unconditional) between-group heterogeneity, and allows for identification of the social interaction effect from group-level data on size and variance in outcomes. Graham (2005) extends this idea to conditional between-group heterogeneity, using a model that is similar to ours and data from the Project STAR class-size experiment. In this experiment, students and teachers were randomly assigned to classes of varying size. Graham notes that under random assignment small classes will have more variation in average student ability than large classes but will not have more variation in teacher quality, and uses this to develop a GMM estimator of peer effects that exploits the implied relationship between class size and variability of outcome. The key difference between our approach and Graham's is that his identification is based on the properties of random assignment, while ours is based on self-selection. While our approach requires more restrictive assumptions on the nature of unobserved heterogeneity and identifies a more limited set of parameters, it has the advantage of being applicable in cases where random assignment does not occur.

By providing information on the magnitude of peer effects, both the natural experiment and excess variance approaches can be informative about the relationship between sorting and inequality. At the same time, these studies have limitations. First, those group characteristics for which there is exogenous variation are not necessarily those whose distribution can be changed by public policy. For example, a number of costly policies to reduce sorting such as the Moving To Opportunity program (Katz, Kling and Liebman 2001) are aimed specifically at reducing residential sorting by income. However, none of the studies referenced above speaks directly to the effect of peer family income on academic outcomes, simply because their data lack credibly exogenous sources of variation in peer family income. A second limitation is that estimates of peer effects by themselves tell us little about the relationship between sorting and inequality. Positive school-level peer effects imply a positive relationship, but further quantitative analysis along the lines of Kremer (1997) is needed to say whether sorting explains a large or small proportion of the variation in test scores across students. Kremer develops a calibration methodology to estimate the long-run impact on inequality of what had been previously described as a large increase in marital and residential sorting. Despite calibrating his model with seemingly large social interaction effects, Kremer finds that the observed increase in sorting has a minimal long run effect on inequality. This finding suggests that there is a clear need for direct quantitative evidence on the relationship between sorting and inequality.

Our research uses cross-community data on sorting of students into schools based on their family characteristics to fill this gap in the literature. The essential feature of the cross-community (or cross-city) approach is that it does not require the assumption of random assignment; indeed it exploits the fact that families sort. This feature enables us to estimate the impact of sorting along a much wider and potentially more policy-relevant set of dimensions than are addressed in peer effect studies, including income, parental education, and other characteristics. A second benefit of the cross-community approach is that it will capture effects of sorting on inequality that arise through mechanisms other than peer effects. For example, suppose that political processes operate to move resources from poor schools to rich schools. In that case, a community in which schools are more strongly sorted on income may have a more unequal distribution of resources across schools, and thus a greater dispersion in student outcomes, even in the absence of peer effects. In some sense, this "contaminates" our results if the primary goal is to estimate peer effects; however, it is an advantage if the goal is to estimate the effect of sorting on inequality.

Related applications of the cross-community approach include work by Cutler and Glaeser (1997) and Card and Rothstein (2004) on the effects of racial segregation in U.S. cities. Cutler and Glaeser characterize segregation by a dissimilarity index calculated across the census tracts in a city, and include both this measure of segregation and an interaction term for segregation and race in regressions for various individual outcomes. Card and Rothstein characterize segregation by the black-white gap in percentage of black schoolmates, and include this measure of segregation along with measures of black-white gaps in various socioeconomic characteristics in a regression for the black-white SAT score gap in the city. Our work uses a similar research design to these studies, while considering sorting on a wider variety of characteristics.

2 Methodology

The overall methodology in the paper involves estimating the coefficients of a simple linear regression model using community-level data. The dependent variable is the variance in exam scores across the students in the community, and the explanatory variables include measures of the variance in background characteristics of the students' families, as well as measures of the degree of sorting across the community's schools with respect to these same characteristics. This basic regression is estimated using both cross-sectional and first-difference approaches, and the estimated coefficients on the sorting variables are interpreted as providing information on the effect of sorting on inequality. To motivate this approach, and discuss some of the econometric and conceptual issues, this section introduces a simple model of peer effects and sorting.

2.1 Aggregation and notation

Because we are working with aggregate variables at two different levels of aggregation (school and community) and at different grades, the notation is unavoidably somewhat complex. Communities are indexed by c. We have data for a single cohort of students within each community. The number of schools in grade g of community c is given by S_{cg}, and schools within that community are indexed by $s = 1, 2, \ldots, S_{cg}$. Let N_{scg} be the number of students taking the grade-g exam in school s, and let $N_{cg} = \sum_{s=1}^{S_{cg}} N_{scg}$ be the total number of students taking the exam in community c.

Let x_{iscg} be an arbitrary individual-level random variable describing an individual i in grade g of school s of community c. Let $\hat{E}[.|.]$ be the average operator, e.g. so that $\hat{E}[x_{iscg}|c, g]$ is the average value of x_{iscg} across individuals in grade g within community c. The within-school average and variance of x are given by:

$$\bar{x}_{scg} \equiv \hat{E}[x_{iscg}|s, c, g] = \frac{1}{N_{scg}} \sum_{i=1}^{N_{scg}} x_{iscg} \tag{1}$$

$$\hat{\sigma}^2_{scg}(x) \equiv \hat{E}[(x_{iscg} - \bar{x}_{scg})^2|s, c, g] = \frac{1}{N_{scg}} \sum_{i=1}^{N_{scg}} (x_{iscg} - \bar{x}_{scg})^2 \tag{2}$$

Note that because the data describe the entire population of exam takers in each community, we do not use any degrees-of-freedom correction for the variance. The within-community average and variance are given by:

$$\bar{x}_{cg} \equiv \hat{E}[x_{iscg}|c, g] = \sum_{s=1}^{S_{cg}} \frac{N_{scg}}{N_{cg}} \bar{x}_{scg} \tag{3}$$

$$\hat{\sigma}^2_{cg}(x) \equiv \hat{E}[(x_{iscg} - \bar{x}_{scg})^2|c, t] = \sum_{s=1}^{S_{cg}} \frac{N_{scg}}{N_{cg}} \hat{\sigma}^2_{scg}(x) + \sum_{s=1}^{S_{cg}} \frac{N_{scg}}{N_{cg}} \left(\bar{x}_{scg} - \bar{x}_{cg}\right)^2 \tag{4}$$

Note that the overall within-community variance can be decomposed, as in the standard ANOVA decomposition, into a term that reflects the average within-school variation and a term that reflects between-school variation in averages. This second term, which we will call the community's cross-school variance in x, will be employed as our measure of sorting:

$$\widehat{CSV}_{cg}(x) \equiv \hat{E}\left[\left(\bar{x}_{scg} - \bar{x}_{cg}\right)^2 |c, t\right] = \sum_{s=1}^{S_{cg}} \frac{N_{scg}}{N_{cg}} \left(\bar{x}_{scg} - \bar{x}_{cg}\right)^2 \tag{5}$$

The cross-school variance is the natural measure of sorting under the model described in the next section. The extreme cases of perfect sorting and perfect integration provide upper and lower bounds on the cross-school variance of a given characteristic. Under perfect sorting (each school in community c is perfectly homogeneous, i.e., $\hat{\sigma}^2_{scg}(x) = 0$) it will be the case that $\widehat{CSV}_{cg}(x) = \hat{\sigma}^2_{cg}(x)$. Under perfect integration (all of the schools in community c have the same value of $\bar{x}_{scg}$), it will be the case that $\widehat{CSV}_{cg}(x) = 0$.

2.2 A simple model of sorting and peer effects

Next we outline a simple linear model of peer effects in academic performance. Let y_{iscg} be the exam score of an individual i attending school s in community c in grade g, and suppose that:

$$y_{iscg} = z_{iscg} + \lambda \bar{z}_{scg} \tag{6}$$

where z_{iscg} is the student's (unobserved) ability, and $\bar{z}_{scg}$ is average (unobserved) ability within the student's school. "Ability" is meant here as simply a catch-all term for any private inputs into student performance: it could include innate talent, prior educational experience, parental inputs, and anything other than the qualities and/or performance of the current peer group. In the terminology of Manski (1993), the peer effect in the model above is of the "contextual" variety, i.e., it operates through average peer characteristics $\bar{z}_{scg}$. Endogenous peer effects, which operate through average peer outcome, could also be included in the model but would yield the same reduced form as in equation (6) and so are omitted for convenience. There are a number of different mechanisms that could generate peer effects, including:

- Simple externalities/spillovers: students may learn from one another and thus learn more when in contact with high-achieving or high-ability peers.
- Provision of public goods: parents with high endowments of time, money, or skills may supply public goods to their child's classrooms.
- Scarcity of instructor's time: students with behavioral or learning problems may take instruction time or energy away from classmates.
- Political economy/resource considerations: parents with high interest or resource levels may divert resources towards their child's school.
- Specialization: homogeneous classes may facilitate efficiency through specialization or reduced discrimination.

As with most of the literature, we will not be able to distinguish between these different mechanisms, nor between endogenous and contextual effects. Finally, the model presented here does not include other sources of school-level heterogeneity such as variations in teacher or principal quality that are not associated with student ability. Section 2.4 describes an extension to the model which allows for additional school-level heterogeneity.

This model is a standard one in the literature on peer effects, and researchers have made numerous attempts to estimate this type of model using various proxy variables for student ability. However, as discussed previously, the model parameters are not identified from standard microeconomic data when schools are sorted on ability (Manski 1993), because this sorting induces positive correlation between the unobserved components of the student's ability (z_{iscg}) and the observed characteristics of the student's peers (some observable predictor of $\bar{z}_{scg}$).

Rather than use microeconomic data to estimate peer effects, this study uses aggregate data to directly estimate the effect of sorting on inequality. To do so, we aggregate the model, and decompose the within-community variance in test scores into a portion that is due to within-community variance in private inputs, and a portion that is due to sorting. Aggregating equation (6) to the school and community level, average scores are given by:

$$\begin{aligned} \bar{y}_{scg} &= (1+\lambda)\bar{z}_{scg} \\ \bar{y}_{cg} &= (1+\lambda)\bar{z}_{cg} \end{aligned} \tag{7}$$

The within-community variance in test scores is given by:

$$\hat{\sigma}^2_{cg}(y) = \hat{E}\left[(y_{iscg} - \bar{y}_{cg})^2 \,|c,t\right] \tag{8}$$

Substituting in (6) and (7), and applying algebra we get:

$$\hat{\sigma}^2_{cg}(y) = \hat{\sigma}^2_{cg}(z) + \left(\lambda^2 + 2\lambda\right)\widehat{CSV}_{cg}(z) \tag{9}$$

Equation (9) shows how the within-community variance in student outcomes[1] can be broken down into a portion that is explained by within-community variance in private inputs, and a portion that is explained by sorting. This decomposition is analogous but not identical to the standard ANOVA decomposition. As equation (9) indicates, sorting will affect the overall variance in outcomes when λ is nonzero. In the absence of peer effects, only the variation in private inputs across individuals matters for the variation in test scores.

2.3 Econometric Implementation

Note that equation (9) is based on an unobserved index of student-specific inputs, and so cannot be estimated directly. Instead, our estimating equations will be based on using proxy variables for both

[1] In this application the student outcome is the level of a test score. An alternative approach would use a value-added model in which current inputs determine the test score gain rather than the test score level. In that case equation (9) would still apply, but with the outcome y being the test score gain. Because we do not have access to student-level longitudinal data, we cannot estimate value-added models.

the cross-individual variation in inputs ($\hat{\sigma}^2_{cg}(z)$), and for the extent of sorting ($\widehat{CSV}_{cg}(z)$).

We assume that we have a vector of observable community/grade-level variables X_{cg} such that:

$$\hat{\sigma}^2_{cg}(z) = X_{cg}\alpha + u_{cg} \tag{10}$$

and a vector of observable community/grade-level variables W_{cg} such that:

$$\widehat{CSV}_{cg}(z) = W_{cg}\delta + v_{cg} \tag{11}$$

where u_{cg} and v_{cg} have statistical properties (to be outlined below) sufficient to identify interesting features of the model. In other words, X contains proxy variables for the variability in private inputs across families in the community. These proxy variables include the total number of students writing the exam, as well as the measured variance across households in Aboriginal status, visible minority status, home language, wage and salary income, government transfer income, and education of primary parent. The vector W contains proxy variables for the cross-school variation in inputs, in particular the cross-school variance in measured household characteristics.

Combining equations (9), (10), and (11), we get the reduced form:

$$\hat{\sigma}^2_{cg}(y) = X_{cg}\alpha + W_{cg}\beta + \epsilon_{cg} \tag{12}$$

where $\beta = \delta(\lambda^2 + 2\lambda)$ and $\epsilon_{cg} = u_{cg} + v_{cg}\left(\lambda^2 + 2\lambda\right)$. Interpretation of the reduced form coefficients is straightforward: the marginal effect of sorting on inequality is given by β. Note that when there is no peer effect ($\lambda = 0$), the true value of the reduced form coefficient is $\beta = 0$. We pursue three strategies for estimating the coefficients in (12): simple cross-sectional regressions, first-difference regressions that allow for a community-specific unobserved effect, and first-difference regressions that also use instrumental variables to isolate one plausibly exogenous source of variation in measured sorting.

2.3.1 Cross-section regressions

Our starting approach estimates an OLS regression using the cross-sectional grade 9 data, i.e. a regression of $\hat{\sigma}^2_{c9}(y)$ on (X_{c9}, W_{c9}). As usual, this regression will produce consistent estimates of β under the assumption that the explanatory variables are exogenous:

$$E(\epsilon_{c9}|X_{c9}, W_{c9}) = 0 \tag{13}$$

Assumption (13) holds if sorting on unobserved dimensions is unrelated to sorting on observed dimensions (i.e., $E(v_{c9}|X_{c9}, W_{c9}) = 0$) and the unobserved variation in heterogeneity across communities is unrelated to the degree of sorting (i.e., $E(u_{c9}|X_{c9}, W_{c9}) = 0$). The first of these two assumptions is relatively innocuous, as the consequences of its violation (a researcher mistakenly attributing the effects of sorting on income to sorting on ethnicity, for example) are relatively minor. In addition, its violation does not imply violation of (13) under the null hypothesis of no peer effects ($\lambda = 0$), and so will not invalidate standard tests of that null. The second assumption, that the unobserved variation in heterogeneity is unrelated to measured sorting, is much stronger. It is not difficult to imagine greater unobserved heterogeneity within a community leading affluent families to take greater steps to segregate from disadvantaged families, so we also follow some slightly more elaborate estimation approaches.

2.3.2 First-difference regressions

In order to relax the assumption that there is no relationship between unobserved variation in community heterogeneity and variation in measured sorting, we use linked grade 9/grade 6 data to develop a simple first-difference estimator. We replace (13) with the weaker assumption of a community unobserved effect:

$$\epsilon_{cg} = a_c + w_{cg} \tag{14}$$

and the assumption that the explanatory variables are strictly exogenous conditional on the community unobserved effect:

$$E(w_{cg}|X_{c6}, X_{c9}, W_{c6}, W_{c9}, a_c) = 0 \qquad \text{for } g = 6, 9 \tag{15}$$

Given these assumptions, we can estimate (by OLS or other means) the first-difference regression:

$$\Delta\hat{\sigma}_c^2(y) \quad = \quad \Delta X_c \alpha + \Delta W_c \beta + \Delta w_c \tag{16}$$

where, for example $\Delta X_c \equiv (X_{c9} - X_{c6})$. The model coefficients are identified by variations in sorting between grades 6 and 9. As will be discussed in Section 3, the primary source of variation in sorting between these two grades is the reallocation of students from a larger number of primary schools in grade 6 to a smaller number of secondary schools in grade 9.

While equation (16) can be estimated by OLS, we instead follow Arellano and Bover (1995) and use GMM based on the moment condition:

$$E\left(\left(\Delta\hat{\sigma}_c^2(y) - \Delta X_c\alpha - \Delta W_c\beta\right)\left[\begin{array}{cccc} X'_{c6} & X'_{c9} & W'_{c6} & W'_{c9} \end{array}\right]'\right) = 0 \tag{17}$$

This method explicitly includes the overidentifying restrictions implied by the strict exogeneity condition (15), and provides a straightforward means of testing those restrictions.

2.3.3 First-difference regressions with instrumental variables

The simple cross-sectional approach described in Section 2.3.1 requires exogeneity in the level of sorting. The first-difference approach described in Section 2.3.2 allows for endogeneity in the level of sorting, but requires that changes in sorting between grades 6 and 9 are exogenous. Our third estimator is also based on an unobserved effects specification but further relaxes the identifying assumptions by considering a particular and more plausibly exogenous source of variation in sorting: the structure of school availability in the community.

Changes in sorting between different grades in a given community can be thought of as driven by two main factors: changes in the private incentive to sort, and changes in the ability to sort. While the incentive for families to sort is driven by a number of unobserved variables including any unobserved variation in school quality, the ability to sort is driven primarily by the school structure of the community, i.e., the number and relative size of available schools. The positive relationship between the number of options (schools and/or districts) available to families and their ability to sort has been the subject of several papers on the school choice debate, with Urquiola (2005) being a recent example. To take an extreme but empirically relevant example of this relationship, a community with only one high school will have a grade 9 cross-school variance of exactly zero for all variables. A community with multiple elementary schools but only one high school necessarily will show a decline in sorting from grade 6 to grade 9. This variation in school structure is more plausibly exogenous than most other sources of variation in sorting, and so might provide a suitable instrumental variable.

Constructing such an instrument is complicated by the difficulty of characterizing the school structure of a community in a suitable manner. The sorting effect can be consistently estimated using simply the number of schools in the community as an instrument. However, the relationship between the number of schools and the ability of families to sort is nonlinear: the opening of an additional school has a bigger effect on the potential to sort when there is initially only one school than when there are several hundred. In addition, there is relevant information in the size of schools: the potential effect of an additional school on the ability to sort is larger if the school itself is relatively large. In order to construct the most informative possible instrument, we calculate the highest possible CSV for each characteristic given the community's school structure and the underlying distribution of the variable in question across the community's households. This community-level variable, which we will denote by $max\widehat{CSV}_{cg}(x)$, incorporates information on both the number and size of schools, and accounts for the potentially nonlinear effect of the number of schools on the ability to sort. In addition $max\widehat{CSV}_{cg}(x)$ has some theoretically-relevant content, as it is exactly the CSV predicted by the early theoretical work on community sorting under educational peer effects (Durlauf 1996, Fernandez and Rogerson 1996) that treats the family locational choice as strictly student-outcome-maximizing subject to income constraints. Subsequent work (Epple and Sieg 1999, for example) allows for cross-family heterogeneity on more than one relevant dimension and implies sorting below the maximal level. However, the equilibrium level of sorting in these models remains closely related to the maximal level.

Letting $\tilde{W}_{cg}$ be the vector of instrumental variables constructed by replacing each $\widehat{CSV}_{cg}(x)$ in W_{cg} by its corresponding maximum value $max\widehat{CSV}_{cg}(x)$, our identifying assumption (15) can then be replaced by:

$$E(w_{cg}|X_{c6}, X_{c9}, \tilde{W}_{c6}, \tilde{W}_{c9}, a_c) = 0 \qquad \text{for } g = 6, 9 \tag{18}$$

In other words, the amount of sorting is no longer required to be strictly exogenous, only that portion of sorting that is explained by the ability to sort. The model is estimated using GMM based on the moment condition:

$$E\left(\left(\Delta\hat{\sigma}_c^2(y) - \Delta X_c\alpha - \Delta W_c\beta\right)\left[\begin{array}{cccc} X'_{c6} & X'_{c9} & \tilde{W}'_{c6} & \tilde{W}'_{c9} \end{array}\right]'\right) = 0 \tag{19}$$

We also evaluate the relevance of each instrument by reporting the F statistic from the "first-stage" regression. That is, we report the F statistic for the null hypothesis that the coefficient on $\Delta\tilde{W}_c$ is zero in an OLS regression of ΔW_c on $(\Delta X_{c6}, \Delta\tilde{W}_c)$. In the case of multiple endogenous variables, we report the Cragg-Donald F statistic (Cragg and Donald 1993).

2.4 A comment on heterogeneity in school quality

The model described above treats all heterogeneity as being student-based. However, unobserved heterogeneity in school quality may also play a role. Incorporating this heterogeneity directly into the model adds some complications, but it is important to understand the potential implications of such heterogeneity. This section adds school-level heterogeneity to the model, considers what assumptions must be made about this heterogeneity in order to identify the effect of sorting, and discusses the likely direction of bias if those assumptions fail to hold.

We start by adding an unobserved school-level term b_{scg} to equation (6)

$$y_{iscg} = b_{scg} + z_{iscg} + \lambda \bar{z}_{scg} \tag{20}$$

Aggregating, we get:

$$\begin{aligned} \bar{y}_{scg} &= b_{scg} + (1+\lambda)\bar{z}_{scg} \\ \bar{y}_{cg} &= \bar{b}_{cg} + (1+\lambda)\bar{z}_{cg} \end{aligned} \tag{21}$$

The within-community variance in test scores is given by:

$$\hat{\sigma}^2_{cg}(y) = \hat{E}\left[(y_{iscg} - \bar{y}_{cg})^2 \,|c,t\right] \tag{22}$$

Substituting in (20) and (21), and applying algebra we get:

$$\begin{aligned} \hat{\sigma}^2_{cg}(y) &= \hat{\sigma}^2_{cg}(b) + \hat{\sigma}^2_{cg}(z) + \left(\lambda^2 + 2\lambda\right)\widehat{CSV}_{cg}(z) + (2+2\lambda)\hat{E}\left[(b_{scg} - \bar{b}_{cg})(\bar{z}_{scg} - \bar{z}_{cg})\right] \\ &= \hat{\sigma}^2_{cg}(b) + \hat{\sigma}^2_{cg}(z) + \left(\lambda^2 + 2\lambda\right)\widehat{CSV}_{cg}(z) + (2+2\lambda)\hat{\rho}_{cg}\left(b_{scg}, \bar{z}_{scg}\right)\sqrt{\hat{\sigma}^2_{cg}(b) + \widehat{CSV}_{cg}(z)} \end{aligned} \tag{23}$$

where $\hat{\rho}_{cg}\left(a,b\right)$ is the community-level correlation between two arbitrary variables a and b. Equation (23) decomposes the cross-student variation in outcome into four components: simple variation in school quality ($\hat{\sigma}^2_{cg}(b)$), simple cross-student variation in private resources ($\hat{\sigma}^2_{cg}(z)$), and two sorting effects. One sorting effect operates through peer effects ($\left(\lambda^2 + 2\lambda\right)\widehat{CSV}_{cg}(z)$), and the other operates through the interaction of sorting with quality variation ($(2+2\lambda)\hat{\rho}_{cg}\left(b_{scg}, \bar{z}_{scg}\right)\sqrt{\hat{\sigma}^2_{cg}(b) + \widehat{CSV}_{cg}(z)}$).

In order for the regression results reported in Section 4 to consistently estimate something interesting about equation (23), we need to consider two issues. First, we need to interpret the entire term $\left(\lambda^2 + 2\lambda\right)\widehat{CSV}_{cg}(z) + (2+2\lambda)\hat{\rho}_{cg}\left(b_{scg}, \bar{z}_{scg}\right)\sqrt{\hat{\sigma}^2_{cg}(b) + \widehat{CSV}_{cg}(z)}$ as representing the "sorting effect." That is, this term is zero if sorting is absent, and positive if it is present. However, this term is no longer linear in $\widehat{CSV}_{cg}(z)$ and depends on random unobserved factors, so our regression needs to be considered a constant-coefficient linear approximation to the true nonlinear model.

Second, the level of sorting in a given community-grade needs to be unrelated to the cross-school variation in quality, at least after adjusting for community-level unobserved effects and controlling for any available proxy variables for the cross-school variation in quality. Specifically, let:

$$\hat{\sigma}^2_{cg}(b) = \gamma B_{cg} + \delta_{cg} \tag{24}$$

where B_{cg} is a vector of proxy variables for cross-school quality variation. In the regressions estimated in this paper, the relevant proxy is the total number of students[2] in the community. The additional identifying assumption in the cross-sectional case is:

$$E(\delta_{c9}|B_{c9}, X_{c9}, W_{c9}) = 0 \tag{25}$$

In the first-difference case, it is:

$$\begin{aligned} \delta_{cg} &= d_c + e_{cg} \\ E(e_{cg}|B_{c6}, B_{c9}, X_{c6}, X_{c9}, W_{c6}, W_{c9}, a_c, d_c) &= 0 \end{aligned} \tag{26}$$

and in the first-difference case with instrumental variables:

$$\begin{aligned} \delta_{cg} &= d_c + e_{cg} \\ E(e_{cg}|B_{c6}, B_{c9}, X_{c6}, X_{c9}, \tilde{W}_{c6}, \tilde{W}_{c9}, a_c, d_c) &= 0 \end{aligned} \tag{27}$$

Although each of these is a distinct condition, the common feature is fairly straightforward: once proxy variables are controlled for and once any unobserved effect has been differenced out, the within-community variation in school quality is unrelated to the amount of sorting. If instead, greater within-community variation in school quality led to increased sorting, this would tend to bias our estimates of sorting effects upwards.

3 Data

3.1 An overview of Alberta's Provincial examination program

Alberta is a large province in western Canada with an economy based primarily on the oil and gas industry. The population is concentrated in two large metropolitan areas (Calgary and Edmonton), several smaller cities and a large number of small agricultural or resource industry towns. Approximately 547,000 students attended Alberta's K-12 public school system in the 2000-2001 school year (Interprovincial Education Statistics Project 2002). Although it is administered mostly by local school boards, education is funded by the province and therefore per pupil funding formulas are the same in

[2]Instead of the total number of students, we could use the number of schools in the community and possibly the average size of schools in the community. We have estimated our model with these variables included, and find that including them does not affect our main results.

all public schools. Although a number of students attend private schools, both the size and degree of independence of the private educational sector are less than in the United States. In particular, 92% of the 192 private schools in 2001 were classified as "accredited" and therefore received provincial funding and were subject to provincial standards and curriculum requirements. The remaining private schools were classified as "registered" and did not receive provincial funding. Alberta reports that 4% of its students attended private schools in 2001, and such schools received per-student funding at approximately 60% of the per-student public school funding (Alberta Learning 2002a, Sections 8.2 and 8.4).

The Achievement Tests are conducted in grades 3, 6, and 9. The exams are in the core academic subjects of Language Arts, Mathematics, Science, and Social Studies, and are intended to be taken by all students. Both accredited and registered private schools participate in these exams, as do public schools. The dependent variable in our regressions is the community-level variance of grade 9 test scores in Language Arts and Mathematics. Because the exams are in principle mandatory, and because there are almost no dropouts at this level, the selection issues involved in this estimation are relatively minor. However, a small fraction of students do not take the exams. Approximately 4% to 7% of grade 9 students are absent from school on the exam day and fail to take a make-up exam. In addition, a student may be excused from the exam if his or her principal judges that he or she is "incapable of responding to the exam," or if "participation will be harmful to the student." Approximately 3% to 6% of students are exempted in this manner (Alberta Learning 2002b).

3.2 Data structure and links with data on family characteristics

The data set was originally assembled by Cowley and Easton (2002) in conjunction with Statistics Canada and the Alberta provincial government. The primary unit of observation in the Cowley and Easton data set is the school. For each exam and each school, the data set provides the mean and variance of exam scores for the school, as well as the average number of students taking the exam and other school-level variables. If five or fewer students at a given school take a given exam in a given year, results are suppressed.

In addition, the Cowley and Easton data set features detailed Census data on the family characteristics of the school population. These authors obtained from Statistics Canada a custom 1996 Census data set describing the population of households with one or more children of secondary school age, organized by Enumeration Area (EA). This data set includes means and variances of a wide variety of background characteristics, including family income from various sources, ethnicity, language spoken at home, and primary parent's education. The Alberta provincial government provides a breakdown of each school's population of enrolled students by EA of residence. The Census EA-level data are then aggregated to the school level using weights based on the number of students in each EA. The resulting data set thus approximates the characteristics of the families who have children attending each school.

3.3 Aggregation to community level

Although the Cowley and Easton data set is at the school level, the relevant unit of observation in this study is the group of schools in a community. For the purposes of this study, we define a community as a collection of one or more municipalities that form an integrated labor market and host at least one school offering grade 9. Two municipalities are considered to be part of the same community if they are fewer than 30 minutes driving time apart and at least one municipality has a population over 10,000. This definition implies that, for example, the city of Calgary and its suburbs are treated as part of the same community, but two farming towns nearby to one another are treated as two distinct communities. We distinguish between small and large municipalities in this way because many of Alberta's small municipalities are agricultural centers, where family residential choices are dictated by the location of farmland, rather than residential amenities such as schools. Driving time is calculated by using the quickest route provided by the *Microsoft Streets and Trips 2005* computer program (Microsoft Corporation 2004). Once communities are defined, the original school level data are aggregated to the community level using weights based on the number of students taking the exam in each school.

This method for defining communities generates 213 distinct communities in Alberta in 2001. For each community, the 2001 grade 9 data are also linked to the community's grade 6 results from the year (1998) when most of the 2001 grade 9 class would have been in grade 6. The assignment of grade 6 schools to grade 9 communities is slightly complicated by some cases where a very small municipality has a local elementary school but sends its students to a high school in another municipality. For example, grade 6 students attending Dunstable School in the town of Busby generally attend Barrhead Composite in Barrhead for grade 9, and so are assigned in our data to the community of Barrhead. For each such grade 6 school, we contacted the school directly to identify the high school that students in grade 6 in 1998 would have been expected to attend in 2001.

3.4 Construction of instrumental variables

The maximum CSV variable is constructed as follows. Let x_{iscg} be an arbitrary individual-level variable drawn from a community-specific frequency distribution with CDF F_c. Let $STRUCT_{cg} \equiv \{N_{1cg}, N_{2cg}, \ldots, N_{(S_{cg})cg})$ be the school structure in grade g of community c. The maximum CSV

$(max\widehat{CSV}_{cg}(x))$ of grade g in community c is defined as the highest value of $\widehat{CSV}(x)$ that could be attained given $(F_c, STRUCT_{cg})$.

Calculation of $max\widehat{CSV}_{cg}(x)$ turns out to be relatively straightforward. The school structure is directly observed in the data. In those cases where the characteristic x is binary, the community-wide distribution F_c can be calculated directly from the data. In cases where x is continuous, we assume a normal distribution for x with mean and variance as given in the data. Given $STRUCT_{gc}$ and F_c, we can calculate $max\widehat{CSV}(x)$ by a simple procedure. Any CSV-maximizing allocation of students to schools must be perfectly stratified, i.e., if a given school has a student with $x = a$ and a student with $x = b$, then it must have all students with $x \in (a, b)$. For $S_{cg} \leq 6$, we simply try out all $S_{cg}!$ perfectly stratified allocations and find the allocation with the highest CSV. For $S_{cg} > 6$ we choose the allocation with the highest CSV among 1,000 perfectly stratified allocations that have been generated randomly.

4 Results

4.1 Descriptive Statistics

Table 1 describes the number of schools at each grade level in the 213 communities in our data. Many of these communities are small, with a single school at each grade level. A substantial number of communities have two schools at one or both grade levels. The two large metropolitan areas of Calgary and Edmonton have over 200 schools offering grade 6 and over 100 schools offering grade 9. Because all communities are weighted equally, our regression results will be largely determined by the many small communities rather than the few large ones.

Table 2 characterizes the data in terms of the difference between the number of grade 6 and grade 9 schools within communities. As discussed in Section 2.3.3, the changing number of schools will play a key role in generating changes in measured sorting. In Alberta, grade 6 students generally attend an elementary (grade 1-7 or 1-6) school, while grade 9 students attend high schools (grade 8-12) or junior high schools (grade 7-9). Elementary schools are generally smaller than high schools so multiple elementary schools will often “feed” a given high school. This reallocation can in principle generate large changes in sorting. The number of grade 6 schools in the community is equal to the number of grade 9 schools in 162 of these communities, and the number of grade 6 schools is higher in 49 communities. In 23 of these 49 communities, students attending two or more schools for grade 6 are merged into a single grade 9 school. These communities experience the largest change in sorting, and so play a key role in identifying our unobserved effects models. Two communities had more schools offering grade 9 than schools offering grade 6.

Table 3 reports community-level summary statistics for both exam scores and family characteristics. Note that the mean of community-level means, as reported in Table 3, is not identical to the provincial mean across individuals because it weighs small and large communities equally. Table 3 shows that the communities in our data exhibit substantial variation in the average exam performance of their students, and that some communities have much more variation in exam performance than others. In both Language Arts and Mathematics, the cross-school variation in test scores is a small proportion of the overall within-community variation in test scores.

The family characteristics considered in this study can be divided into two groups: characteristics related to ethnicity and culture, and characteristics related to socioeconomic status. Each of the three ethno-cultural variables – visible minority status, Aboriginal status, and whether the household primarily speaks a language other than English or French at home – measures a different way in which the student and his or her family might be outside of the majority culture. Aboriginal students in particular have faced substantial and well-documented difficulties in the Canadian school system (British Columbia Ministry of Education 2002). Although only 8.1 percent of the average community in our data is Aboriginal, many small communities include substantial numbers of Aboriginal families: the proportion of families with school age children that are headed by self-reported Aboriginal people in communities in our data ranges from 0 to 98 percent. "Visible minority" households represent a smaller group (Aboriginal families are not counted as visible minorities in the Census), about 3 percent of the average community. Visible minorities are also concentrated in particular communities: the proportion of visible minority families in our communities ranges from zero to over 44 percent. Over one-third of those with visible minority status in the average community report their ethnicity as Chinese. The variable "official language not usually spoken at home" means that a household speaks a language other than English or French (Canada's two official languages) at home. About 4 percent of students in the average community live in such homes: the proportion of homes with school-age children where neither official language is spoken ranges from 0 to 64 percent across communities.

Our indicators of socioeconomic status are family wage and salary income, family government transfer income, and the level of education of the primary parent. Income may affect learning directly if parents with greater financial means are able to provide more resources in support of a child's learning, such as tutoring services. More generally, income may affect health and self-esteem, which in turn influence both learning and behavior. However, it is not obvious that we would expect the relationship between family income and student learning to conform to the linear functional form that underlies our empirical specification. As a result, we also include the level of government transfer income as a proxy for low-income status. Because within-province variation in the generosity of transfer programs is virtually nonexistent, the level of transfer income will be closely and positively related to low-income status.[3] Our third indicator of socioeconomic status, primary parent's education, has been shown elsewhere to be an important predictor of individual student academic performance (see, for example, Hanushek and Raymond 2004). The average community in the data has a mean family income of just over $40,000 per year, mean transfer income of about $4,500 per year, and a mean level of primary parent's education of a few months past high school. All three of these variables exhibit substantial variation across communities, as do their variances.

One complication in measuring the effect of sorting on all six of these dimensions is their relatively high correlation. Table 4 shows the correlation matrix for the sorting variables, with Pearson (linear) correlations reported in the upper triangle and Spearman (rank) correlations in the lower triangle. As the table shows, most of the sorting variables are sufficiently collinear that it might be difficult with limited data to independently separate out their effects. Even in those cases where the sorting variables are not substantially collinear, the Spearman correlations indicate a strong and positive nonlinear relationship between the sorting variables.

4.2 Regression results: Cross section

Tables 5-7 report estimates for the cross-sectional model without a community-level unobserved effect. The specifications vary only in terms of which sorting variables are included in the model. The dependent variable in each regression is the community-level variance in grade 9 Language Arts or Mathematics exam scores, scaled so that the dependent variable has unit standard deviation across communities. In addition to the sorting variables, all specifications include a common set of explanatory variables that includes the average number of students per school, a quadratic term in the number of schools in the community, and the community-level variance in each of the six household characteristics. Heteroskedasticity-robust standard errors are used to calculate asymptotic t-statistics and p-values. The regressions themselves are estimated by weighting each community equally, rather than by population. We report the coefficients on the variance and CSV variables in standardized units. That is, the coefficient of 0.21 on "Var(W/S income)" in the first column of Table 5 means that a one-standard-deviation increase in the variance of wage and salary income is associated with a 0.21 standard deviation increase in the variance of grade 9 test scores. We leave the coefficient on the "total number of students" variable unstandardized because its unit of measurement is straightforward.

[3] Government transfer income includes all transfer payments received from federal, provincial or municipal governments. This variable is the sum of the amounts reported as Old Age Security and Guaranteed Income Supplements, Canada Pension Plan, Employment Insurance, Canada Child Tax benefits, and "other income from government sources."

That is, the coefficient of -0.06 on "(# writing)/1000" in the first column of Table 5 means that a 1000-student increase in the number of students in a community is associated with a 0.06 standard deviation decrease in the variance of grade 9 test scores.

The first column in each of the Language Arts and Math panels includes no sorting variables at all. In these regressions, the variance in Language Arts exam scores is increasing in the variance of wage and salary income and decreasing in the variance of parental education, while the remaining variables have statistically insignificant coefficients. The variance in Mathematics exam scores is increasing in the variance in wage/salary income, and the variance in transfer income, and is decreasing in the variance of parental education.

The remainder of Table 5 presents results of various specifications that include sorting by ethnicity and language, while Table 6 presents specifications that include sorting by socioeconomic status. Because of the high collinearity in the sorting variables and the close relationship between them (any policy which affects sorting on education will also affect sorting on income), we follow the strategy of estimating the model with each sorting variable on its own in Tables 5 and 6, then estimating the model with all sorting variables in Table 7. The results in these tables suggest a strong and positive cross-sectional relationship between sorting on Aboriginal status and variance in Language Arts exam scores, whether or not we control for other forms of sorting. They also suggest a positive relationship between sorting on visible minority status and variance in both-subject exam scores, though this relationship is not always statistically significant. Other relationships between sorting and the variance in test scores – positive between sorting on Aboriginal status and variance in Mathematics scores, negative between sorting on wage/salary income and variance in both Language Arts and Mathematics scores – appear only in the Table 7 regression with all sorting variables and not in the one-variable-at-a-time regressions in Tables 5 and 6.

The sorting effects in the regressions reported in Tables 5-7 are identified from cross-community variation in sorting, and can be given causal interpretation only if sorting is not correlated with unobserved community-level effects that play a role in determining the overall variance of test scores. Because this is a strong assumption, we next pursue two alternative identification strategies that exploit more credibly exogenous sources of variation in the degree of sorting within communities.

4.3 Regression results: First-difference

Tables 8-10 report estimates for the model with a community-specific unobserved effect, as described in Section 2.3.2. These estimates are constructed by linking the 2001 grade 9 test results to the 1998 grade 6 test results from the same community, and estimating a regression of the change in the variance of test scores on the change in measured sorting and other variables. Note that the community-level variance variables do not change, and so their coefficients are not identified. Coefficients are reported using the same scaling as in the cross-sectional results in Tables 5-7. As with the cross-sectional results, the model is estimated weighting all communities equally, and the reported asymptotic t-statistics and p-values are robust to heteroskedasticity. In addition to coefficient estimates each table reports the Hansen (1982) J statistic for the overidentifying restrictions. As the tables show, the overidentifying restrictions fail to be rejected at any conventional significance level in any of the specifications. As a further robustness check, we also estimated an OLS regression of the change in test score variance on the change in sorting and the levels of the variance variables. Doing so had little effect on the results, and an F test of the coefficients on these level variables was unable to reject the null that these levels all have coefficients of zero.

Tables 8 and 9 show that when the sorting variables are included one at a time, sorting on primary parent's education and on home language are associated with increased variance in scores on both exams, and these associations are statistically significant. Sorting on Aboriginal status has a positive and statistically significant association with the variance in Language Arts exam scores, but not with Mathematics exam scores. Table 10 shows that the positive association between sorting on Aboriginal status and variance in Language Arts exam scores remains when other sorting variables are included, as does the positive association between sorting on parental education and variance in both Language Arts and Mathematics exam scores. The association between sorting on home language and variance in Language Arts exam scores becomes statistically insignificant when other sorting variables are included. Surprisingly, sorting on government transfer income is negatively and significantly associated with variance in exam scores when other sorting variables are included, though its association is near-zero when included on its own in Table 9.

Next we compare the first-difference results reported in Tables 8-10 to the cross-section results reported in Tables 5-7. Some patterns are consistent across the two approaches: sorting on Aboriginal status is positively associated with variance in Language Arts scores, but has a much weaker association with variance in Mathematics scores. In both sets of estimates, sorting on wage and salary income seems not to have a positive association with variance in exam scores for either subject. However, many parameter estimates differ substantially across the cross-section and first-difference results. These differences may reflect several factors: point estimates may differ because social interactions have a different effect at different grade levels or because of community level fixed effects; statistical significance may differ because variables exhibit greater independent variation in the cross-section or the first-differenced data.

4.4 Regression results: First-difference with instrumental variables

Tables 11-13 show the results from estimating the first-difference model using the change in maximum CSV induced by changes in school structure as an instrument for the change in actual CSV. This approach, described in Section 2.3.3, isolates changes in sorting that occur due to changes in the ability of families to sort (as multiple local elementary schools feed into a single high school). These changes may be more plausibly exogenous than changes in sorting that occur due to changes in the incentive of families to sort.

In addition to coefficient estimates, these tables report test statistics for the overidentifying restrictions and for instrument relevance. The overidentifying restrictions are tested using the Hansen (1982) J statistic. As the tables show, these restrictions cannot be rejected at any conventional confidence level in any of the reported specifications. Instrument relevance is tested in Tables 11 and 12 using the "first-stage" F statistic, i.e., the F statistic for the joint null hypothesis of zero coefficients on all excluded instruments. The regressions in Table 13 feature multiple endogenous explanatory variables, so we instead report the Cragg-Donald (1993) F statistic . Note that unlike all other test statistics reported in this paper, the Cragg-Donald F statistic is not robust to heteroskedasticity. We can interpret the first-stage or Cragg-Donald F statistics both as indicators of both instrument relevance and instrument strength. The F statistic is significant at the 5% level in all but one of the regressions, indicating that the instruments are relevant and the model is identified. Instrument strength can be assessed for the regressions in Tables 11 and 12 by applying the Staiger and Stock (1997) rule of thumb that an instrument can be considered weak if the first-stage F statistic is less than ten. By this rule of thumb, we can reject the hypothesis of a weak instrument in 4 of the 12 regressions in Tables 11 and 12. For Table 13, we are able to reject the null of nonidentification as well. However, the test statistic is quite low, suggesting that identification in this particular specification is quite weak and the results in Table 13 should be interpreted with more caution than those in Tables 11 and 12.

The coefficient estimates in Tables 11 and 12 are roughly consistent with those found using the simple first-difference method of Tables 8 and 9. Sorting on Aboriginal status is positively and significantly associated with the variance in Language Arts exam scores, and sorting on parental education is positively and significantly associated with the variance in exam scores for both subjects. The point estimates of the association between sorting on home language and exam scores are actually larger than in the simple first-difference estimates, but are now statistically insignificant due to the loss of precision from IV estimation. Overall, each of the 95% asymptotic confidence intervals for each coefficient estimated in Tables 11 and 12 overlaps the corresponding confidence interval in Table 8 or 9.

4.5 How Big Are the Estimated Sorting Effects?

Next we consider how to evaluate the magnitude of the estimated sorting effects. To do so, we consider a simple counterfactual. Suppose that we took a given multi-school community and completely eliminated cross-school sorting on a particular family characteristic. By how much, in percentage terms, do our models predict the variance in exam scores would decrease?

Table 14 provides the answer to these questions using all three estimation schemes. The sorting variables considered are the three variables that frequently appeared with statistically significant coefficients: Aboriginal status, home language, and parental education. The table reports the predicted change in variance from the elimination of sorting for three communities: the major cities of Edmonton and Calgary, as well as a hypothetical community whose CSV is equal to the average CSV among multi-school communities. The predicted changes are calculated using the coefficients from those models that include only one sorting variable at a time. This is because it is unclear what it would mean, for example, to substantially reduce sorting on parental education while keeping sorting on income unchanged.

As the table shows, changes to sorting would have limited effect in a typical Alberta community, reducing the variance in test scores by no more than 5.4%. However, changes in sorting would potentially have large effects in more heterogeneous major cities. Sorting on Aboriginal status is fairly low in both Calgary and Edmonton, and so its elimination would only reduce the variance in test scores in those communities by a few percentage points. Sorting on home language and on parental education are both much higher in Edmonton and Calgary than in the typical Alberta community. As a result, our model estimates predict that elimination of sorting in these communities would lead to quite large reductions in the variance of test scores. Using the first-difference and IV estimates as our base, the model predicts that the elimination of sorting on home language in Edmonton schools will reduce the variance in Language Arts exam scores by 18-45% and the variance in Mathematics exam scores by 19-22%. Elimination of sorting on parental education in Edmonton is predicted to reduce the variance in Language Arts exam scores by 31-48% and the variance in Mathematics exam scores by 8-12%. Similar results appear for Calgary, though the estimated effect of eliminating sorting on home language is somewhat lower.

5 Conclusion

The results reported in this paper suggest that certain types of sorting across schools are associated with increased community-level variance in academic outcomes. This evidence contributes to the literature on educational peer effects, sorting, and inequality in two ways. First, it provides information on educational peer effects with respect to contextual variables such as family income and parental education that are often unavailable in the student-level administrative data usually seen in this literature. Second, it provides measures of the magnitude of sorting's effect on inequality in student outcomes, and thus quantifies the potential for integration of schools on income, etc. to reduce current levels of inequality in student outcomes.

Our cross-sectional estimates indicate a sorting effect that might be associated directly with ethnicity – specifically Aboriginal status – rather than other characteristics such as family wage and salary income, government transfer income, primary parent's education and language spoken in the home. This result suggests that factors such as differences in cultural norms or discrimination may be generating peer effects within school communities. However, endogenous sorting may be contaminating these results if communities that are more diverse in unobserved ways are more likely to sort themselves along ethnic/cultural lines.

Our first-difference estimates suggest associations between inequality and sorting on several family characteristics. Communities in which the degree of sorting by home language falls as students move from elementary to high school exhibit a greater-than-average reduction in inequality in both Language Arts and Mathematics test scores between grades 6 and 9, as do communities in which the degree of sorting by primary parent's education falls. These results seem intuitively plausible. Parental education may serve as a proxy for other characteristics that might affect peers, such as the degree of parental provision of public goods or political influence, or it may be correlated with ability and therefore picking up simple externalities. Language spoken in the home may also serve as a proxy for public goods or political influence. Furthermore, students from homes where English is not spoken may themselves have relatively weak English language skills. The presence of students in the classroom with weak language skills may alter teaching methods in ways that affect their fellow students' progress.

We also estimate a variation on the first-difference model that uses changes in community school structure – the number and size of schools – to generate plausibly exogenous instrumental variables for changes in sorting. This approach yields similar results to those found for the basic first-difference models: point estimates of the effect of sorting on home language and parental education are generally higher, though higher standard errors lead to the coefficients on home language sorting to be statistically insignificant.

The negative results we obtain are also interesting. In particular, our estimates do not produce evidence indicating a substantial social interaction effect arising directly from family wage and salary income.

The point estimates suggest that within the average community studied, the magnitudes of the effects of sorting on the variance of test scores are not especially large. However, the larger cities in the data exhibit high degrees of both community-level diversity but school-level homogeneity (i.e., high rates of sorting). In those larger cities the estimated total effects of sorting are much larger. The significance of these results from a policy perspective would be clearer if we were able to say more about the effect of sorting on other moments of the test score distribution. In particular, it would be helpful to know whether sorting has a symmetric effect on the upper and lower tails of the distribution. While the data set used in this study does not allow us to investigate higher moments or different quantiles of the test score distributions, this would be an interesting question for future work.

This research could be extended in a number of ways. First, as both more years of data and data from more provinces are becoming available, increased sample sizes may allow for the estimation of richer models. The approach can also be applied to data from outside of Canada, although extension to U.S. data will require addressing the selection issues with private schools. Finally, if peer effects are nonlinear, sorting may affect average student performance as well as the variance in performance. Future research may determine whether this is the case.

References

Alberta Learning, *2001-2002 Funding Manual for School Authorities*, Province of Alberta, 2002.

— , *Achievement Test Multiyear Report 1997-2998 to 2001-2002*, Province of Alberta, 2002.

Arellano, Manuel and Olympia Bover, "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 1995, *68*, 29–51.

Boozer, Michael A. and Stephen E. Cacciola, "Inside the 'black box' of Project STAR: Estimation of peer effects using experimental data," Working Paper, Yale University 2001.

British Columbia Ministry of Education, *How are we doing? Demographics and performance of aboriginal students in BC public schools*, Queen's Printer, 2002.

Card, David and Jesse Rothstein, "Racial segregation and the Black-White test score gap," Working Paper, Princeton University 2004.

Coleman, James S., Ernest Q. Campbell, Carol F. Hobson, James M. McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York, *Equality of Educational Opportunity*, U.S. Department of Health, Education, and Welfare, 1966.

Cowley, Peter and Stephen T. Easton, *Report Card on Alberta's Elementary Schools*, The Fraser Institute, 2002.

Cragg, John G. and Stephen G. Donald, "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 1993, *9*, 222–240.

Cutler, David M. and Edward L. Glaeser, "Are ghettos good or bad?," *Quarterly Journal of Economics*, 1997, *112*, 827–872.

Ding, Weili and Steven F. Lehrer, "Do peers affect student achievement in China's secondary schools?," *Review of Economics and Statistics*, 2007, *89* (2).

Durlauf, Steven N., "A theory of persistent income inequality," *Journal of Economic Growth*, 1996, *1*, 75–93.

Epple, Dennis and Holger Sieg, "Estimating equilibrium models of local jurisdictions," *Journal of Political Economy*, 1999, *107* (4), 645–680.

Fernandez, Raquel and Richard Rogerson, "Income distribution, communities, and the quality of public education," *Quarterly Journal of Economics*, 1996, *111*, 135–164.

Glaeser, Edward L., Bruce I. Sacerdote, and José A. Scheinkman, "Crime and Social Interactions," *Quarterly Journal of Economics*, 1996, *111* (2), 507–548.

— , — , and — , "The Social Multiplier," *Journal of the European Economic Association*, 2003, *1* (2-3), 345–353.

Graham, Bryan S., "Identifying social interactions through excess variance contrasts," Working Paper, University of California - Berkeley 2005.

Hansen, Lars P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 1982, *50*, 1029–1054.

Hanushek, Eric A. and Margaret E. Raymond, "The effect of school accountability systems on the level and distribution of student achievement," *Journal of the European Economic Association*, 2004, *2*, 406–415.

— , John F. Kain, Jacob M. Markman, and Steven G. Rivkin, "Does peer ability affect student achievement?," *Journal of Applied Econometrics*, 2003, *18* (5), 527–544.

Hoxby, Caroline M., "Peer effects in the classroom: Learning from gender and race variation," Working Paper 7867, NBER 2000.

Interprovincial Education Statistics Project, *Summary of school statistics from the provinces and territories*, British Columbia Ministry of Education, 2002.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman, "Moving to Opportunity in Boston: Early results of a randomized mobility experiment," *Quarterly Journal of Economics*, 2001, *116* (2), 607–654.

Kremer, Michael, "How much does sorting increase inequality?," *Quarterly Journal of Economics*, 1997, *112*, 115–140.

Manski, Charles F., "Identification of endogenous social effects: The reflection problem," *Review of Economic Studies*, 1993, *60* (3), 531–542.

Microsoft Corporation, *Microsoft Streets and Trips, 2005 edition* 2004. Computer program.

Staiger, Douglas and James H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 1997, *65*, 557–586.

Urquiola, Miguel, "Does School Choice Lead to Sorting? Evidence from Tiebout Variation," *American Economic Review*, 2005, *95*, 1310–1326.

# Schools	**# communities Grade 6**	**# communities Grade 9**
1	139	162
2	37	28
3	13	8
4	8	1
5	1	2
6	1	1
7	1	1
8	2	2
9	0	2
10 to 99	9	4
100 or more	2	2
Total # Schools	1025	604
Total # Communities	213	213

Table 1: Frequency count of communities by number of schools, Grade 6 and Grade 9.

Change in # Schools (Grade 9 - Grade 6)	**# Communities**
+1	2
0	162
-1	27
-2	9
-3	2
-4	1
-5	2
-6	1
-8	2
-10 to -99	3
-100 or more	2
Total	213

Table 2: Frequency count of communities by difference between number of Grade 6 and Grade 9 schools in community.

Variable	Mean	Std. Dev.	Min	Max
Grade 9 Language Arts exam score				
Total # writing	177.98	1039.44	6.00	10817.00
Average # writing	39.90	38.75	6.00	206.60
Community mean	67.27	5.80	32.71	82.29
Community variance	155.78	61.75	16.57	497.90
Cross-school variance	8.27	8.75	0.01	44.86
Grade 9 Math exam score				
Total # writing	170.24	977.80	6.00	10001.00
Average # writing	39.08	37.30	6.00	188.53
Community mean	32.76	4.33	16.33	43.08
Community variance	83.67	29.72	7.60	208.80
Cross-school variance	4.69	5.41	0.00	22.07
Visible minority (%)				
Community mean	2.94	5.57	0.00	29.90
Community variance	254.18	442.88	0.00	2096.16
Cross-school variance	10.57	39.22	0.00	252.81
Aboriginal (%)				
Community mean	8.11	18.86	0.00	98.09
Community variance	391.01	583.68	0.00	2487.14
Cross-school variance	8.90	40.27	0.00	282.35
Official language not usually spoken at home (%)				
Community mean	3.91	8.97	0.00	63.21
Community variance	295.74	537.85	0.00	2477.65
Cross-school variance	4.00	10.67	0.00	62.44
Annual family wage and salary income ($1,000)				
Community mean	40.24	12.40	8.80	80.57
Community variance	204.69	220.26	6.96	1896.45
Cross-school variance	26.89	59.18	0.00	265.53
Annual family transfer income ($1,000)				
Community mean	4.54	1.78	1.76	14.90
Community variance	3.50	4.53	0.19	28.47
Cross-school variance	0.20	0.36	0.00	1.50
Education of primary parent (years)				
Community mean	12.43	0.90	9.57	15.50
Community variance	1.00	1.15	0.04	8.67
Cross-school variance	0.09	0.17	0.00	0.79

Table 3: Summary statistics for exam scores and family characteristics. Summary statistics are calculated from subset of multi-school communities for cross-school variances, and from entire data set for all other variables.

	CSV of:					
	Vis. min.	Aborig.	Home lang.	W/S income	Transfer inc.	Parent educ.
CSV of:						
Vis. min.		0.01	0.61	0.31	0.28	0.41
Aborig.	0.67		0.24	0.49	0.40	0.04
Home lang.	0.76	0.64		0.73	0.64	0.84
W/S income	0.56	0.55	0.49		0.66	0.76
Transfer inc.	0.50	0.55	0.48	0.46		0.69
Parent educ.	0.73	0.72	0.72	0.74	0.74	
	ΔCSV of:					
	Vis. min.	Aborig.	Home lang.	W/S income	Transfer inc.	Parent educ.
ΔCSV of:						
Vis. min.		0.03	0.25	0.48	0.53	0.10
Aborig.	0.57		0.07	0.17	0.23	0.32
Home lang.	0.83	0.53		0.19	0.09	0.21
W/S income	0.68	0.71	0.64		0.45	0.27
Transfer inc.	0.53	0.60	0.49	0.61		0.39
Parent educ.	0.72	0.57	0.63	0.67	0.57	

Table 4: Correlation matrices for sorting variables. The upper triangle of each matrix reports Pearson (linear) correlation, the lower triangle reports Spearman (rank) correlation.

	Language Arts				Mathematics			
(# writing)/1000	-0.06	-0.07	-0.06	-0.13	0.00	-0.01	0.00	-0.02
t-statistic	-1.70	-2.35	-1.69	-2.23	0.11	-0.29	0.12	-0.26
p-value	0.09	0.02	0.09	0.03	0.91	0.77	0.91	0.80
Var(Vis.min)	0.15	0.14	0.15	0.14	0.10	0.10	0.10	0.10
	1.43	1.33	1.43	1.37	1.27	1.17	1.27	1.24
	0.15	0.19	0.15	0.17	0.21	0.24	0.21	0.22
Var(Aboriginal)	0.09	0.10	0.08	0.09	0.06	0.06	0.05	0.06
	1.49	1.50	1.19	1.43	0.96	0.98	0.83	0.94
	0.14	0.14	0.24	0.15	0.34	0.33	0.41	0.35
Var(Home lang.)	-0.00	-0.00	-0.00	-0.00	-0.10	-0.10	-0.10	-0.10
	-0.05	-0.03	-0.04	-0.04	-1.48	-1.46	-1.47	-1.47
	0.96	0.97	0.97	0.96	0.14	0.15	0.14	0.14
Var(W/S income)	0.21	0.21	0.20	0.21	0.18	0.18	0.17	0.18
	2.27	2.28	2.18	2.27	2.39	2.40	2.34	2.38
	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02
Var(Transfer inc.)	0.12	0.12	0.12	0.12	0.15	0.14	0.15	0.14
	1.11	1.09	1.11	1.09	2.05	2.02	2.05	2.03
	0.27	0.28	0.27	0.28	0.04	0.05	0.04	0.04
Var(Parent educ.)	-0.26	-0.26	-0.26	-0.26	-0.19	-0.19	-0.19	-0.19
	-2.97	-2.97	-2.94	-2.97	-2.13	-2.13	-2.12	-2.12
	0.00	0.00	0.00	0.00	0.03	0.03	0.04	0.03
CSV(Vis.min)		0.03				0.03		
		1.52				1.66		
		0.13				0.10		
CSV(Aboriginal)			0.05				0.01	
			2.16				0.48	
			0.03				0.63	
CSV(Home lang.)				0.08				0.02
				1.30				0.35
				0.19				0.73
Constant	2.33	2.33	2.34	2.33	2.66	2.66	2.66	2.66
	19.71	19.63	19.55	19.56	20.69	20.64	20.51	20.58
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5: Regression results: Community level variance of Grade 9 test scores, 2001 cross-sectional data, with sorting by ethnicity and language.

	Language Arts			Mathematics		
(# writing)/1000	-0.04	-0.06	-0.04	0.03	0.00	0.01
t-statistic	-0.92	-1.78	-0.75	0.90	0.08	0.17
p-value	0.36	0.08	0.46	0.37	0.94	0.87
Var(Vis.min)	0.15	0.15	0.15	0.10	0.10	0.10
	1.43	1.42	1.43	1.26	1.26	1.27
	0.16	0.16	0.15	0.21	0.21	0.21
Var(Aboriginal)	0.10	0.09	0.09	0.06	0.06	0.06
	1.50	1.46	1.48	1.01	0.95	0.96
	0.14	0.15	0.14	0.31	0.34	0.34
Var(Home lang.)	-0.00	-0.00	-0.00	-0.10	-0.10	-0.10
	-0.05	-0.04	-0.06	-1.49	-1.47	-1.47
	0.96	0.96	0.95	0.14	0.14	0.14
Var(W/S income)	0.21	0.21	0.21	0.19	0.18	0.18
	2.14	2.26	2.26	2.32	2.37	2.35
	0.03	0.03	0.02	0.02	0.02	0.02
Var(Transfer inc.)	0.12	0.12	0.12	0.14	0.14	0.15
	1.10	1.06	1.11	2.02	1.98	2.05
	0.27	0.29	0.27	0.04	0.05	0.04
Var(Parent educ.)	-0.26	-0.26	-0.26	-0.19	-0.19	-0.19
	-2.93	-2.92	-2.97	-2.15	-2.12	-2.12
	0.00	0.00	0.00	0.03	0.04	0.04
CSV(W/S income)	-0.02			-0.04		
	-0.32			-0.88		
	0.75			0.38		
CSV(Transfer inc.)		0.00			0.00	
		0.04			0.08	
		0.97			0.94	
CSV(Parent educ.)			-0.03			-0.00
			-0.43			-0.08
			0.67			0.94
Constant	2.33	2.33	2.33	2.66	2.66	2.66
	19.66	19.61	19.66	20.64	20.62	20.64
	0.00	0.00	0.00	0.00	0.00	0.00

Table 6: Regression results: Community level variance of Grade 9 test scores, 2001 cross-sectional data, with sorting by income and education.

	Language Arts	Mathematics
(# writing)/1000	0.05	0.14
t-statistic	0.70	1.26
p-value	0.49	0.21
Var(Vis.min)	0.14	0.10
	1.30	1.14
	0.19	0.25
Var(Aboriginal)	0.08	0.05
	1.10	0.74
	0.27	0.46
Var(Home lang.)	-0.00	-0.10
	-0.04	-1.42
	0.97	0.16
Var(W/S income)	0.23	0.20
	2.24	2.36
	0.03	0.02
Var(Transfer inc.)	0.12	0.14
	0.99	1.82
	0.32	0.07
Var(Parent educ.)	-0.28	-0.20
	-2.98	-2.19
	0.00	0.03
CSV(Vis.min)	0.06	0.07
	1.82	1.94
	0.07	0.05
CSV(Aboriginal)	0.17	0.12
	2.68	1.90
	0.01	0.06
CSV(Home lang.)	-0.12	-0.17
	-1.21	-1.26
	0.23	0.21
CSV(W/S income)	-0.18	-0.15
	-2.29	-2.19
	0.02	0.03
CSV(Transfer inc.)	-0.03	-0.01
	-0.58	-0.20
	0.56	0.85
CSV(Parent educ.)	0.12	0.11
	1.60	1.61
	0.11	0.11
Constant	2.33	2.66
	19.34	20.35
	0.00	0.00

Table 7: Regression results: Community level variance of Grade 9 test scores, 2001 cross-sectional data, with all sorting variables.

	Language Arts				Mathematics			
Δ(# writing)/1000	1.16	2.39	1.05	0.77	0.42	1.09	0.39	-0.13
t-statistic	1.66	2.01	1.69	1.12	0.63	1.39	0.59	-0.20
p-value	0.10	0.04	0.09	0.26	0.53	0.16	0.56	0.84
ΔCSV(Vis.min)		-0.30				-0.21		
		-1.23				-1.29		
		0.22				0.20		
ΔCSV(Aboriginal)			0.12				-0.00	
			2.23				-0.01	
			0.03				0.99	
ΔCSV(Home lang.)				0.04				0.07
				1.94				3.76
				0.05				0.00
Constant	0.12	0.11	0.14	0.15	0.39	0.34	0.39	0.39
	1.53	1.36	1.69	1.86	6.39	4.95	6.25	5.93
	0.13	0.17	0.09	0.06	0.00	0.00	0.00	0.00
Overidentifying restrictions								
J-statistic	0.31	0.99	0.24	1.39	1.64	1.20	2.74	2.57
p-value	0.58	0.61	0.89	0.50	0.20	0.55	0.25	0.28

Table 8: Regression results: Difference in community level variance of Grade 9 test scores in 2001 and Grade 6 test scores in 1998, with sorting by ethnicity and language.

	Language Arts			Mathematics		
Δ(# writing)/1000	1.92	1.53	0.74	1.65	0.50	0.07
t-statistic	1.78	2.20	1.29	1.50	0.70	0.12
p-value	0.07	0.03	0.20	0.13	0.48	0.91
ΔCSV(W/S income)	-0.09			-0.18		
	-0.84			-1.43		
	0.40			0.15		
ΔCSV(Transfer inc.)		-0.03			-0.02	
		-1.05			-0.65	
		0.29			0.52	
ΔCSV(Parent educ.)			0.10			0.03
			3.71			2.24
			0.00			0.03
Constant	0.12	0.12	0.21	0.35	0.36	0.38
	1.44	1.49	3.31	5.27	5.33	6.37
	0.15	0.14	0.00	0.00	0.00	0.00
Overidentifying restrictions						
J-statistic	0.32	0.46	1.50	2.56	2.97	2.70
p-value	0.85	0.79	0.47	0.28	0.23	0.26

Table 9: Regression results: Difference in community level variance of Grade 9 test scores in 2001 and Grade 6 test scores in 1998, with sorting by income and education.

	Language Arts	**Mathematics**
Δ(# writing)/1000	1.93	0.74
t-statistic	2.20	0.78
p-value	0.03	0.43
ΔCSV(Vis.min)	-0.12	-0.24
	-1.01	-2.24
	0.31	0.02
ΔCSV(Aboriginal)	0.10	-0.03
	3.01	-0.65
	0.00	0.51
ΔCSV(Home lang.)	0.02	0.07
	0.54	2.84
	0.59	0.00
ΔCSV(W/S income)	-0.06	-0.06
	-1.00	-0.59
	0.32	0.55
ΔCSV(Transfer inc.)	-0.05	-0.03
	-2.00	-1.81
	0.05	0.07
ΔCSV(Parent educ.)	0.10	0.04
	7.05	4.28
	0.00	0.00
Constant	0.17	0.31
	2.52	4.66
	0.01	0.00
Overidentifying restrictions		
J-statistic	5.60	4.21
p-value	0.59	0.76

Table 10: Regression results: Difference in community level variance of Grade 9 test scores in 2001 and Grade 6 test scores in 1998, with all sorting variables.

	Language	Arts		Mathematics		
Δ(# writing)/1000	1.51	1.00	-0.04	1.32	0.30	-0.08
t-statistic	0.65	1.66	-0.04	1.60	0.43	-0.10
p-value	0.51	0.10	0.97	0.11	0.67	0.92
ΔCSV(Vis.min)	-0.07			-0.32		
	-0.12			-1.45		
	0.91			0.15		
ΔCSV(Aboriginal)		0.14			-0.02	
		2.25			-0.46	
		0.02			0.65	
ΔCSV(Home lang.)			0.11			0.06
			1.42			0.88
			0.16			0.38
Constant	0.13	0.14	0.12	0.33	0.36	0.36
	1.60	1.70	1.57	4.52	5.96	5.36
	0.11	0.09	0.12	0.00	0.00	0.00
Overidentifying restrictions						
J-statistic	0.66	0.19	0.82	0.07	4.44	3.64
p-value	0.72	0.91	0.66	0.97	0.11	0.16
Instrument relevance						
First stage F statistic	3.35	5.03	4.09	4.33	6.44	4.21
p-value	0.07	0.03	0.04	0.04	0.01	0.04

Table 11: Regression results: Difference in community level variance of Grade 9 test scores in 2001 and Grade 6 test scores in 1998, with sorting by ethnicity and language. Change in maximum CSV used as instrument for change in CSV.

	Language	Arts		Mathematics		
Δ(# writing)/1000	2.73	1.40	0.21	1.48	0.48	0.02
t-statistic	1.69	2.27	0.31	0.96	0.69	0.04
p-value	0.09	0.02	0.76	0.34	0.49	0.97
ΔCSV(W/S income)	-0.16			-0.17		
	-0.75			-0.83		
	0.45			0.41		
ΔCSV(Transfer inc.)		-0.02			-0.01	
		-0.33			-0.29	
		0.75			0.77	
ΔCSV(Parent educ.)			0.15			0.05
			2.49			1.21
			0.01			0.23
Constant	0.14	0.13	0.20	0.33	0.37	0.38
	1.72	1.55	2.74	4.54	5.14	6.76
	0.09	0.12	0.01	0.00	0.00	0.00
Overidentifying restrictions						
J-statistic	1.20	1.13	0.91	1.73	1.46	2.15
p-value	0.55	0.57	0.64	0.42	0.48	0.34
Instrument relevance						
First stage F statistic	22.85	5.00	11.81	25.08	5.71	12.42
p-value	0.00	0.03	0.00	0.00	0.02	0.00

Table 12: Regression results: Difference in community level variance of Grade 9 test scores in 2001 and Grade 6 test scores in 1998, with sorting by income and education. Change in maximum CSV used as instrument for change in CSV.

	Language Arts	Mathematics
Δ(# writing)/1000	3.01	0.89
t-statistic	1.42	0.67
p-value	0.16	0.50
ΔCSV(Vis.min)	-0.11	-0.40
	-0.21	-1.67
	0.83	0.09
ΔCSV(Aboriginal)	0.14	-0.02
	1.56	-0.44
	0.12	0.66
ΔCSV(Home lang.)	0.10	0.09
	1.32	1.52
	0.19	0.13
ΔCSV(W/S income)	-0.37	-0.11
	-1.56	-0.54
	0.12	0.59
ΔCSV(Transfer inc.)	-0.02	-0.00
	-0.63	-0.04
	0.53	0.96
ΔCSV(Parent educ.)	0.13	0.06
	2.03	1.17
	0.04	0.24
Constant	0.17	0.33
	2.04	4.88
	0.04	0.00
Overidentifying restrictions		
J-statistic	4.40	7.42
p-value	0.73	0.39
Instrument relevance		
Cragg-Donald F statistic	1.80	2.73
p-value	0.04	0.01

Table 13: Regression results: Difference in community level variance of Grade 9 test scores in 2001 and Grade 6 test scores in 1998, with all sorting variables. Change in maximum CSV used as instrument for change in CSV.

	Language Arts			Mathematics		
	OLS	FD	IV	OLS	FD	IV
Average multi-school community:						
CSV(Aboriginal)	0.8%	2.0%	2.4%	0.2%	0.0%	-0.3%
CSV(Home lang.)	2.2%	1.2%	3.0%	0.5%	1.6%	1.4%
CSV(Parent educ.)	-0.9%	3.5%	5.4%	-0.1%	1.0%	1.5%
Edmonton:						
CSV(Aboriginal)	1.1%	2.8%	3.4%	0.3%	0.0%	-0.4%
CSV(Home lang.)	33.5%	18.4%	45.0%	7.2%	21.8%	18.7%
CSV(Parent educ.)	-8.2%	31.4%	47.6%	-1.1%	8.1%	11.7%
Calgary:						
CSV(Aboriginal)	0.1%	0.4%	0.4%	0.0%	0.0%	0.0%
CSV(Home lang.)	22.2%	12.2%	29.8%	4.9%	15.0%	12.8%
CSV(Parent educ.)	-7.9%	30.4%	46.1%	-1.1%	7.9%	11.4%

Table 14: Predicted reduction in variance of test score from completely desegregating communities on specified dimension.