INSTITUT C.D. HOWE INSTITUTE
Essential Policy Intelligence | Conseils indispensables sur les politiques

E-BRIEF APPENDIX

November 13, 2012

Methodology for "Are Middle Schools Good for Student Academic Achievement? Evidence from Ontario"

by David R. Johnson

This Appendix presents details of the author's methodology for measuring the association between students' pass rates on secondary school assessments and the paths they take from elementary school to secondary school.

The Data

I use three sources of data: student performance and characteristics from the Education Quality and Accountability Office (EQAO), student postal codes from the Ministry of Education and Census data from Statistics Canada.

Student Education Data

I link student results on the Grade 9 math assessments and school information from 2008/09 and 2009/10 for each individual student who wrote secondary school exams in those years with that student's matching mathematics assessment result and school information in Grade 6 in 2005/06 and 2006/07 respectively. Similarly, I link student results on the 2009/10 Grade 10 literacy test results (the OSSLT) and school information in 2009/10 for each individual student with that student's matching result and school information on the Grade 6 reading assessment in 2005/06. The OSSLT is an examination where the result is reported only on a pass/fail basis. The Grade 9 mathematics assessment is scored with 4 levels: a Level 4 or Level 3 is considered to meet or exceed provincial standards, in this paper, a pass. All other students are treated as failing.

I divide students into four paths by type of school attended in Grade 6. If a student writes his or her Grade 6 assessment in a school ending in Grade 8, I assume the student avoided either the standard or extended middle school path. In this study, I label all students who exit a school after Grade 6 as taking the standard middle school path.[1] Some of these students complete Grade 7 and 8 by joining another elementary school that operates to the end of Grade 8. Most students who exit a school after Grade 6 attend a standard middle school where

1 The data do not track students through specific schools on a year-by-year basis. I only know the type of school where the Grade 6 test was written. I do not know exactly which students attended a Grade 7 and 8 middle school and which students completed Grade 7 and 8 at another elementary school that included those grades as well as earlier grades. 28,620 students in 2005/06 wrote the Grade 6 EQAO assessment in a school that ended in Grade 6. In 2005/06, 23,150 Grade 7 students were enrolled in schools where only Grade 7 and 8 were taught. Thus, most students leaving a school as of Grade 6 are likely attending a school where only Grade 7 and 8 are taught. In some jurisdictions what I call extended middle school is called middle school and what I call standard middle school is called junior high school.

grades 7 and 8 are taught. A minority of these students move to complete Grade 7 and 8 in another elementary school. The data do not allow these two paths to be separated. If a student was in an extended middle school in Grade 6, I assume that student followed the extended middle school path to Grade 9. My analysis compares only students following these three paths and I drop all other students (about 4 percent of the total in each year) from the analysis.

Socio-Economic Data

There are four controls related to the social and economic background of students, two explicit and two implicit. For the first control variable, I link student postal codes – provided by the Ministry of Education – to data from the long form 2006 Census to create a complete picture of the social and economic background of the neighbourhoods of students who attend a school. A student drawn from a specific Grade 6 school is assigned that school's measure of parental education. This is clearly only an approximation for the individual educational background of the student. The second control variable is the student's gender.

The third control variable is the student's outcome on the Grade 6 assessment. This outcome reflects a combination of both the student's individual ability, the student's social and economic background at the individual level and the quality of teaching experienced by that student up to and including Grade 6.[2]

The fourth control is a dummy variable for the secondary school attended by the student. These school fixed effects capture the quality of education experienced by the group of students attending that secondary school as well as the average quality of schooling received during Grade 7 and 8, the period since the Grade 6 assessment. To the extent that social and economic effects might vary by year of assessment and secondary schools vary in the social and economic background of their students, this variable also acts as a further control for the social and economic background of a student.

The Model

The model I use to analyse the effect of middle school paths on Grade 9 mathematics assessments is:

$$\begin{aligned}P_{i,j} &= a + b_1\,Dum_{year2} + b_2\,(Dum_{year2} \times P_{i,Grade6})\\ &+ b_3\,Sex_i + b_4\,P_{i,Grade6}\\ &+ b_5\,Dum_{i,678} + b_6\,(Dum_{i,678} \times Educ_{i,s6}) + b_7\,(Dum_{i,678} \times P_{i,Grade6})\\ &+ b_8\,Dum_{i,78} + b_9\,(Dum_{i,78} \times Educ_{i,s6}) + b_{10}\,(Dum_{i,78} \times P_{i,Grade6})\\ &+ s_j\,Dum_{i,j} + u_{i,j} \qquad (1)\end{aligned}$$

In equation (1) $P_{i,j}$ is an indicator (dummy) variable that takes on a value 1 when student "i" at secondary school "j" passes the secondary school assessment and 0 if the student fails. The remainder of equation (1) is a series of variables which help predict if a student will pass or fail the secondary school assessment in question.

2 I cannot use the IV approach in Rockoff and Lockwood (2010) to control for student background because I do not know the configuration of the Grade 3 school attended by the student. This is the first use of EQAO data that links Grade 6 and 9 results. This is the first Grade 6 cohort with a student number linked to later exams. There are no links to earlier exams.

The coefficient on an indicator variable Dum_{year2} measures, for the Grade 9 mathematics assessments, whether the assessment in the second year of data had a different pass rate than the assessment in the first year. The reason for this is that even though these exams cover the same topics, they posed different questions to the students in each of the years, meaning the exam could have been slightly easier or harder in one year versus the other. This variable and the interaction terms associated with it are not included in the OSSLT model because there is only one year of OSSLT results. If b_2 is positive, then a student in the second year of data is more likely to pass simply because of the year in which they wrote the examination. There is an interaction variable between the second year indicator variable and the indicator variable for passing the Grade 6 assessment. The students who write the Grade 9 mathematics assessment in the second year also write a different mathematics examination in Grade 6. If the Grade 9 examination was very easy and the Grade 6 exam of 'normal' difficulty, then passing the Grade 6 examination is not as useful a signal that a student will also pass the Grade 9 examination. The coefficient b_2 would be negative in this example. It could also be a positive coefficient if, for example, the Grade 6 assessment that matched the second year mathematics assessment had been very hard so that passing the Grade 6 assessment in a "hard" year was a stronger predictor of secondary school success than passing the Grade 6 assessment in the "easy" year.

Sex is an indicator variable where the value 1 indicates a male student. If b_3 is positive, a male student is more likely to pass the assessment in question. There is an indicator variable set equal to 1 if the student passed the Grade 6 assessment. I expect the coefficient b_4 to be large and positive, that is, a student who passed the Grade 6 assessment would be much more likely to pass the Grade 9 or Grade 10 assessment measuring similar skills. As already stated above, the ability of a student to pass the Grade 6 assessment will be a combination of student ability, student social and economic background and the quality of teaching received by the student to the end of Grade 6.

$Dum_{i,\,678}$ and $Dum_{i,78}$ are indicator variables where the indicator takes the value of 1 when student "i" takes either the extended or standard middle school path – these are labelled with the relevant combination of grades. These are the variables of direct interest in measuring the association of middle school path and secondary school outcomes. The direct coefficients on these variables, b_5 and b_8, if they appeared alone in the equation, would measure the change in the probability of passing the secondary assessment associated with following a middle school path. However they do not appear alone in the equation. They are interacted with two other variables, the average education level at the Grade 6 school of this student and whether the student passed the Grade 6 assessment.

The Interaction Between Student Characteristics and Middle School Path

I include the interaction between the variable $Educ_{i,s6}$, the logarithm of the percentage of adults with some university education living in the student's Grade 6 school community, and the two dummies for the middle school paths. The coefficient on these interaction terms are b_6 (with the mathematics results) and b_8 (with the OSSLT result) respectively. A high value of $Educ_{i,s6}$ means the student comes from a relatively well-educated school community. The direct effect of the extended middle school path is measured by b_5 and the standard middle school by b_8. The interacted effect of the extended middle school path is measured by b_6. The total effect of the extended middle school path depends on both coefficients and the student's position on the distribution of the education variable. If b_5 is negative and b_6 is positive, the extended middle school path has less effect on the probability of passing the secondary school assessment as the Grade 6 school's education level rises. The coefficients b_8 and b_9 are related in the same way to measure the effect of the standard middle school path at different levels of education.

Students who switch schools between Grade 5 and Grade 6 perform substantially worse on the Grade 6 assessments. The extended middle school is, on average, an adverse situation for a Grade 6 student. This means that passing the Grade 6 assessment in an extended middle school (represented in the equation by an interaction variable, where the coefficient of interest is b_7) should signal higher individual ability than passing the same assessment in another setting. I apply the same interaction effect of the dummy variable for the standard middle school path with Grade 6 performance (where the coefficient of interest is b_9) although I have no *a priori* reason to expect this coefficient to be significant.

Secondary School Fixed Effects

The last term in equation (1) is a series of secondary school indicator variables and the associated coefficients s_j. These are usually called school fixed effects. Secondary school fixed effects play several roles in the analysis.[3] To be more concrete and understand better, think of a large positive value of s_j so that students attending secondary school j have a higher probability, all else equal, of passing a secondary school assessment. This large positive value could indicate better than average quality of teaching at the secondary school j. A high value of s_j could indicate higher than average quality of teaching received by the student in secondary school j at the schools that those students attended in Grade 7 and 8, whether a middle school or a continuation of the student's Grade 6 school. Finally, there is evidence (see Johnson (2005) and Johnson and Brydon (2012)) that the association of assessment results and social and economic characteristics does vary by grade level of assessment. The secondary school fixed effects would control for variation in the social and economic background of students not captured in the education variable from the Grade 6 school and not captured by the Grade 6 assessment. Johnson (2010b) shows that in Alberta the relationship between social and economic variables does vary by level of assessment.

The most useful way to summarize the model in equation (1) is to recognize that it compares two students of the same gender at the same secondary school with the same Grade 6 test result at the same value of the education measure and asks if, on average, the probability of passing the Grade 9 mathematics assessment or the OSSLT depends on the path through grades 6, 7 and 8. Do students who take the middle school routes have lower pass rates?

Results

The estimated coefficients from equation (1) are presented in Table A-1. I use a linear probability model, meaning all coefficients reported in the table are changes in probability from a one unit change. The first two rows pertain only to the Grade 9 mathematics assessments. In 2009/10, the second year of mathematics results, students had a 6.8 percent higher likelihood of passing the mathematics test than the year before. When the Grade 9 mathematics test became easier to pass in the second year, passing the associated Grade 6 mathematics assessment is a weaker signal that the student will pass the test, making the estimate of b_2 negative and quite large.

Boys are more likely to pass the mathematics assessment than girls. Girls are more likely to pass the OSSLT. Thus, for mathematics the coefficient b_3 is positive and in the OSSLT equation b_3 is negative. Both effects are large and statistically significant.

3 I have to choose between which set of fixed effects to include. If I had board-or-city fixed effects, then I would have to drop the secondary school fixed effects since all secondary schools belong to a board or are in a city. The association of middle school paths and secondary school results is stronger (more negative) when the secondary school fixed effects are excluded from the model.

The coefficient on the indicator variable that the student passed the matching assessment in Grade 6, b_5, is extremely large, 0.424 for mathematics and 0.317 for the OSSLT. This variable incorporates the three effects that complicate its interpretation: the quality of schools attended by the student to the point of the Grade 6 assessment; the student's individual-specific social and economic background; and individual-specific academic ability of this student that is not associated with either school quality or social and economic factors. Its presence in the equation controls for all three factors and helps isolate the effects of the middle school path.

The secondary school fixed effects that control for characteristics of the secondary school also help isolate the effects of the middle school path. Johnson (2005) and Johnson and Brydon (2012) show that school assessment results vary by the social and economic background of students attending a school as well as by the quality of teaching at a school. The coefficients on secondary school fixed effects are unreported in Table A-1. They are a combination of student background and measures of the quality of all schools attended by the student up to and including the secondary school attended at the time of the assessment. In my analysis of mathematics results, I included students from a school if there were test results from 50 students in both years from that school. Likewise, in my analysis of OSSLT results, I included students from a school if there were test results from 50 students in Grade 10 in that school. Thus, I estimate the school fixed effects using a large number of students where students from all three paths attend the same secondary school.

The presence of these control variables in equation (1) means that we are comparing the pass rates across middle school paths using students who are attending the same secondary school but arrived at that school through the two different middle school paths as well as, for most students, no middle school at all. Further, the students we are comparing are of the same gender and have the same result on the relevant Grade 6 assessment.

Summary of Path Effects

The path effects through middle school are estimated using the coefficients on the dummy variables as well as their interactions. There are two sets of effects estimated, one for extended middle school and the other for standard middle school, representing coefficients b_5 through b_{10}. The average effect of the extended middle school path on the Grade 9 mathematics pass rates is -0.015. This means that students on that path are 1.5 percentage points less likely to pass the Grade 9 assessment than other students at the same secondary school who share the same gender and success on the Grade 6 assessment after controlling for where (and when) the assessments were written. The averages are calculated over all the values of the education variable and whether the student passed or failed the Grade 6 assessment. Those students on a standard middle school path have a 1.7 percentage point higher failure rate in Grade 9 mathematics. Both middle school paths reduce Grade 9 mathematics success for the average student. Those students who pass through a standard middle school have a 0.9 percentage points higher failure rate on the OSSLT than the average student. Passing through an extended middle school does not change the probability of passing the OSSLT in this model.[4]

4 If I estimate the model without secondary school fixed effects, then both middle school paths reduce OSSLT pass rates and the reduction in Grade 9 mathematics results associated with the middle school path are larger. The interpretation of this coefficient pattern is that some secondary schools receive a lot more middle school students than others. Part of the school fixed effect picks up differences between secondary schools that receive a lot of middle school students and secondary schools with very few middle school students. In this sense the modelling strategy with secondary school fixed effects is a lower bound estimate of the effects of middle school paths on pass rates. The effects could be larger.

The interaction of the path variable and the education variable provides measures of the effect of each middle school path on passing the secondary school assessment at different points of the socio-economic distribution of students. The education variable is measured by the logarithm of the percent of adults with a university degree associated with the student's school in Grade 6. In both the extended and standard middle school paths, it is clear that the middle school transition is associated with a larger drop in mathematics pass rates for students from Grade 6 school communities with less educated adults. The table is read as follows: at the 10^{th} percentile of parental education, the drop in pass rates is 5.4 percentage points following the extended middle school path and 6.2 percentage points for a student following the standard middle school path. As the level of education rises, the positive value of the coefficient b_6 multiplied by the positive value of the education variable offsets the large negative value of b_5. The negative effects of middle school on mathematics pass rates are found for all students at or below the median level of parental education.

There are small gains in results for students who come from neighbourhoods in the 90^{th} percentile of the socio-economic indicator (2.3 and 2.9 percentage points). The pattern of the interactions between the middle school path and the education measure for the OSSLT results and the standard middle school path are similar to the mathematics results – students from more disadvantaged backgrounds are further disadvantaged by attending a middle school. There is no measureable path effect of the extended middle school path to secondary school on the pass rate on the OSSLT at any point on the social and economic distribution although the pattern of coefficients does move from negative to positive.

The model allows for a similar interaction between the path indicator and the Grade 6 pass indicator. This might be needed because we know that the extended middle school path is associated with lower Grade 6 pass rates. Thus we might expect that the association of a Grade 6 pass and passing the secondary school assessment would vary by middle school path. In particular, it ought to be the case that a pass in Grade 6 in a more difficult circumstance, a new school, should have a marginal positive association with passing the secondary assessment (b_7 is predicted to be positive). However neither the coefficient estimates of b_7 and b_{10} are statistically significant. These interactions are not important.

Conclusion

The overall conclusion is clear. Both middle school paths reduce the probability of passing the Grade 9 mathematics assessment. The reduction is larger when students come from a more disadvantaged background. The standard middle school path reduces OSSLT results in the same way. The extended middle school path does not affect pass rates on the OSSLT.

Table A-1: Estimating the Association of Middle School Paths to Secondary School and EQAO Secondary School Pass Rates				
Regression Dependent Variable:	**Grade 9 Mathematics Pass Rate**		**OSSLT Pass Rate**	
Effects associated with the second year of assessments in mathematics				
Dummy for second year of assessments	0.068 (0.004)**		NA	
Dummy variable if passed Grade 6 interacted with second year dummy	-0.029 (0.004)**		NA	
Individual characteristics of student				
Male	0.032 (0.001)***		-0.025 (0.002)**	
Dummy variable if passed Grade 6	0.424 (0.004)**		0.317 (0.004)**	
Extended middle school path				
Extended middle school dummy	-0.200 (0.040)**		-0.042 (0.040)	
Extended middle school dummy interacted with adult education measure	0.054 (0.013)**		0.009 (0010)	
Extended middle school dummy interacted with passed Grade 6 dummy	0.001 (0.008)		0.013 (0.010)	
Standard middle school path				
Standard middle school dummy	-0.224 (0.020)**		-0.104 (0.010)**	
Standard middle school dummy interacted with education measure	0.063 (0.006)**		0.032 (0.006)**	
Standard middle school dummy interacted with passed Grade 6 dummy	0.001 (0.006)		-0.013 (0.008)	
Summary of path effects				
Average marginal effect of attendance in extended middle school	-0.015 (0.006)**		-0.001 (0.006)	
Average marginal effect of attendance in standard middle school	-0.017 (0.003)**		-0.009 (0.003)*	
Effect of middle school path at different percentiles of education measure (percent of adults with some university at Grade 6 school)	Standard	Extended	Standard	Extended
10th (9 percent)	-0.054 (0.010)**	-0.062 (0.006)**	-0.008 (0.01)	-0.032 (0.007)**
25th (15 percent)	-0.036 (0.009)**	-0.041 (0.005)**	-0.005 (0.009)	-0.021 (0.005)**
50th (25 percent)	-0.013 (0.005)**	-0.014 (0.0063)**	-0.001 (0.005)	-0.007 (0.003)*
75th (44percent)	0.005 (0.005)	0.007 (0.003)*	0.002 (0.005)	0.003 (0.004)
90th (75 percent)	0.024 (0.008)**	0.029 (0.004)**	0.005 (0.007)	0.014 (0.005)**
Regression characteristics				
Number of students	245,107		129,165	
Number of secondary schools	684		696	
R^2	0.24		0.21	

* This coefficient is statistically significant at 5%. ** This coefficient is statistically significant at 1%.
Dependent variable is the indicator for passing the secondary school assessment. All regressions have school fixed effects at the secondary school level. There are 684 secondary schools with more than 50 students in both years in Grade 9 and 696 secondary schools with more than 50 students in Grade 10. Standard errors are corrected for clustering at the secondary school level.

Sources: Author's calculations from EQAO, Ministry of Education and Statistics Canada data.