How Important Are School Principals in the Production of Student Achievement?

Elizabeth Dhuey
Department of Management
Centre for Industrial Relations and Human Resources
University of Toronto

416.978.2721
elizabeth.dhuey@utoronto.ca

Justin Smith
Department of Economics
Wilfrid Laurier University

519.884.0710 ext.4554
jusmith@wlu.ca

Abstract

As school leaders, principals can influence student achievement in a number of ways, such as: hiring and firing of teachers, monitoring instruction, and maintaining student discipline, among many others. We measure the effect of individual principals on gains in math and reading achievement between grades 4 and 7 using a value added framework. We estimate that a one standard deviation improvement in principal quality can boost student performance by 0.289 - 0.408 standard deviations in reading and math, while the principal at the 75th percentile improves scores by 0.170 - 0.193 relative to the median principal. Our results imply that isolating the most effective principals and allocating them accordingly between schools can have a significant, positive effect on reducing achievement gaps.

JEL Classification Code: I2

1. Introduction

Many researchers and policy makers believe that a central component of school quality is teacher quality. Indeed, in recent years an extensive literature on the importance of teachers on student achievement has been amassed. This literature has established the importance of teacher quality on student achievement and has also shown that teacher quality can have long-term effects on student achievement (Chetty, Friedman, and Rockoff, 2011). Despite the extensive literature on teacher quality, relatively little attention has been paid by economists to the importance of principals in the production of student achievement. As the main administrator in a school, the principal is responsible, among many other things, for maintaining and improving teacher quality, monitoring and enforcing student conduct, and ensuring the curricula are properly implemented. Therefore, it is important to examine principal quality to find out whether it also is a central component of school quality.

In this paper we quantify the importance of principals in the production of student test score gains between fourth and seventh grades using principal fixed effects estimated from longitudinal administrative data from the province of British Columbia (BC), Canada. Principal fixed effects measure the impact of all unchanging characteristics of a principal that are transferrable between schools, such as leadership ability, personality, and gender. We obtain these estimates for math and reading scores using a value-added model that controls for observable, time-varying student, school, and neighbourhood factors, along with fixed school effects. The frequent mobility of principals between schools in BC allows us to identify principal effects using within school variation, which is important because school fixed effects remove the influence of fixed school characteristics that might bias estimated principal effects if principals sort across schools based on those attributes. We also estimate the effect of experience as a principal within a school (which we call tenure), and overall experience as a principal (which we call experience) on student achievement.

We estimate that moving one standard deviation up the distribution of principal quality improves math scores by 0.408 standard deviations and reading scores by 0.289 standard deviations, and that the principal at the 75th percentile improves math scores by 0.193 standard deviations and reading scores by 0.170 standard deviations relative to the median principal. We also find that the extent of principal experience and length of tenure in a school have no significant impact on student performance. This implies that, at least in BC, what matters for student achievement is finding a principal with good fixed attributes and assigning that individual to the correct school.

2. Existing Literature on the Effect of Principals

The existing economics literature related to our research can be separated into three strands. The first focuses on estimating the overall effect of the principal on achievement, the second on the effects of specific principal attributes on achievement and the third on the effects of specific principal training programs.

Most closely related to our research are two recent papers by Coelli and Green (2012) and Branch, Hanushek, and Rivkin (2012). Coelli and Green (2012) use a sample of students entering grade 12 drawn from administrative data in BC to estimate the variance of the effect of principals on high school graduation rates and grade 12 provincial exam scores in English. They find that moving up the distribution of principal effects can improve both outcomes, but has a particularly significant effect on grade 12 English exam scores. They also estimate a dynamic model of principal effects and find that that a principal’s full impact increases gradually over time. Even though we both draw our samples from BC data and estimate the dispersion of principal effects on outcomes, our research is best viewed as a complement to Coelli and Green (2012) because of fundamental differences. The main difference is in the populations under analysis, in that we study elementary school students and they study high school students. In addition, though our empirical strategies rely on similar identifying assumptions, they are otherwise different. For example, we estimate a fixed effect for each principal within a value-added model and report the dispersion in those effects, whereas they estimate the variance directly and do not use value-added modeling.[1]

Branch, Hanushek, and Rivkin (2012) use data from Texas to estimate the importance of principals on student math and reading test score gains. In their main analysis, they first estimate an implied standard deviation in principal by school fixed effects in models with and without school effects, then directly estimate a lower bound variance using a method similar to Coelli and Green (2012). In models without school fixed effects, they estimate a standard deviation in principal effects for a year of gains of roughly 0.2, which shrinks by about half when school effects are added to the model. In their main estimates, they focus on the first three years of a principal's tenure out of concern that simply adding tenure controls is insufficient to account for tenure effects. This concern may be unwarranted however, because when they re-estimate their model on principals of all tenure lengths, the results do not vary substantially. Lower bound estimates are significantly lower than their main estimates. Finally, they find that the variance of principal effects increases with the school poverty rate. While our empirical strategy is similar to Branch et al. (2012), one fundamental difference is that we estimate pure principal effects, rather than principal by school effects. This is an important distinction as principal by school effects do not allow researchers to separate the independent influences of principals and schools.

The second strand of the literature on principal effects focuses on the relationship between observable characteristics of principals and school performance. Many of these papers focus on the role of principal education and experience on school performance, and have found mixed results. In terms of education, some researchers find relationships such that a more educated principal is associated with a worse school performance (see Ballou and Podgursky, 1993 and Eberts and Stone, 1988), whereas other researchers find no relationship (see Clark, Martorell, and Rockoff, 2009). In addition, Eberts and Stone (1988) and Clark, Martorell, and Rockoff (2009) find a positive correlation between teaching experience and school performance, but Brewer (1993) finds no correlation.

The third strand of the literature examines the effects of principal training programs. For example, Corcoran, Schwartz, and Weinstein (2012) show that principals trained in the New York City Aspiring Principals Program have positive effects on school performance. Clark, Martorell, and Rockoff (2009), however, find mixed evidence on the relationship between formal principal training/development programs and school achievement.

3. Principals and Student Performance

3.1 How do Principals Affect Academic Achievement?

To explain the link between principal leadership and student learning, we follow the conceptual framework of Leithwood et al. (2004). The framework outlines the pathways through which leadership ultimately affects student achievement, which is mainly indirectly through schools, classrooms, and teachers. It also outlines how a principal's leadership method is determined by outside factors, which may themselves have an impact on achievement.

As mentioned above, principals affect student achievement mainly through schools, classrooms, and teachers. Leithwood et al. (2004) note that principals mainly affect school “conditions,” through such avenues as developing a governance structure, creating a school culture (for example, one that is inclusive), and developing school-wide policies about retention, adherence to the curriculum, and working conditions for teachers. Such changes in school conditions developed at the behest of the principal may lead to variations in student achievement. Principals may also separately affect classrooms within schools, by manipulating such variables as class size, efficient allocations of teachers to students, student ability grouping, and by monitoring the content and nature of instruction and student assessments. Policies with respect to these classroom conditions will also influence student achievement. Changes to both school and classroom conditions will affect the way in which teachers interact with students, which in turn will have its own effect on achievement. For example, a teacher assigned to a small class by the principal may be able to provide more individualized instruction to students compared to another teacher assigned to a large class. The takeaway is that policies set by the principal at the school and classroom levels, and how teachers interact with those policies, will combine to exert an influence over student achievement.

Principals do not form their leadership style within a vacuum. There are many outside factors that influence the way principals lead, some of which may also affect student achievement directly. Examples of the main factors influencing the way principals lead schools include province and district policies, and principals’ prior professional learning and training (e.g., education, experience as a teacher, experience as an administrator, mentoring, etc.). Province and district leadership and policies are particularly important as they directly affect how the principal is required to lead and supervise the development, delivery, assessment, and improvement of the education for all students in their school. Principal leadership is also clearly influenced by their education and experience, and may also be shaped in part by the interests of stakeholders. Leadership style may also be influenced by the family background of the students in the school, impacting perhaps the policies and programs principals enact or the financial resource available to the school. At the same time, a student's family background can influence learning directly by shaping the educational culture of the community and school, or by shaping the assumptions, norms and beliefs held by the family regarding the importance of academic learning. Therefore, we cannot view the principals as independent of student background, making it important to control for these factors in any analysis of how principals affect student achievement. In our econometric model below, we add many student background variables at the individual, school, and neighbourhood level, in order to absorb the potential confounding influence of family background.

It is important to consider that the indirect effects of principals on achievement may have a dynamic component. For instance, changes in the makeup of family background of a school's students may immediately affect the principal's directives regarding the nature of instruction but also may slowly affect the school and improvement planning. In addition, principal tenure and experience is inherently dynamic in the sense that a principal may draw on past experiences in the current or past schools when they set school and classroom conditions. Principals may continue to have an effect on current student performance through policies set in years past that continue to linger after a new principal arrives. In our econometric model below, we allow for such dynamics in two ways. First, we explicitly add controls for principal tenure and experience. Second, we allow for the effects of past principals to persist into the future, but to decay geometrically. This should allow us to capture the main dynamics through which principals continue to affect student performance.

3.2 Education in British Columbia

The public school system in BC is very similar to that of other jurisdictions across Canada and the United States. There are 60 school districts in BC, many of which correspond to city boundaries in the densely populated areas (e.g., Vancouver and its surrounding area, and Victoria). In the more rural areas, districts cover a much wider area. Prior to 2002, students attending public schools were required to attend the school determined by their catchment area. If they wanted to attend a school outside their designated area, students needed permission from both the catchment school principal and the out-of-catchment principal. Since 2002, legislation has existed that allows parents to choose any public school, regardless of catchment, with the exception that if a student wants to transfer to a school which has excess demand, the student needs to obtain permission from the principal. In addition, 30 percent of students attend a middle or junior high school that begins in either sixth or seventh grade, and the other 70 percent attend a school with grades extending through seventh grade and higher (Dhuey, 2013). Therefore, the majority of the students in our sample will have the opportunity to stay at the same school between grades 4 and 7.[2]

Each year since 1999, students in fourth and seventh grades are tested in reading, writing, and math using the Foundation Skills Assessment (FSA) tests. All students are expected to participate in the tests, with the exception of some ESL students and students with special needs. These tests are low stakes in the sense that no funding is tied to the outcomes, nor do they contribute to student course grades. Therefore, despite recent work regarding teacher and principal cheating, principal cheating on these exams seems unlikely, mainly due to the fact that principal compensation and school resources are set according to guidelines that are not based on student outcomes (Coelli and Green, 2012).

Principals in BC are appointed by school boards, and their duties are outlined in various parts of the School Act from the Revised Statutes of BC and are specifically listed in BC Regulation 265/89 and the many amendments thereof. These duties of BC principals are typical of most principals across jurisdictions; they are responsible for carrying out orders from the school district or from the Ministry of Education, and they are also responsible for ensuring that the instructional practices of their school conform to the School Act. In addition, principals are responsible for the smooth functioning of the province's various standardized testing programs. Unlike some other jurisdictions, BC principals do not directly hire and fire teachers; instead they provide information to the school district about teacher performance, and any disciplinary action is then decided at this upper level. However, the environment they create within their school may affect teachers' decisions. Finally, BC principals carry out numerous administrative and operational tasks such as making teacher timetables, maintaining school records, and monitoring the conduct of students.

4. Estimation

4.1 Estimation Framework

Suppose that in some grade g, a student's test score is determined within the following model:

$$y_i^g = \sum_{h=0}^{g}\left[x'_{ih}\beta_h^g + z'_{s(i,h)h}\gamma_h^g + \delta_{p(i,h)}^g + \phi_{s(i,h)}^g\right] + \eta^g + \varepsilon_i^g \qquad (1)$$

where x'_{ih} is a vector of student-level demographic characteristics including family inputs that may vary across grades, $z'_{s(i,h)h}$ are school-level factors for student *i* in grade *h* that vary across grades, $\delta^{g}_{p(i,h)}$is the effect of student *i*'s principal from grade *h* on outcomes during grade *g*, $\phi^{g}_{s(i,h)}$is the effect of student *i*'s school from grade *h* on outcomes during grade *g*, $\eta^{g}$is a grade fixed effect, and $\varepsilon^{g}_{i}$is a student-level error term. The index *h* runs from the beginning of the schooling process up until grade *g*.

In this model, family and school inputs from the past continue to have an effect during grade *g*. Available data rarely include information on all past inputs, so estimating this model is generally not feasible. In light of this, researchers typically make assumptions about how inputs from past grades continue to affect student outcomes in current grades. One common assumption is that the effects of all past inputs decay geometrically into higher grades at the same rate. This implies that lagged test scores are a sufficient statistic for all factors which affect student outcomes prior to grade *g* and as such, researchers need only control for lagged scores on the right hand side of a regression to identify the effects of current inputs on achievement. To see this, suppose that all family and school inputs decay geometrically between time periods at rate λ, i.e. $\beta^{g}_{g-1} = \lambda\beta^{g-1}_{g-1};\ \delta^{g}_{p(i,g-1)} = \lambda\delta^{g-1}_{p(i,g-1)}$, etc. Lagging equation 1 and multiplying by λ, we have

$$\lambda y^{g-1}_{i} = \lambda\left(\sum_{h=0}^{g-1}\left[x'_{ih}\beta^{g-1}_{h} + z'_{s(i,h)h}\gamma^{g-1}_{h} + \delta^{g-1}_{p(i,h)} + \phi^{g-1}_{s(i,h)}\right] + \eta^{g-1} + \varepsilon^{g-1}_{i}\right) \qquad (2)$$

If we then subtract equation 2 from equation 1, the model becomes

$$y^{g}_{i} - \lambda y^{g-1}_{i} = x'_{ig}\beta^{g}_{g} + z'_{s(i,g)g}\gamma^{g}_{g} + \delta^{g}_{p(i,g)} + \phi^{g}_{s(i,g)} + (\eta^{g} - \lambda\eta^{g-1}) + (\varepsilon^{g}_{i} - \lambda\varepsilon^{g-1}_{i}) \quad (3)$$

This is the lagged score value-added model (Rothstein, 2010) that appears frequently in the literature. With student data across two or more contiguous grades, researchers can estimate the parameters of this equation by regressing current test scores on the set of relevant variables, and controlling for lagged test scores on the right hand side.[3]

Our data do not contain scores across contiguous grades for each student. Instead, we observe scores from grade 4 and grade 7. Using the same logic as above, but instead subtracting scores from g-3, and defining $\pi = \lambda^3$ for notational simplicity, we get

$$y_i^7 - \pi y_i^4 = \sum_{h=5}^{7}\left[x'_{ih}\beta_h^7 + z'_{s(i,h)h}\gamma_h^7 + \delta_{p(i,h)}^7 + \phi_{s(i,h)}^7\right] + (\eta^7 - \pi\eta^4) + (\varepsilon_i^7 - \pi\varepsilon_i^4) \quad (4)$$

This implies that even if the effect of past inputs decays geometrically at the same constant rate, subtracting the g-3 test scores does not eliminate the influence of inputs from the previous two years. We must therefore directly control for the principals, school and family characteristics for grade 5 and grade 6 in our estimating equation.[4]

4.2 Empirical Specification

We estimate a version of equation 4 by Ordinary Least Squares, treating the principal and school effects as parameters. Our primary interest is the grade 7 principal fixed effects, which we will present in the form of summary statistics, with a focus on their standard deviation and the gap between the principal at the 75th percentile and the median principal. As we outlined in our conceptual framework, these principal fixed effects are interpreted as the effect of each individual grade 7 principal on grade 4 to grade 7 achievement gains that are invariant across time, students, and schools conditional on other time-varying and time-invariant student and school-level factors. These effects will measure the impact of any fixed attribute of the principal, such as leadership ability, gender, previous education and training.

While it is theoretically possible to estimate equation 4 in its current form, trying to separately identify principal and school effects across three grades puts extremely high demands on the data, and relies on variation that is questionably exogenous. To obtain the cleanest possible estimates, we restrict our sample to only students who do not change schools between grade 5 and grade 7. This restriction serves two purposes. First, it allows us to control for the effects of the school across all grades by including a fixed effect for the grade 7 school. Second, it eliminates student mobility as a source of variation when estimating principal effects, which is key because student moves are unlikely to be an exogenous source of variation. The remaining identifying variation is based on principal mobility between schools, the exogeneity of which we examine in detail in Section 6.

Even after restricting the sample in this way, we need to control for the effect of the grade 5 and grade 6 principal in order to identify the effect of the grade 7 principal. Separately identifying principal effects in each grade is not feasible using only variation based on principal moves, so we instead choose to control for their effects, along with the school effect, by including a school by grade 5 principal by grade 6 principal fixed effect - i.e. a fixed effect based on the three way interaction between the school and the principals from grades 5 and 6. To be specific, we include a dummy variable for each combination of grade 5 principal, grade 6 principal, and school that exists in the data. This strategy will control for the effect of each previous principals and school because none vary within each combination. To simplify the exposition below, we continue to refer to the principal 5/6 by school interaction as a “school effect," and the grade 7 principal fixed effect as a "principal effect." We will be clear when we depart from these simplifications.[5]

There are additional important statistical issues to keep in mind when estimating models with two high-dimensional levels of fixed effects. Normally, fixed effects are used to control for unobserved heterogeneity when researchers are trying to identify other coefficients in the model. In that scenario, the fixed effects are generally never explicitly estimated, but rather swept away by a within transformation. Our primary focus however, is on the estimates of the principal fixed effects themselves. Even if we had a model containing only grade 7 principal effects (i.e., no school or past principal dummies), we could not separately identify all of the principal effects when the model contains a constant (the "dummy variable trap"); thus, an identifying restriction must be imposed. The most obvious restriction is to simply set one principal effect equal to zero. This is unsatisfactory because it forces the interpretation of the remaining fixed effects to be deviations from the dropped principal, which is sensitive to which principal is dropped. Instead, we restrict the fixed effects to sum to zero as in Mihaly et al. (2010). With this restriction, there is no dropped principal but rather, the fixed effects are parameterized so that their coefficients are deviations from the average of the principal effects.

The problem becomes more complicated when we include both principal and school effects. Principals and schools are implicitly separated into many disconnected groups because not every principal works at every school. Each group contains all principals that have ever worked at any school in the group, and all schools that have ever employed any principal in the group. Principals in one group are never observed working at a school in any other group, and schools in one group never employ a principal from any other group. When including school fixed effects, one principal effect in each group is not identified, and additional restrictions must be imposed.[6] Following Mihaly et al. (2010), when the model contains principal and school effects, we restrict the principal effects to sum to zero within each group. The principal effects in this case are interpreted as deviations from the within-group mean.

One consequence of including school effects in the model is that principal effects must vary within schools in order to be identified. Thus, in our case, at least one school in each group must employ multiple principals over time to contribute to estimation. This arises mainly through principal mobility across schools, but can also occur when new principals enter the sample. In the extreme case where a group contains exactly one principal and one school (i.e., the principal does not move and the school does not employ any other principal), principal effects are not identified and therefore not estimated. The reported standard deviations based on the school fixed effects specification use only identified principal fixed effects. Fortunately, there is enough mobility to identify an effect for most principals in the school effects model (see Table 1). Despite the fact that we can only identify principal effects for the subset of identified principals, it is our preferred specification because controlling for fixed school and past principal factors is crucial. We examine the difference between "identified" and "unidentified" principals in Section 5.3.

As explained in Kane and Staiger (2002) the variance of estimated value added measures could be due to differences in ability across principals (i.e. an actual principal effect) and could also be affected by other non-persistent variation. Such variation could be due to things such as: sampling error, disruptive students, cohort-specific school or school-by-grade shocks, or any other random unobserved factor that affects student achievement. To correct for this potential bias, we adjust the standard deviation downward using an estimate of the variance of the random factors that we derive from the standard errors of the principal fixed effect estimates.[7] Following

Aaronson, Barrow, and Sander (2007) and Jacob and Lefgren (2005 and 2008), assume the estimated principal effect is the sum of the true effect plus mean-zero, independent, normally distributed random factors $\hat{\delta}_p = \delta_p + \nu_p$. Due to independence, the variance of the estimated principal effect is $\sigma^2_{\hat{\delta}} = \sigma^2_{\delta} + \sigma^2_{\nu}$. We obtain the true principal effect variance by subtracting the variance of the random factors from the variance of the estimated principal effect. Because σ^2_{ν} is unknown, we estimate it by taking the average of the square of the standard errors of $\delta_{p(i,h)}$.[7] Estimating the variance of the random factors using the OLS standard errors is appropriate because those standard errors are based on all random unobserved factors that affect student achievement but are uncorrelated with principal effects, which includes sampling error, disruptive students, cohort-specific school and school-by-grade shocks. We provide two different adjusted standard deviations: one based on σ^2_{ν} computed from unclustered standard errors, and another based on multi-way clustered errors across principals and schools.[8] We also check the robustness to other adjustment methods used in the literature, and to the method of Coelli and Green (2012), which provides a direct estimate of the lower bound of the variance.[9]

5. Data and Analysis Sample

5.1 Data Sources

Our main datasets are three linked files obtained from the Ministry of Education in BC. The first file contains observations on all students writing the FSA test in the period 1999 through 2010. For these students, we know the percentage score on each test, the school in which the test was written. Student scores are linked over time via an encrypted student identifier.

Student test scores are linked via the student identifier to a file containing the administrative records of all public school students in BC from 1999–2010 in grades 4 to 7. These data include information on gender, aboriginal status, participation in Special Education or English as a Second Language (ESL) programs, and each student's residential postal code. We use postal codes to link Dissemination/Enumeration Area (DA) level information from the Canada Census as proxies for students' socio-economic status (SES). We attach information on household income, average dwelling value, education levels, unemployment rates, ethnic and immigrant composition, and the age distribution of each DA.

The third file contains information on all public schools in British Columbia from 1995 through 2010. For each school we know what grades are offered, the number of teachers, and each school's exact address and postal code. Most importantly, we also know the name of the principal, which we use to create principal fixed effects in the econometric specification. To this dataset we append principal names and school information dating back to 1989, as transcribed from the *Public and Independent Schools Book*, a document produced by the Ministry of Education in BC.

5.2 Descriptive Statistics on Principals

We begin by analyzing descriptive statistics for the principals in our data, with a focus on mobility, which is key for our estimation strategy. In this section, we restrict our attention to the years 1999–2010 and we also restrict the sample to schools that offer grade 7 or lower grades to keep the focus on the student population of interest. Table 1 contains basic statistics about the number, mobility and experience of principals in BC. Between 1177 and 1295 principals are employed each year, and between 99 and 167 of these principals are newly hired into the system.

Between 69 and 78 percent of principals stay in the same school from one year to the next. Nine to 16 percent moved to a different school within British Columbia and 9–17 percent leave the sample. This latter group includes those principals who retired, those who moved out of province, those who move to a non elementary school, and those who stop working as a principal but remain in the system in some other capacity.

Each year, roughly one quarter of the sampled principals are in their first year of tenure at a school, and another quarter have remained at the same school for five or more years. The other half have remained at the same school for between two and four years. This shows considerable turnover of principals across schools, which helps our estimation strategy because we need multiple principals at each school to be able to identify the principal fixed effects. Finally, about one third of principals are in their first three years of employment as a principal, and another third have been a principal for 10 or more years.

5.3 Regression Sample

Because we estimate a value-added model in test scores, we restrict our focus to students who wrote both the fourth- and seventh-grade mathematics and reading exams. We observe test scores between 1999 and 2010, and because three years pass between each test, most such students are observed between 2002 and 2010. There are 397,619 students observed in the grade 7 test data between 2002 and 2010. We drop 48,051 (12.1 percent) who are not observed in all years between grades 4 and 7, 56,666 students (14.3 percent) who do not have a valid test score in both years, 5,831 students (1.5 percent) who are in schools that have fewer than ten test takers, and a final 809 observations (0.2 percent) who progress irregularly through the grades or are missing demographic information. From the remaining 286,262 students, our regression sample excludes an additional 124,909 students (43.6 percent of the remaining 286,262) who switch schools between grades 5 and 7, leaving a total of 161,353 students in our estimation sample.[10]

The summary statistics in Table 2 regarding test scores, demographics, and neighborhood census characteristics are based on this sample of 161,353 students. The fourth- and seventh-grade math and reading test scores are standardized to have a mean of zero and standard deviation of one in the population.

As outlined in Section 4, in the school fixed effects model, principal effects are only identified for a subset of principals. In Table 3 we compare sample means of observable characteristics for identified and unidentified principals from the regression sample. Column 1 pools all principals, and columns 2 and 3 restrict the sample to identified principals and unidentified principals, respectively. On average, the identified principals have slightly more experience and tenure in schools. The other differences are that unidentified principals are more likely to be in rural areas (where it is more difficult to move between schools), which in turn means the percent visible minority, immigrant, and ESL are lower. These principals also tend to belong to schools with slightly lower test scores.

6. Principal Mobility

In our preferred specification, principal effects within schools are estimable because of principal mobility. The identifying assumption is that mobility is exogenous, conditional on all the variables that appear in our model. In particular, if principals prefer schools located in high-income areas, or prefer schools with certain fixed attributes, this will not affect the causal interpretation. Rothstein (2010) shows that such random assignment is tenuous in the context of identifying *teacher* effects, because, for example, students may be sorted into classrooms based on lagged test score gains.[11] The identifying assumption, however, is much less tenuous in the context of *principal* effects. First, identification of a principal effect is not compromised when students are non-randomly assigned to teachers within a school; indeed, we posit that this is likely one of the mechanisms driving the principal effect. Second, the sorting of principals across schools is apt to be driven by school factors that are relatively fixed over the time period we consider in this analysis, and because we control for fixed characteristics, this does not compromise identification. Finally, we control for many time-varying observed characteristics that may be related to principal mobility.

In spite of the above, one may still worry that principal mobility is endogenous and we may mistakenly attribute a subsequent improvement in scores after a principal leaves a school to the incoming principal when it fact was just an improvement due to the scores reverting to the mean. If principal mobility was related to low school scores, our estimates may be biased upwards. Conversely, if mobility was related to high school scores, our estimates may be biased downwards.

The administrative rules governing principal mobility in BC are discussed in detail in Coelli and Green (2012). They find that mobility is determined at the school district level and is generally overseen by the Superintendent of Schools. Of the 60 districts in BC, 14 are found to mention principal rotation somewhere in their district policies; these districts are larger than average in size. The main example of a district policy towards rotation comes from the Vancouver School Board, whose written procedure is to transfer principals between schools roughly every five years. Other districts are sometimes less precise, stating in their district policies the minimum and maximum times until principals transfer, leaving much discretion to the school district. Smaller districts that do not have written rotation procedures are found not to have included them in the district policy because there are too few schools to make such a strategy work.

In Table 4 we empirically evaluate the determinants of principal mobility. We take a sample of all schools that offer the grade 7 FSA test between 1999 and 2010, and use a regression to relate principal mobility between those schools to observable characteristics of the principals and schools. We regress an indicator for a principal move between t and t+1 on school and neighbourhood characteristics at time t (and in some specifications, t-1 or t-2). The dependent variable equals 1 only if the principal moves from one school to another school in the sample.[12] Column 1 includes a full set of controls and shows no evidence that principal turnover is related to the level of math and reading test scores. Note in particular that the test score coefficients are very small and imprecise.[13] Column 2 adds lagged test scores, and there continue to be no significant relationships. Column 3 uses gain scores instead of level scores as principal mobility may be related to test score valued added from grade 4. Again there is no significant relationship. Column 4 uses the absolute value of the test score gains and finds no significant relationships. Columns 5 and 6 use a measure of whether the test score level significantly declines or improves in the past year or two years, respectively. Here we find a significant relationship between having large gains for two years in a row on level scores and principal mobility. Columns 7 and 8 repeat Columns 5 and 6 but use gain scores instead of level scores. Using gain scores, there is a marginal significant relationship between large one year loss or large gains for two years in gain scores and principal turnover.[14]

The above results tell us that, while we cannot be certain that mobility is unrelated to unobservable factors that change during our time period, there is very little evidence showing that school performance determines moves. In addition, this speaks to the issue of mean reversion as there is very little evidence that there is a decrease or increase of test scores before a principal leaves a school, and therefore, there is little evidence that mean reversion would affect our value-added estimates.

Another interesting exercise is to examine the characteristics of schools that switch principals over the course of their career. Table 5 shows all the principals who appear in our data from 1999-2010, divided into two groups. The first group includes the principals who switch from their first school in their career as a principal to their second school in their career as a principal. The second group comprises the principals who switch from their second school to their third school. Next we examine the attributes of the first school at which the principal was employed, measured the year before the principal arrived, along with the attributes of the second school at which the principal was employed, again measured the year before the principal arrived. This is repeated for the second and third schools. In Panel A we use all schools in our sample and display the mean characteristics of the two schools along with the difference. We find that principals move to schools that are more urban and located in neighbourhoods with higher SES. This pattern is also found for the second to third school switch. Panel B examines only schools that are not rural to see if the pattern holds up for urban to urban school switches. Here we find similar patterns as in Panel A.

7. Results

7.1 Variation in Principal Quality

Table 6 presents the baseline estimates from equation 4, and includes details on the distribution of $\delta^{g}_{p(i,h)}$ including the standard deviation, associated standard error, along with the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles. Both the standard deviation and the two adjusted standard deviations (derived from estimates of sampling error based on unclustered and multi-way clustered errors by principal and school) are included, in addition to the gap between the 75^{th} percentile and the median principal fixed effect. The latter measure is included as a representation of the difference between a "good" principal and an average principal, and because it is a measure of dispersion that is potentially less sensitive to outliers than the standard deviation. The number of principals and the number of schools used in the regression are listed along with the p-value for the F-test of the joint significance of $\delta^{g}_{p(i,h)}$ (i.e., $\delta^{g}_{p(i,h)} = 0$ for all p).

Columns 1 and 5 present estimates of the distribution of the principal fixed effects from a specification which includes no demographic or census control variables, and no school fixed effects. This specification essentially computes the deviation of each principal's average score across students, from the mean principal. The estimates in column 1 imply that moving one standard deviation up the distribution of principals increases test scores over a three-year period by 0.323 standardized units in math. The standard error of the standard deviation is 0.001.[15] After adjusting our estimates for sampling error, the standard deviation drops to 0.322 with clustering and 0.307 without clustering, implying that only a small fraction of the estimate is due to sampling error. Comparing the principal at the 75^{th} percentile to the median leads to a more modest increase of 0.211 standardized student points. In addition, we also list the fourth grade lagged score coefficient. In column 1, this coefficient is 0.616, which indicates that past achievement is a good predictor for current achievement. We find slightly smaller results in column 5, which shows the results from the reading test. The F-test of the joint significance of

$\delta^{g}_{p(i,h)}$ rejects the null hypothesis of no principal fixed effects at a high level of significance in both cases. Columns 2 and 6 add demographic and census control variables.[16] Estimates shrink slightly, but the story is largely unchanged.

In Columns 3 and 7 we drop the principal effects that would be unidentified in a school fixed effects specification, to assess the effect of limiting the sample of principals before explicitly including school fixed effects. The standard deviation and gap between the 75th percentile and the median shrink compared to columns 2 and 6. This implies that our sample of principals that are identified in a school fixed effect framework have less variation than principals who are not identified.

Columns 4 and 8 report estimates when the principal 5/6 by school fixed effects are added to the model. The estimated adjusted standard deviations using clustered standard errors are 0.408 and 0.289 for math and reading, respectively.[17] The difference in the 75th percentile and the median also increases to 0.193 in math and 0.170 in reading. While it may seem counterintuitive that the standard deviation rises after including fixed effects, this occurs because the principal and school effects are negatively correlated, which is to say that high value added principals tend to be in low value added schools.[18]

The results in Table 6 show that overall, individual principals can have a substantial impact on student achievement in reading and math.[19] How large are these effects? One first needs to note that we estimate gains over a three-year period for an individual principal, whereas most of the literature estimates gains over one year. To interpret the size of our estimates against other studies, the three-year time period must be kept in focus. It is also important to note that although the standard deviation is the most widely used statistic in this literature, it is not necessarily the most intuitive. The difference between the 75th percentile and the median is a potentially more informative way to gauge the difference between a good and an average principal, and is a more robust statistic in the sense that it is less sensitive to outliers. Indeed, in Table 6 we see less variation across specifications in the difference between the 75th percentile and the median than we do with the standard deviation.

For comparison, using a similar research design, Rockoff (2004) estimates teacher fixed effects using one-year test score gains. He estimates that the unadjusted standard deviation of estimated teacher effects is roughly 0.21 in reading and 0.29 in math, and the adjusted effect is 0.10 for reading and math. Also, Aaronson, Barrow, and Sander (2007) estimate a standard deviation of 0.13 for teacher effects on math scores in Chicago. Again, note that compared against these teacher effects, our magnitudes are larger in part because we measure the effect of principals on the gains over a three year period rather and as noted in the conceptual framework, there are many channels through which principals affect school quality and student learning.

7.2 Robustness Checks

To check the robustness of our results, Table 7 replicates the results of Table 6 using different specifications or different samples.[20] We report the standard deviation and gap between the 75th percentile and median, using only specifications that include school fixed effects. These estimates correspond to column 4 for math and column 8 for reading in Table 6.[21]

The estimates presented in Table 6 come from a model that includes grade 4 test scores as a control variable. This inclusion may come at a cost when estimating principal fixed effects because if a principal leads a school prior to a student taking their grade 4 test, they would have also impacted the grade 4 exams and therefore it might be harder for them to alter the gains of that student from grade 4 to grade 7. As can be seen in Panel A of Table 7, the estimates are roughly similar in magnitude to the main estimates if grade 4 exams are not included as controls.

We next check the robustness of our estimates by adding students who move between schools into the sample. In our main specifications, we only use students who do not change schools from grade 5 to 7. Therefore to check the robustness of our estimates to using a broader sample, we reconfigure the school fixed effects to be the interaction of the grade 5/6 principal and the grade 5/6/7 school (i.e., a 5-way interaction) so that we can properly control for the fixed characteristics of the schools the students attended and the former principals. As reported in Panel B, estimates of the standard deviation from this specification are higher than the corresponding adjusted standard deviation in Table 6, but the difference in the median and 75th percentile is only slightly different.[22]

Panel C presents estimates from a different specification in which we first estimate a grade 7 principal by school fixed effect and then demean these estimates by the school average. This specification compares principals only to other principals within the same school, which almost always amounts to comparing two or three individuals in each connected group. To demean by the school average, it was necessary to compute principal by school effects rather than pure principal effects.[23] In this specification, the principal by school effects are smaller than our main estimates. This indicates that the variation between principals who work in the same school is smaller than the variation between principals in the larger connected groups estimated from equation (1). This may be due to schools generally hiring similar quality principals or similar quality principals applying for positions at the same school.

The estimates presented in Table 6 and Panel A-C in Table 7 are summary statistics based on estimating a fixed effect for each principal.[24] An alternative approach, used by Coelli and Green (2012), is to estimate the lower bound of the variance in the principal effects directly, without estimating a fixed effect for each principal. We estimate this by regressing the variance in test scores across cohorts within each school on variable indicating principal turnover plus other control variables. The turnover term (explained in detail in Coelli and Green, 2012) will equal zero if one principal leads a school for the entire sample period. If there are multiple principals, this term will be positive and increasing in the number of principal turnovers.
The coefficient on the principal turnover can be interpreted as the within school variance of principal effects. In Panel D, we follow this estimation strategy using our sample. We find that the lower bound estimate of the standard deviation is 0.293 for math and 0.218 for reading and the adjusted standard deviation is 0.286 and 0.198 for math and reading, respectively.[25]

7.3 Principal Experience and Tenure

As previously mentioned, a principal's professional experiences affect the principal and this can influence the other factors such as school and classroom conditions. It might be expected that more experienced principals would influence these conditions in a different way as compared to less experienced principals, and that this in turn would affect teachers and student learning. In addition, tenure in a school may be related to changes in school and classroom conditions due to having more time to make changes to these conditions (e.g., school culture, community, or nature of instruction).

In our regression analyses, we separately analyze the effect of principal tenure and principal experience on student performance. We model tenure with a set of dummy variables for three years, four years, and any amount of tenure five years and over, with one to two years of tenure as the excluded group. We model experience using dummies for groups of four to six, seven to ten, and ten or more years, with one to three years as the excluded group.

Panel A of Table 8 reports the results from specifications that include our measures of principal tenure. Columns 1, 3, 5 and 7 display estimates when school fixed effects are not included in the model. Columns 1, 2, 5, and 6 all include fourth grade scores and columns 3, 4, 7, and 8 do not include fourth grade scores as a control variable. We include both specifications because the longer a principal is at a particular school, the more likely to are to be leading that school when students are in fourth grade. A good principal may boost fourth grade scores for students, but then may find it more difficult to produce gains in scores from grade 4 to grade 7. Therefore, it is important to check our estimates for robustness by not including grade 4 test scores as controls.

In all specifications in Panel A, we find little evidence of any statistically significant relationship. Including school fixed effects, and removing grade 4 test scores as a control does little to change the pattern across years of tenure. One explanation for this finding is that we are measuring the average effect of tenure across all principals. It could be that poor decisions made by a principal cumulate over time, and likewise for good decisions. On average, we might then observe no effect of tenure on student outcomes. Panel B of Table 8 presents the results of using principal experience rather than tenure. Across all specifications we find a slightly negative effect of additional experience, but none of the estimates are statistically significant.

In Table 8 we include linear tenure effects. However, in Table 9 we further explore tenure effects by separately estimating the variance in principal effects among principals with 3 or fewer years of tenure, and principals with more than 3 years. From those results, it is clear that there is no substantial difference in the standard deviations of the principal effects whether principals have been at the school for 3 years or less versus more than 3 years. This is consistent with the results in Branch et. al. (2012), who show that restricting the sample to principals with 3 or fewer years of tenure produces a principal effect variance that is not very different from the one they produce with all principals. However, this is different than what is found by Coelli and Green (2012) as they find that tenure is important when examining high school principals.

These results from Table 8 and 9 are interesting as they suggest that neither school tenure nor experience for an average principal matters very much. This is important from a policy perspective since it suggests that if the goal is to boost performance of underperforming students, and reduce achievement gaps, a more effective approach would be to identify high-ability principals and then allocate them to schools, rather than relying on a principal to boost performance by gaining experience.

8. Conclusion

We estimate the impact of fixed principal characteristics on student performance. Results show that principals have a substantial impact on both math and reading scores. Our main results show that a one standard deviation shift up the principal quality distribution can increase achievement by approximately 0.289 - 0.408 standard deviations in math and reading, while shifting to the 75th percentile improves scores by 0.170 - 0.193 relative to the median principal. We also estimate that a principal with more tenure or with more experience overall has little effect on student achievement.

These results have important implications for policy. The main implication is that shifting principals between schools has the potential to significantly reduce achievement gaps. However, the optimal reassignment of existing principals, with respect to a gap-minimizing objective function, might affect the supply of effective principals. Some of this improvement depends on where the principal works, but a sizable portion is portable across schools. Policy makers could identify the most effective principals, and allocate them between schools to potentially reduce achievement gaps. Alternatively, researchers could identify the best principals and dig deeper to learn more about the attributes that make them so effective. Information gleaned from that exercise could be used to train underperforming principals. What are the fixed characteristics that make one principal better than another? We cannot identify these in our data, but prior literature can shed some light on this issue. The results of Clark, Martorell, and Rockoff (2009) show that education level and pre-principal experience are not likely to be among the factors, while the evidence on training programs is mixed. Principals are most likely to affect student outcomes via their policies. Figlio and Sass (2010) show that when a new principal enters a school, that individual is most likely to change the policies on teacher incentives, curriculum, and those which boost performance of low-achieving students. They further show that the most effective principals focus on policies that boost performance of low-achieving students. Finally, Jacob (2010) shows that principals are likely instrumental in the hiring and firing decisions regarding teachers, which may trickle down to affect student achievement.

9. References

Aaronson, Daniel, Lisa Barrow, and William Sander (2007) 'Teachers and student achievement in the Chicago Public High Schools,' *Journal of Labor Economics* 25, 95–135

Ballou, Dale, and Michael J. Podgursky (1993) ‘What makes a good principal? How teachers assess the performance of principals,’ *Economics of Education Review* 14, 243–52

Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin (2012) ‘Estimating the effect of leaders on public sector productivity: The case of school principals,’ National Bureau of Economic Research Working Paper No. 17803, NBER, Cambridge, MA

Brewer, Dominic J. (1993) ‘Principals and student outcomes: Evidence from U.S. high schools,’ *Economics of Education Review* 12, 281–93

Cameron, A. Colin, Jonah B. Gelbach, and Douglas Miller (2011) “Robust inference with Multi-Way Clustering,” *Journal of Business and Economic Statistics* 29(2) 238-49

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2011) ‘The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood,’ National Bureau of Economics Research Working Paper No. 17699. NBER, Cambridge, MA

Clark, Damon, Paco Martorell, and Jonah Rockoff (2009) ‘School principals and school performance,’ The Urban Institute Working Paper No. 38, Washington, DC

Coelli, Michael, and David Green (2012) ‘Leadership effects: School principals and student outcomes,’ *Economics of Education Review* 31, 92-109

Corcoran, Sean P., Amy E. Schwartz, and Meryle Weinstein (2012) ‘Training your own: The impact of New York City’s Aspiring Principals Program on student achievement,’ *Education Evaluation and Policy Analysis* 34 (2), 232-53

Dhuey, Elizabeth (2013)’Middle school or junior high? How grade level configurations affect student achievement,’ *Canadian Journal of Economics* 46(2), 469-96

Eberts, Randall W., and Joe A. Stone (1988) ‘Student achievement in public schools: Do principals make a difference?’ *Economics of Education Review* 7, 291–99

Figlio, David, and Tim Sass (2010) ‘The policy choices of effective principals’ Unpublished manuscript

Jacob, Brian A. (2010) ‘Do principals fire the worst teachers?’ National Bureau of Economic Research Working Paper No.15715, NBER, Cambridge, MA

Jacob, Brian, and Lars Lefgren (2005) ‘Principals as agents: subjective performance measurement in education’ National Bureau of Economic Research Working Paper No. 11463, NBER, Cambridge, MA

Jacob, Brian, and Lars Lefgren (2008) ‘Can principals identify effective teachers? Evidence on subjective performance evaluation in education’ *Journal of Labor Economics* 26, 101–36

Kane, Thomas J., and Douglas O. Staiger (2002) 'Volatility in school test scores: Implications for test-based accountability systems,' In *Brookings Papers on Education Policy 2002*, ed. Diane Ravitch, 235–69 (Washington, DC: Brookings)

Leithwood, Kenneth, Karen Seashore Louis, Stephen Anderson, and Kyla Wahlstrom (2004) 'How leadership influences student learning' Wallace Foundation Learning from Leadership Project Report, New York, NY

Mihaly, Kata, Daniel McCaffrey, J. R. Lockwood, and Tim R. Sass (2010) 'Centering and reference groups for estimates of fixed effects: Modifications to felsdvreg,' *The Stata Journal* 10, 82–103

Rockoff, Jonah E. (2004) 'The impact of individual teachers on student achievement: Evidence from panel data,' *American Economic Review* 94, 247–52

Rothstein, Jesse (2010) 'Teacher quality in educational production: Tracking, decay, and student achievement,' *The Quarterly Journal of Economics* 125, 175-214

Table 1
Principal characteristics in British Columbia, 1999-2010

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total Employed	1280	1291	1295	1272	1238	1234	1231	1233	1218	1205	1185	1177
New Hires	124	123	134	99	109	132	167	160	164	146	110	124
Mobility												
Same school	75%	78%	77%	72%	73%	73%	71%	69%	73%	78%	73%	
Different school	16%	12%	12%	16%	15%	13%	15%	14%	13%	9%	11%	
Out of sample	9%	10%	10%	12%	12%	14%	14%	17%	14%	13%	15%	
Tenure at a school												
1 yr	21%	25%	23%	21%	24%	26%	27%	29%	29%	26%	20%	24%
2 yrs	18%	19%	23%	21%	18%	21%	22%	22%	24%	26%	24%	17%
3 yrs	20%	14%	15%	20%	17%	15%	17%	17%	17%	18%	22%	20%
4 yrs	13%	15%	10%	12%	15%	12%	11%	11%	11%	12%	14%	16%
5+ yrs	28%	27%	29%	27%	25%	26%	23%	21%	19%	19%	20%	23%
Experience as principal												
1 to 3 yrs	28%	26%	27%	26%	26%	26%	31%	34%	36%	35%	33%	29%
4 to 6 yrs	25%	30%	23%	23%	22%	24%	21%	20%	20%	24%	27%	29%
7 to 9 yrs	17%	12%	19%	19%	23%	18%	18%	16%	16%	14%	14%	16%
10+ yrs	31%	31%	31%	31%	29%	32%	30%	30%	29%	27%	26%	26%

Notes: Numbers in this table are based on 11,342 principal-year observations, 2939 principals, and 1488 schools. The sample includes all schools in British Columbia that offer grade 7 or below. Some principals are observed in multiple schools in the same year in certain rural locations. For these cases, we kept the principal's "main" school, which we define as the school in which the principal is observed most often. New hire is defined as a principal who is observed in the data for the first time as a principal.

Table 2
Descriptive statistics for analysis sample

	Mean	Std dev.
4th grade test scores		
Math	0.059	0.977
Reading	0.062	0.955
7th grade test scores		
Math	0.122	0.982
Reading	0.113	0.951
Demographics		
Male	0.503	0.500
Aboriginal	0.083	0.277
ESL	0.267	0.442
In special education	0.050	0.217
Attends rural school	0.131	0.337
Neighbourhood census characteristics		
Percent over 65 yrs old	0.131	0.086
Percent visible minority	0.274	0.276
Percent immigrant	0.285	0.189
Percent unemployed	0.053	0.044
Percent no high school degree	0.215	0.094
Percent high school degree	0.280	0.062
Percent university degree	0.223	0.125
Average household income	64,755	26,909
Average dwelling value	354,988	207,479
Number of principals	1468	
Number of schools	823	
Number of observations	161,353	

Notes: Test scores represent the z-score by year, grade, and subject among the population of students prior to sample exclusions. Neightbouhood census characteristics are the attributes of the Dissemination area surrounding a child's home postal code.

Table 3

Average principal, school, and neighbourhood characteristics for identified and unidentified principals

	All principals (1)	Identified principals (2)	Unidentified principals (3)
Principal characteristics			
Principal tenure	3.104	3.191	2.796
Principal experience	6.680	6.981	5.617
Number of years in data	3.920	4.340	2.440
Number of principals	1468	1144	324
School characteristics			
7th grade math score	0.015	0.036	-0.061
7th grade reading score	-0.006	0.007	-0.050
Did not take 7th grade math test	0.107	0.108	0.105
Did not take 7th grade reading test	0.103	0.104	0.101
Percent male	0.514	0.515	0.507
Percent aboriginal	0.140	0.128	0.183
Percent ESL	0.262	0.276	0.211
Percent special education	0.101	0.099	0.107
Rural school	0.210	0.167	0.362
Neighbourhood census characteristics			
Percent over 65 yrs old	0.132	0.131	0.136
Percent visible minority	0.222	0.243	0.149
Percent no high school degree	0.229	0.225	0.244
Percent high school degree	0.279	0.279	0.279
Percent university degree	0.205	0.212	0.179
Percent immigrant	0.249	0.264	0.194
Percent unemployed	0.060	0.058	0.067
Average household income	60,914	62,670	54,713
Average dwelling value	316,734	331,662	264,023

Notes: Data based on regression sample of 161,353 observations. Average school characteristics and average neighbourhood characteristics are computed by averaging all students observed in grade 7 in a school. The number of connected groups for the identified principals is 244.

Table 4

Estimates of the relationship between principal turnover on school and neighbourhood characteristics using school to school switches

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
7th grade math score	0.021	0.009						
	(0.020)	(0.023)						
7th grade reading score	-0.011	-0.001						
	(0.023)	(0.026)						
Lagged 7th grade math score		-0.023						
		(0.019)						
Lagged 7th grade reading score		0.006						
		(0.023)						
4th to 7th grade math gain scores			0.007					
			(0.014)					
4th to 7th grade reading gain scores			0.002					
			(0.015)					
Absolute value of math gain scores				-0.001				
				(0.020)				
Absolute value of reading gain scores				-0.019				
				(0.022)				
Large one year gain in level scores					-0.010			
					(0.014)			
Large one year loss in level scores					-0.010			
					(0.014)			
Large gains for two years in level scores						0.071**		
						(0.032)		
Large losses for two years in level scores						0.041		
						(0.032)		
Large one year gain in gain scores							-0.013	
							(0.014)	
Large one year loss in gain scores							-0.026*	
							(0.014)	
Large gains for two years in gain scores								0.050*
								(0.030)
Large losses for two years in gain scores								0.046
								(0.033)
R squared	0.021	0.021	0.018	0.018	0.028	0.032	0.024	0.033
Number of observations	8519	7302	6844	6844	5292	4426	5681	4672

Notes: ***p<0.01; **p<0.05; *p<0.1. Data for this regression are at the school by year level, and include all schools that offer the grade seven numeracy and reading tests from 1999-2010. A small number of principals who serve at two schools in the same year are exluded from the regression. School to school switch refers to all principal mobility from one school in the sample to another. All models include both year and school fixed effects. See Section 7.1 for a description of the demographic and census variables included in the regressions. In addition, indicator variables for four to six years experience, seven to nine years experience, ten+ years of experience, school enrollment, and number of elementary teachers in school are also included.

Table 5
Characteristics of schools by principal school career

	Switch 1st to 2nd			Switch 2nd to 3rd		
	1st school	2nd school	Difference	2nd school	3rd school	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All schools						
4th grade math score	-0.029	-0.003	0.026	-0.073	-0.029	0.044
4th grade reading score	-0.051	-0.027	0.024	-0.088	-0.033	0.055
Male	0.514	0.513	-0.001	0.512	0.517	0.006
Aboriginal	0.168	0.140	-0.029*	0.162	0.124	-0.038**
ESL	0.216	0.254	0.038*	0.243	0.267	0.024
In special education	0.083	0.085	0.001	0.080	0.068	-0.012*
Attends rural school	0.248	0.164	-0.085***	0.141	0.123	-0.018
Percent over 65 yrs old	0.119	0.124	0.005	0.123	0.135	0.011**
Percent visible minority	0.170	0.217	0.047***	0.194	0.238	0.044**
Percent immigrant	0.223	0.247	0.024**	0.235	0.261	0.026*
Percent unemployed	0.079	0.060	-0.019***	0.077	0.059	-0.018***
Percent no high school degree	0.265	0.225	-0.041***	0.262	0.225	-0.037***
Percent high school degree	0.266	0.282	0.015***	0.266	0.281	0.014***
Percent university degree	0.177	0.203	0.026***	0.186	0.205	0.019*
Average household income	54,371	60,619	6247***	55,077	59,455	4378**
Average dwelling value	238,556	307,557	69001***	243,510	313,708	70198***
Number of principals	330	330		227	227	
Panel B: Urban schools						
4th grade math score	0.048	0.026	-0.022	-0.065	-0.030	0.035
4th grade reading score	0.003	-0.005	-0.008	-0.072	-0.057	0.015
Male	0.507	0.517	0.009	0.511	0.523	0.011
Aboriginal	0.112	0.100	-0.011	0.141	0.116	-0.025
ESL	0.232	0.303	0.071***	0.274	0.315	0.041
In special education	0.074	0.075	0.001	0.069	0.066	-0.003
Percent over 65 yrs old	0.119	0.123	0.004	0.120	0.128	0.008
Percent visible minority	0.218	0.280	0.062***	0.230	0.283	0.053**
Percent immigrant	0.260	0.295	0.035**	0.260	0.291	0.031*
Percent unemployed	0.064	0.052	-0.011***	0.074	0.057	-0.017***
Percent no high school degree	0.245	0.210	-0.035***	0.254	0.218	-0.035***
Percent high school degree	0.267	0.281	0.015***	0.268	0.283	0.014***
Percent university degree	0.200	0.226	0.026***	0.199	0.219	0.020*
Average household income	59,317	63,911	4593***	58,210	60,907	2696
Average dwelling value	267,625	342,482	74857***	259,806	331,794	71988***
Number of principals	239	239		184	184	

Notes: ***p<0.01; **p<0.05; *p<0.1. Test scores represent the z-score by year, grade, and subject among the population of students prior to sample exclusions. Rural schools are those whose postal codes contain a zero in the second digit. Average school characteristics and average neighbourhood characteristics are computed by averaging all students observed in grade 7 in a school.

Table 6
Student level estimates of principal fixed effects, 4th to 7th grade gains

	Math				Reading			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard deviation	0.323	0.297	0.271	0.415	0.234	0.227	0.200	0.303
Standard error	0.001	0.004	0.003	0.010	0.001	0.004	0.002	0.016
Adjusted standard deviation	0.307	0.278	0.256	0.358	0.209	0.202	0.178	0.213
Adjusted standard deviation with clustered SE's	0.322	0.293	0.267	0.408	0.233	0.224	0.197	0.289
10th percentile	-0.390	-0.346	-0.331	-0.454	-0.272	-0.261	-0.226	-0.347
25th percentile	-0.207	-0.177	-0.176	-0.215	-0.131	-0.124	-0.110	-0.164
50th percentile	-0.005	-0.001	0.001	0.008	0.013	0.014	0.008	-0.009
75th percentile	0.206	0.172	0.161	0.202	0.150	0.138	0.122	0.161
90th percentile	0.389	0.346	0.327	0.508	0.253	0.248	0.220	0.327
(75th percentile - median)	0.211	0.173	0.160	0.193	0.136	0.124	0.115	0.170
4th grade score coefficient	0.616	0.603	0.603	0.615	0.615	0.593	0.593	0.599
Number of principals	1468	1468	1144	1144	1468	1468	1144	1144
Number of schools	823	823	823	823	823	823	823	823
P-value on F-test	0	0	0	0	0	0	0	0
<u>Control variables:</u>								
Demographics	no	yes	yes	yes	no	yes	yes	yes
Census	no	yes	yes	yes	no	yes	yes	yes
School fixed effects	no	no	no	yes	no	no	no	yes

Notes: Figures in this table represent the distribution among all 1468 or 1144 principal effects estimated via equation 4. See Section 7.1 for a description of the demographic and census variables included in the regressions. Coefficients are measured in student level standard deviations. The p-value on the F-test is the test for the joint significance of all principal fixed effects in the model, using unclustered errors. Standard errors are computed using the Delta Method and clustered errors. Clustering is across grade 7 principals and 3-way interaction school effects.

Table 7
Robustness checks of estimates of principal fixed effects, 4th to 7th grade gains

	Math (1)	Reading (2)
Panel A: No grade 4 exam controls		
Standard deviation	0.413	0.334
Standard error	0.008	0.013
Adjusted standard deviation	0.312	0.196
Adjusted standard deviation with clustered SE's	0.403	0.320
(75th percentile - median)	0.206	0.164
Number of principals	1144	1144
Number of schools	823	823
Panel B: Includes students who move and grade 5 and 6 school fixed effects		
Standard deviation	0.529	0.381
Standard error	0.018	0.016
Adjusted standard deviation	0.477	0.298
Adjusted standard deviation with clustered SE's	0.481	0.305
(75th percentile - median)	0.216	0.176
4th grade score coefficient	0.609	0.602
Number of principals	1356	1356
Number of schools	944	944
Panel C: Principal by school effects demeaned by school average		
Standard deviation	0.225	0.200
Standard error	0.003	0.002
Adjusted standard deviation	0.185	0.150
Adjusted standard deviation with clustered SE's	0.222	0.197
(75th percentile - median)	0.144	0.107
4th grade score coefficient	0.615	0.599
Number of principals*schools	1579	1579
Number of schools	823	823
Panel D: Coelli and Green (2012) estimates		
Standard deviation	0.293	0.218
Standard error	0.019	0.026
Adjusted standard deviation	0.286	0.198
Standard error	0.019	0.026
Number of principals	1468	1468
Number of schools	823	823
Control variables:		
Demographics	yes	yes
Census	yes	yes
School fixed effects	yes	yes

Notes: See Section 7.1 for a description of the demographic and census variables included in the regressions. Panels A, C, and D use the main sample described in Table 2. See Appendix Table 3 for a summary statistics on the sample used for Panel B. Standard errors are computed using the Delta Method, except for those Panel B which are bootstrapped using a simple nonparametric bootstrap, resampling from the set of principal fixed effects. Clustering is across grade 7 principals and 3-way interaction school effects.

Table 8
Regression coefficients for tenure and experience measures

	Math				Reading			
	With grade 4 scores		Without grade 4 scores		With grade 4 scores		Without grade 4 scores	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Years of tenure								
3 yrs school tenure	-0.005	-0.015	0.007	-0.017	0.008	-0.015	0.009	0.000
	(0.011)	(0.052)	(0.012)	(0.061)	(0.009)	(0.035)	(0.009)	(0.052)
4 yrs school tenure	0.006	-0.033	-0.007	-0.059	0.003	-0.024	-0.009	-0.027
	(0.014)	(0.048)	(0.015)	(0.058)	(0.012)	(0.033)	(0.012)	(0.050)
5+ yrs school tenure	0.006	-0.045	0.005	-0.073	0.003	-0.025	0.005	-0.025
	(0.016)	(0.047)	(0.018)	(0.058)	(0.013)	(0.032)	(0.014)	(0.050)
Panel B: Years of experience								
4 to 6 yrs principal experience	0.005	-0.007	-0.014	-0.031	0.003	-0.003	-0.009	-0.013
	(0.016)	(0.027)	(0.017)	(0.027)	(0.013)	(0.023)	(0.014)	(0.025)
7 to 9 yrs principal experience	0.006	-0.011	-0.010	-0.180	-0.012	-0.028	-0.016	-0.019
	(0.023)	(0.046)	(0.025)	(0.048)	(0.018)	(0.037)	(0.020)	(0.041)
10+ yrs principal experience	-0.011	-0.016	-0.016	-0.005	-0.012	-0.037	-0.019	-0.030
	(0.029)	(0.060)	(0.032)	(0.061)	(0.023)	(0.049)	(0.025)	coo
Control variables:								
Principal fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Demographics	yes	yes	yes	yes	yes	yes	yes	yes
Census	yes	yes	yes	yes	yes	yes	yes	yes
School fixed effects	no	yes	no	yes	no	yes	no	yes

Notes: ***p<0.01; **p<0.05; *p<0.1 Coefficients in this table correspond to the models in Table 7. Coefficients are measures in standard deviations. See Section VII.A. for a full list of demographic and census measures. Standard errors are clustered by principal and school.

Table 9
Principal by Tenure Fixed effects

	Math	Reading
	(1)	(2)
Panel A: 3 years of tenure or less		
Mean	0.009	-0.008
Standard deviation	0.404	0.343
number of principals x tenure	1403	1403
Panel B: More than 3 years of tenure		
Mean	0.000	0.006
Standard deviation	0.407	0.299
number of principals x tenure	1670	1670
Control variables:		
Demographics	yes	yes
Census	yes	yes
School fixed effects	yes	yes

Notes: See Section 7.1 for a description of the demographic and census variables included in the regressions. Separate principal effects were estimated for those with 3 or fewer years of tenure, and more than 3 years of tenure. All principal effects are centered to have mean zero among connected groups

Appendix Table 1

Regression of principal turnover on school and neighbourhood characteristics on any switches

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
7th grade math score	-0.001	-0.006						
	(0.023)	(0.026)						
7th grade reading score	-0.014	-0.029						
	(0.027)	(0.030)						
Lagged 7th grade math score		-0.018						
		(0.025)						
Lagged 7th grade reading score		0.022						
		(0.029)						
4th to 7th grade math gain scores			-0.002					
			(0.019)					
4th to 7th grade reading gain scores			-0.010					
			(0.020)					
Absolute value of math gain scores				0.010				
				(0.026)				
Absolute value of reading gain scores				-0.028				
				(0.028)				
Large one year gain in level scores					-0.018			
					(0.018)			
Large one year loss in level scores					-0.002			
					(0.017)			
Large gains for two years in level scores						0.064*		
						(0.039)		
Large losses for two years in level scores						0.115***		
						(0.042)		
Large one year gain in gain scores							-0.015	
							(0.018)	
Large one year loss in gain scores							-0.010	
							(0.018)	
Large gains for two years in gain scores								0.053
								(0.039)
Large losses for two years in gain scores								0.101**
								(0.043)
R squared	0.052	0.056	0.056	0.056	0.071	0.082	0.067	0.080
Number of observations	8519	7302	6844	6844	5292	4426	5681	4672

Notes: ***p<0.01; **p<0.05; *p<0.1. Data for this regression are at the school by year level, and include all schools that offer the grade seven numeracy and reading tests from 1999-2010 A small number of principals who serve at two schools in the same year are exluded from the regression. Any switches refers to school switches plus switches to schools outside the sample or exits from the sample. All models include both year and school fixed effects. See Section 7.1 for a description of the demographic and census variables included in the regressions. In addition, indicator variables for four to six years experience, seven to nine years experience, ten+ years of experience, school enrollment, and number of elementary teachers in school are also included.

Appendix Table 2

Principal by school effects across different school characteristics

	Math		Reading	
	(1)	(2)	(3)	(4)
Panel A: Low socioeconomic status				
Mean	-0.031	0.000	-0.015	0.000
Standard deviation	0.294	0.232	0.228	0.199
number of principals	690	690	690	690
Panel B: High socioeconomic status				
Mean	0.024	0.000	0.012	0.000
Standard deviation	0.288	0.220	0.207	0.200
number of principals	889	889	889	889
Panel C: Low student achievement				
Mean	-0.029	0.000	-0.020	0.000
Standard deviation	0.298	0.235	0.220	0.195
number of principals	761	761	761	761
Panel D: High student achievement				
Mean	0.027	0.000	0.018	0.000
Standard deviation	0.282	0.216	0.213	0.203
number of principals	818	818	818	818
Control variables:				
Demographics	yes	yes	yes	yes
Census	yes	yes	yes	yes
School fixed effects	no	yes	no	yes

Notes: See Section 7.1 for a description of the demographic and census variables included in the regressions. High and low socioeconomic status is defined by being in above or below the median level of household income. High and low student achievement is defined as being above or below the median grade 4 test score across schools.

Appendix Table 3
Descriptive statistics for estimation sample in Table 7, Panel B

	Mean	Std dev.
4th grade test scores		
Math	0.049	0.977
Reading	0.057	0.960
7th grade test scores		
Math	0.024	0.977
Reading	0.065	0.961
Demographics		
Male	0.502	0.500
Aboriginal	0.098	0.298
ESL	0.223	0.416
In special education	0.054	0.226
Attends rural school	0.117	0.322
Neighbourhood census characteristics		
Percent over 65 yrs old	0.136	0.089
Percent visible minority	0.222	0.254
Percent immigrant	0.252	0.176
Percent unemployed	0.052	0.043
Percent no high school degree	0.214	0.090
Percent high school degree	0.283	0.061
Percent university degree	0.211	0.118
Average household income	62,754	24,520
Average dwelling value	332,705	188,224
Number of principals	1675	
Number of schools	944	
Number of observations	286,262	

Notes: Test scores represent the z-score by year, grade, and subject among the population of students prior to sample exclusions. Neightbouhood census characteristics are the attributes of the Dissemination area surrounding a child's home postal code.

Lead Footnote: We thank Kelly Bedard, Mati Dubrovinsky, David Johnson, Harry Krashinsky, Abigail Payne, Simon Woodcock, various seminar participants, and two anonymous referees for helpful comments.

---

[1] We run a robustness check in Table 7 that estimates the specification used by Coelli and Green (2012).

[2] Our main specification only includes students who do not switch schools. We also a provide robustness check, which includes students who do move schools during the sample period. However, in our full sample we find that approximately 11 percent of students change schools from fourth to fifth grade, 27 percent of student change schools from fifth to sixth grade, and 21 percent of students change schools from sixth to seventh grade. In addition, approximately 40 percent of students move schools only one time, eight percent move two times, and one percent moves three times.

[3] It is also common to use the difference between g and g-1 test scores as the dependent variable, which is the gain score model (Rothstein, 2010). We find the gain score model to be too restrictive, as it assumes there is no decay in the effect of past inputs. While the lagged score model is also restrictive, we must trade off restrictions against model tractability, and we believe that this is the least restrictive model in the literature that we can feasibly estimate with our data.

[4] Note in equation 4 that the composite error term is correlated with lagged test scores through ε_i^4. This is common to value-added models of this type. While we do not directly account for this issue, we do present models without lagged test scores in the model and

show that results are similar, suggesting its correlation with the error does not cause substantial bias.

[5] Essentially we treat schools with different sequences of grade 5 and grade 6 principals over time as different schools. Relabeling the interaction term in this way makes the later descriptions much simpler.

[6] In the estimation, within each group the sum of the principal dummies equals the sum of the principal 5/6 by school dummies. Thus, one principal dummy in each group is dropped. Alternatively, one could drop one principal 5/6 by school dummy from each group. Because this model does not include a constant, we keep all principal 5/6 by school dummies in the regression.

[7] While shrinking our estimates is important in this estimation, shrinking of estimates will generally be much more important in the case of teachers as they have more variation in class size and in the number of students tested each year which will generate more variation in the precision of teacher fixed effects versus principal fixed effects.

[8] We provide estimates based on clustered errors to account for possible correlation of student errors within both principals and schools. For the clustered standard errors, we follow the multi-way cluster method of Cameron, Gelbach, and Miller (2011).

[9] In particular, we examine methods used by Rothstein (2010), which is similar to the method we use, but it weights the average of the standard errors and by Rockoff (2004), which makes assumptions about the distribution of the underlying true principal effects and estimates the variance using maximum likelihood.

[10] We present summary statistics for the full sample of students, including the switchers, in Appendix Table 3. Removing the switchers leaves us with a sample that is largely

similar to the full sample, though children in the subsample have higher test scores, and fewer Aboriginal, ESL, Special Education, and rural students. Thus, our main estimates are therefore based on a set of slightly less “at-risk” students. A priori, adding these students back into the analysis would likely add more variation to principal effects, which we show in a robustness check in Table 7, Panel B.

[11] Rothstein (2010) tests the identifying assumptions for teacher effects by checking whether current teacher assignment is related to past test score gains, conditional on past teacher assignment. In theory, one could repeat that exercise with principals. Due to data limitations (having only 2 time observations for each student), we cannot replicate that test in our paper because we cannot construct a lagged gain score.

[12] Appendix Table 1 uses an indicator that equals 1 for any move, including switches to schools out of the sample or exits from the sample for the dependent variable.

[13] The unit of measurement of each test score variable is student-level standard deviations. A one standard deviation increase in student test scores is roughly equivalent to a 2 standard deviation increase in test scores at the school level. Thus, these results should be interpreted as roughly half of their reported size. The estimates for the coefficients for the rest of the model are available upon request. Overall there is little relationship found between the other covariates with the exception of principal experience. Principal experience is significantly positively related to principal mobility.

[14] A statistically significant effect for the two year gain and loss in level scores is found in the sample of any principal switches along with a statistically significant effect for large losses for two year in gain scores. These coefficients can be found in Appendix Table 1.

[15] Standard errors for each standard deviation are computed via the delta method. The standard errors computed via a simple nonparametric bootstrap, with 99 bootstrap replications are available from the authors upon request and are generally somewhat larger than the errors calculated via the delta method.

[16] These variables include indicators for whether the student is male, aboriginal, or in special education, or has repeated a grade. It also includes the number of students enrolled in the school and the number of teachers, along with the number of students excused from the math test and reading test. In addition, it includes the percentage of students at each school who are males, who are in special education, and who are in ESL. Finally, the following DA level community characteristics are also included: average household income, percentage of individuals with a high school diploma, percentage of individuals with a university degree, percentage of individuals 25 years or older who are unemployed, percentage of the population who are immigrants, percentage of the population who are visible minorities, average value of dwellings, and percentage of the population who are older than 65 years.

[17] Based on the Rothstein (2010) method for adjusting the standard deviation, we get .411 for math and .257 for reading. Using the Rockoff (2004) method, we get .402 for math and .276 for reading.

[18] To show this, we use a simplified example. Suppose we run an OLS regression of test scores on only principal effects, and obtain $y_i = \hat{\delta}_p + e_i$. Aggregating to the principal level, we get $\bar{y}_p = \hat{\delta}_p$ because the OLS residual sums to zero for each principal. Now suppose we include school effects in the regression and get $y_i = \hat{\delta}_p^* + \phi_p + e_i^*$.

Aggregating to the principal level, we get $\bar{y}_p = \hat{\delta}_p^* + \bar{\bar{\phi}}_p$ again because the residual sums to zero for each principal. Combining the two equations, we get $\hat{\delta}_p = \hat{\delta}_p^* + \bar{\bar{\phi}}_p$. The variance is $VAR(\hat{\delta}_p) = VAR(\hat{\delta}_p^*) + VAR(\bar{\bar{\phi}}_p) + 2COV(\hat{\delta}_p^*, \bar{\bar{\phi}}_p)$. We can use this to show that $VAR(\hat{\delta}_p^*) > VAR(\hat{\delta}_p)$ if $COV\left(\hat{\delta}_p^*, \bar{\bar{\phi}}_p\right)/VAR(\bar{\bar{\phi}}_p) < -1/2$, which will occur if the principal and school effects are negatively correlated.

[19] In Appendix Table 2, we estimate the mean and standard deviation of principal by school effects for schools having above and below the median level of test scores and household income. We use principal by school effects because principals work at multiple schools, and therefore the exercise would not be feasible with pure principal effects. In models without the school effect, the mean principal effect is higher in high income and high scoring schools. The dispersion is higher in low achieving and low income schools. With school effects, dispersion is higher in low income and low achieving schools for math, and vice versa for reading. The mean effect is mechanically zero when school effects are in the model, because when estimating principal by school effects, schools serve as the connected groups.

[20] Appendix Table 3 displays the descriptive statistics for the sample of students used in Panel B. Estimates of principal fixed effects without the inclusion of school fixed effects for this table are available from the authors upon request.

[21] In addition to the robustness checks described in this section, we also conducted a falsification exercise based on randomizing grade 7 principals across schools. The intent was to measure the amount of sampling error relative to the overall variation in principal

fixed effects, though the test did not provide very reliable results. We refer readers to the online appendix for a discussion of this test and the results.

[22] Due to computational limitations, the standard error in this panel is calculated using a simple nonparametric bootstrap, with 99 bootstrap replications.

[23] Because principals work at different schools over time, we could not demean pure principal effects by the within school average.

[24] Another possible sensitivity check would be to restrict the sample to include schools with two or more principals and where students had the same principal in all years and were in the same school for all three years. The estimated principal fixed effects in this specification that includes school fixed effects is 0.37 for math, and 0.25 for reading. The method produces principal effects that measure each principal's contribution to student test scores over 3 years. However, these estimates are not comparable to our main estimates due to using different variation and a highly selected sample.

[25] Their "adjusted" standard deviation does not have the same interpretation as ours. In their methodology, the adjusted standard deviation first nets out the effect of school and individual demographics and community characteristics.