PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
MAGISTER EN ECONOMIA

TESIS DE GRADO
MAGISTER EN ECONOMIA

Poblete, Coddou, Sebastián Jesús

Julio, 2018

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
MAGISTER EN ECONOMIA

School Value-added and the Math Gender Gap

Sebastián Jesús Poblete Coddou

Comisión

Constanza Fosco
Tomás Rau
Gert Wagner

Santiago, Julio de 2018

School Value-added and the Math Gender Gap

SEBASTIÁN POBLETE*

M.A. in Economics Thesis, Pontificia Universidad Católica de Chile

This Draft: July, 2018

Abstract

We study gender differences in levels of school value-added and their role in explaining the math gender gap in Chile. Using rich and representative panel data on students' test scores, we fit two-way (student and school) fixed effect models for each gender. We interpret school fixed effects as a value-added measure. We find that schools on average do not play a role in the determination of the math gender gap. However, there is heterogeneity in this effect with private schools helping to close this gap by 38.6%. In studying the mechanisms, we rule out the possibility that this average result for private schools is driven by girls being over-represented in high value-added schools. The effect found is explained by a gender-specific effect: at a given private school, more value-added is obtained by girls relative to boys. Finally, we find that the presence of a higher share of female teachers and attending a private school correlates with higher returns to girls' value-added, and that average teachers' expectations regarding students' future outcomes play a more favorable role for boys.

*Thesis written as a Master student at PUC–Chile, Department of Economics. I thank Tomás Rau, for thoughtful guidance and unconditional support. I thank committee members Gert Wagner and specially Constanza Fosco for their advice. I am grateful to Gastón Illanes, Rodrigo Soares and Seth Zimmerman for discussion and useful suggestions. I am grateful to José Diego Salas, Vicente Jimenez, León Guzmán, Fernando Ochoa, Francisca Pinto, Felipe Rodriguez, Nicolás Figueroa, Tomás Croxatto, Patricio Rodríguez, Martín Canessa and my family and friends for comments and support. I thank the Agencia de la Calidad de la Educación from Chile the access to data crucial to this study. I thank the funding provided by Fondecyt, Project No. 1171128. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). All mistakes are solely my responsibility. sjpoblete@uc.cl

1 Introduction

A substantial gender gap found in the results of mathematical tests has been thoroughly documented in the literature. This difference is present in developed (Mead, 2006), low, and middle income (Bharadwaj et al., 2016) countries. The magnitude of this difference is estimated to be valued at roughly 0.2 standard deviations, it is present across every stratum of society, and starts to appear in early years of schooling and then broadens with age (Fryer and Levitt, 2010). The importance of this math gender gap lies in the fact that math test scores are consistently found to be a strong predictor of wage levels (Paglin and Rufolo, 1990; Altonji and Blank, 1999; Murnane et al., 2000, 1995; Weinberger, 1999, 2001) and career choices (see Altonji et al. (2012) for a review). These results imply that the math gender gap may play a significant role in the determination of the well-documented wage gender gap in the labor market.

In this paper, we examine whether schools secure different value-added levels for students depending on gender. We achieve this by estimating two-way fixed effects models with student and school effects using the framework developed by Abowd, Kramarz and Margolis (1999) (hereafter, AKM) and using student-level Chilean data on test scores and school characteristics. Following Card et al. (2016), we obtain for each school one fixed effect for girls and another for boys. We interpret school effects as value-added measures as in similar frameworks (Angrist et al., 2017; Bharadwaj et al., 2016; Abdulkadiroglu et al., 2017). There are two main advantages of using an AKM-style model. First, it allows for one to estimate school fixed effects while controlling for student fixed effects. This limits the emergence of selection biases resulting from time-invariant student characteristics. Second, we can test whether matching effects at school or test score shocks correlate with school mobility, and therefore also infuse selection bias into our estimates. This is done by adapting exercises developed by Card et al. (2013) and Card et al. (2016) to the educational context. We find no evidence of patterns showing that student mobility affects our estimates. Controlling for student characteristics and testing for mobility shocks and matching effects is important because the mentioned forms of selection bias often affect school value-added estimates (Angrist et al., 2017).

The Chilean case is useful for understanding the math gender gap for two main reasons. First, Chile shares gender differences present in other parts of the world in the school and labor markets. Women in Chile are severely under-represented in STEM[1] fields. They also exhibit the lowest levels of participation in the labor force and lower mean and median salaries when compared to the rest of the region and other OECD countries. (OECD, 2017). The gap found in math scores has been estimated at roughly 0.17-0.19 across several cohorts (MINEDUC, 2013; Paredes, 2014; Bharadwaj et al., 2016). As in other countries, it appears in the early stages of schooling around 4th grade and widens with age (Bharadwaj et al., 2016). These facts render the Chilean case informative regarding the math gender gap.

The second key feature of the Chilean case concerns rich and representative individual test score data developed by the Ministry of Education of Chile (MINEDUC, 2013). The advantage of using this dataset is threefold. First, as students are evaluated across several periods, we are able to build a panel and therefore use the AKM model to simultaneously estimate school and student fixed effects, which is crucial for our purposes. Second, student mobility levels in Chile are high (Larroulet, 2011), and in our particular dataset, 73% of students have changed between institutions within the sample. School fixed effects are only identified for the schools that belong to the largest connected set, which is the subset of schools connected by student mobility (i.e., have received or "sent" students to/from schools within the sample). Given the high student mobility in our dataset, we can compute value-added measures for almost all institutions included in the sample. Finally, the dataset offers rich information on schools and teachers that allow for us to test for heterogeneity in value-added differential impacts of the gender gap and to explore mechanisms that may explain a differential value-added secured by students depending on gender.

Our first set of results shows that schools on average do not have a significant impact on the math gender gap. We compute the average contribution of schools to the gender gap measured from the difference between school value-added results for boys and girls. The result of this exercise show that the mean value-added differential is valued at almost zero, confirming that, on average, schools are not helping to close or enlarge the

[1]STEM: Science, Technology, Engineering and Mathematics fields.

gender gap. This result is consistent with previous evidence showing that the gender gap observed within schools is not different from the overall average gender gap Bharadwaj et al. (2016); Carrell et al. (2010). However, we do find substantial heterogeneity in this effect. While effects observed are almost negligible for public and voucher schools, we find that in private schools, girls obtain on average 0.07 standard deviations more value-added than boys. This means that with respect to the overall gap observed in the sample, private schools are closing the gap by 38,6%.

To investigate the potential mechanisms shaping these results, we apply two exercises. First, motivated by evidence on the effect that gender sorting at school could have on gender sorting for college enrollment in the US (Long and Conger, 2013a,b), we replicate Card et al. (2016) Oaxaca-style decomposition to separate the mean value-added differential into "sorting" and "gender-specific" channels. The sorting channel corresponds to average gender disparities that arise when female students are less likely to be enrolled at high value-added schools. The gender-specific channel arises when female students obtain, in expectation, more value-added than males, in a given school. For the sorting channel, we find negligible effects for the overall school sample and when restricting our focus to voucher, public and private schools. This means that gender sorting across schools does not contribute to the average gender gap observed in test scores. While small for public and voucher schools, we find that for the private sector, the gender-specific channel explains almost all of the value-added differential (favorable to girls) found across these schools. This result could help explain existing evidence showing that high-performing girls in the US (in math) come from a limited number of elite schools (Ellison and Swanson, 2010). Elite schools could be helping high-performing girls develop their skills by delivering more value-added that compensates for other differences. This explanation opposes to their average performance being explained only by an over-representation of girls in elite schools.

A second exercise for finding mechanisms through which value-added impacts the gender gap involves regressing value-added measures for each gender over a set of school characteristics. We find that the average share of female math teachers correlates with stronger value-added outcomes for girls. This result is consistent with previous results showing that gender-matching in the classroom favors girls' performance in math (Dee, 2007; Carrell et al., 2010; Paredes, 2014; Bharadwaj et al., 2016). We also find that the average expectations of math teachers regarding boys' future educational outcomes tend to favor boys. These results are similar to those found in the literature on the role of teachers' expectations. In particular, other gaps between racial/demographic groups (Ferguson, 2003; Figlio, 2005; Gershenson et al., 2016a) relate to the self-fulfilled views of teachers regarding their students. Finally, even when controlling for teacher and school characteristics, private schools deliver more value-added to female students than to male students.

Our paper makes three main contributions. First, we use panel data to address the school contributions to the gender gap. This allows for us to control for time-invariant student characteristics as opposed to many articles that study the gender gap that use cross-sectional data and that control only for observable characteristics. To the best of our knowledge, this is the first paper that uses an AKM-style model to calculate school value-added while controlling for a student fixed effect. Second, we test for heterogeneity in the impact of gender-specific school value-added on the gap. We find that different schools (private schools) effectively deliver more value-added to female students. Finally, we use a comprehensive set of variables and different exercises to find the mechanisms that shape our results. These exercises show that private schools systematically deliver more value-added to girls than to boys, as the "gender-specific" effect is substantial across these schools. This means that our result showing that private schools help close the gap on average is not driven by an over-representation of females in schools with higher value-added characteristics. Additionally, we study the effects of several school characteristics on the gender gap (in value-added) other than those present in the literature such as teachers' expectations and the religious orientations of schools.

The rest of the paper proceeds as follows. Section 2 describes the econometric model used to estimate gender-specific value-added measures and empirical exercises used to assess the contributions of schools to the math gender gap. Section 3 describes the data used. Section 4 presents our primary results and an analysis of the mechanisms that shape them. Section 5 concludes.

2 Methods

To estimate the gender-specific value-added of each school, we fit the following additively separable and gender-specific education production function:

$$
\begin{aligned}
y_{it} &= \alpha_i + \psi^{G(i)}_{\boldsymbol{J}(i,t)} + X'_{it}\beta^{G(i)} + \eta^{G(i)}_{i\boldsymbol{J}(i,t)} + \xi^{G(i)}_{\boldsymbol{J}(i,t)t} + \epsilon_{it} \\
&= \alpha_i + \psi^{G(i)}_{\boldsymbol{J}(i,t)} + X'_{it}\beta^{G(i)} + r_{it}
\end{aligned}
\tag{1}
$$

where y_{it} is the test score obtained by student i in year t. $G(i) \in \{M, F\}$ denotes the gender of student i. We also denote $j = \boldsymbol{J}(i,t)$ as the school at which student i is enrolled in period t. The term α_i denotes time-invariant characteristics of student i that are fully portable between schools; elements such as a student's family background, socio-economic status, and natural abilities are measured by this term. X_{it} denotes time-variant covariates that are portable between schools for student i. $\psi^{G(i)}_j$ are gender-specific school fixed effects. As in Card et al. (2016) study of the wage gender gap, which is based on the AKM method, we estimate gender-specific equations to obtain effects for girls and boys for each school.

The error r_{it} is composed of three terms. Term η_{ij} represents a time-invariant score "premium" for student i at school j relative to the score obtained by $\alpha_i + \psi_j$. This component can arise from matching-related score gains (i.e., a school employs some practices that improve the test scores of a particular kind of student). Term ξ_{jt} denotes time-varying factors that raise or lower the average school impact on scores for all students. Finally, ϵ_{it} accounts for other unobservable factors.

The critical concern in this paper is to correctly identify school-specific effects $\psi^{G(i)}_j$. This term captures the gain or loss in scores that is common for all students of gender $G(i)$ enrolled at school j. We will also explore if expectations of gains/losses for students vary across different types of schools. As in similar models, we interpret school effects as value-added measures (Angrist et al., 2017). Conventional value-added models can be biased due to their selection of student characteristics and other factors (Angrist et al., 2017). We argue below that our estimates are not biased for two reasons. First, our data and framework allow for us to control for time-invariant student characteristics α_i to limit selection biases relating to student characteristics. On the other hand, we show that temporary shocks affect test scores and school matching effects by applying the same tools of the AKM framework used in the labor literature. This ensures that selection bias related to students sorting themselves into certain schools (due to matching premiums at schools or other temporary mobility shocks) does not affect our estimates.

Note that we need school mobility to identify a unique solution to the OLS estimation of (1). This is true because when there were no mobility in a sample, school identifiers are perfectly collinear with student identifiers. This in turn means that we can only estimate (ψ_j^F, ψ_j^M) for schools that have "received" or "sent" students of the sample, on the observed period. This subset of schools linked by student mobility is called the "largest connected set" of schools. As we indicate below in our data section, the largest connected set represents almost the entire sample of schools and students tested in 2007, 2011 and 2013, so the restriction to the largest connected set is unlikely to affect our results.

2.1 Identification and exogenous mobility

To identify parameters of our model included in Equation (1), the following orthogonality condition must hold:

$$\mathbb{E}[r_{it}|\alpha_i, \psi_j^{G(i)}, X_{it}] = 0$$

This condition is called the "exogenous mobility" assumption. Intuitively, the condition rules out any correlation between school identifiers, student identifiers and time-variant characteristics being correlated with rit. We will assume that school identifiers and characteristics of x_{it} are independent of r_{it}. Therefore, to identify school value-added measures $\psi_j^{G(i)}$, we need school identifiers to be independent of $(\eta_{ij}, \xi_{jt}, \epsilon_{it})$. This means that matching effects η_{ij}, temporary school shocks ξ_{jt} and other shocks to test scores ϵ_{it} do not predict school mobility.

Card et al. (2013) and Card et al. (2016) develop two ways to test for sorting based on school matching effects η_{ij}, which we apply to the context of this study. First, we examine the test score trajectories of students who change schools. If families are choosing schools based on a matching premium in test scores, we should observe to find that gains in scores obtained by students who move from one school to another should be higher in magnitude than losses observed for students who move in the opposite direction. Within a given limit, these matching effects should offset any losses in scores resulting from differences in school value-added measures and should lead to score gains for all students who change schools, as families change schools in pursuit of "matching gains." By contrast, under exogenous mobility, score gains resulting from moving to one school to another should be equal in magnitude to score losses associated with moving in the opposite direction. We find this "symmetrical" pattern in the data.

A second way to test whether match effects correlate with school mobility involves examining the fit of a fully saturated model that includes dummies for each school-effect match. If match effects are important in predicting test scores, then we should expect this model to explain more or the variance in scores than the additively separable model described in (1). Specifically, the root mean square error (RMSE) of our model should be higher than that of the match-effects model, and the adjusted R^2 of the model should be smaller. We do not find such improvements in fit by using the match-effects model.

Another threat to validity concerns the presence of a potential connection between a school-wide shock ξ_{jt} and mobility. For example, when a school experiences a massive exit of valuable teachers or starts to apply policies that affect test scores negatively, parents could choose to change schools. If this is true, we should find students who change schools to experience a drop in test scores just before moving and also unusual gains in test scores for these students, thus breaking the symmetry in gains and losses of movers. We find no such patterns in the data.

The final mobility pattern that may bias our estimations derives from the possibility that mobility could be correlated with a transitory test score shock ϵ_{it}. Parents or schools could make decisions based on the actual performance of students. If grades are correlated with test scores, it could be that students who are performing poorly should move to the "worst" (lower test scores) schools, while students who are performing well should transfer to better schools. Therefore, we should observe a drop in scores before a move and an increase in scores for students who move to "better" schools. We do not find systematic evidence for these trends except for a particular group of students who move to similar schools regarding test scores. However, if individual shocks or matching effects shape the mobility of such students, then we should find that students who stay at the same school in the observed period follow a different test score trajectory than school movers. We find that "stayers" exhibit the same pattern of test scores as students who move to similar schools.

If elements of r_{it} do not shape school mobility, then what drives it? Note that under the exogenous mobility assumption, the expected test score of a student conditional on him or her being enrolled in a particular school and on his or her characteristics is $\alpha_i + \psi_j + x'_{it}\beta$. Parents may be selecting schools merely from average test scores obtained at schools or depending on different costs of schooling or concerns related to socio-economic status (Gallego and Hernando, 2009). This type of sorting/selection does not affect our estimation of school effects, as we control for household (and other time-invariant) characteristics of the time-invariant student component α_i.

2.2 Decomposing inequality in test scores by gender

To analyze the distribution of the estimated school effects, we divide variance in test scores to separate and quantify the share of variance in test scores explained by student and school effects and we return to time-variant characteristics. We run this analysis for girls and boys to determine if such contributions depend on gender. This decomposition is possible because of a variance of test scores modeled by (1), and under the exogenous mobility assumption, it is:

$$\begin{aligned} Var(y_{it}) \quad = \quad & Var(\alpha_i) + Var(\psi_j^{G(i)}) + Var(x'_{it}\beta^{G(i)}) + \\ & 2[Cov(\alpha_i, \psi_j^{G(i)}) + Cov(\psi_j^{G(i)}, x'_{it}\beta^{G(i)}) + Cov(\alpha_i, x'_{it}\beta^{G(i)})] + Var(r_{it}) \end{aligned} \tag{2}$$

2.3 The impact of value-added differentials

We now turn our attention to our main concern to study contributions of schools to the gender gap. We accomplish this by computing the impact that differences in $\left(\psi^F_{J(i,t)}, \psi^M_{J(i,t)}\right)$ for each school on the gender gap. To illustrate the exercise, let *male* be shorthand for $G(i) = M$, and let $female$ be shorthand for $G(i) = F$. Using this notation and the model given in (1), we have the following expression:

$$\begin{aligned}\mathbb{E}[y_{it}|male] - \mathbb{E}[y_{it}|female] \quad = \quad & (\mathbb{E}[\alpha_i|male] - \mathbb{E}[\alpha_i|female]) + (\mathbb{E}[\psi_j^M|male] - \mathbb{E}[\psi_j^F|female]) \\ & + (\mathbb{E}[x_{it}'\beta^M|male] - \mathbb{E}[x_{it}'\beta^F|female])\end{aligned} \tag{3}$$

Note that the gender gap can be explained by the contributions of between-gender differences in students' characteristics, by a value-added differential (the "school component" of the gender gap), and by gender differences in the return to time-variant characteristics. Such decomposition allows for us to quantify the contributions of between-gender differences observed in schools on the gender gap.

As we obtain the complete distribution of $\left(\psi^F_{J(i,t)}, \psi^M_{J(i,t)}\right)$, we can use the decomposition given in (3) to check for heterogeneities in the contributions of $\mathbb{E}[\psi_j|male] - \mathbb{E}[\psi_j|female]$ to the gender gap. This is achieved by further conditioning (1) based on time-invariant school characteristics. We do so to determine if the average impact of the value-added differential changes when a school is public, voucher or private in format. In addition, obtaining the full distribution allows for us to apply several exercises to address mechanisms shaping the results as we show below.

2.4 Decomposing the contribution of the between-gender differential to the gender gap

Through a simple framework, we can divide the average value-added differential depicted in Equation (3) into two effects: an average sorting effect and an average "gender-specific effect." The sorting effect is interpreted as the share of the differential explained by an under/over representation of female students in schools with stronger value-added characteristics[2]. This effect may be relevant when sorting at the school choice level causes girls to end up in the worst (or better) schools on average. This may also serve as a channel through which gender sorting affects gender sorting in college, as some existing evidence shows (Long and Conger, 2013b,a). The "learning effect" is the residual or the average

[2]This means, for example, that when we observe that girls obtain more value-added than boys in some schools, this could be explained by girls being over-represented in schools with strong value-added features relative to males.

difference in school effects observed across the distribution of males or females. This effect is interesting because it reveals an average "inclination" of schools to deliver more (or less) value-added to males than to females. Via Oaxaca (1973) style decomposition and following Card et al. (2013), we present the following expressions:

$$\begin{aligned}\mathbb{E}[\psi^M_{J(i,t)}|male] - \mathbb{E}[\psi^F_{J(i,t)}|female] &= \mathbb{E}[\psi^M_{J(i,t)} - \psi^F_{J(i,t)}|male] \\ &\quad + \mathbb{E}[\psi^F_{J(i,t)}|male] - \mathbb{E}[\psi^F_{J(i,t)}|female]\end{aligned} \tag{4}$$

$$\begin{aligned}\mathbb{E}[\psi^M_{J(i,t)}|male] - \mathbb{E}[\psi^F_{J(i,t)}|female] &= \mathbb{E}[\psi^M_{J(i,t)} - \psi^F_{J(i,t)}|female] \\ &\quad + \mathbb{E}[\psi^M_{J(i,t)}|male] - \mathbb{E}[\psi^M_{J(i,t)}|female]\end{aligned} \tag{5}$$

The first terms shown on the right-hand side of Equation (4) is the "learning" effect: the average difference in value-added that female students obtain relative to males across the distribution of male students. The second term is the sorting effect, which is obtained from the difference in the value-added of females observed across the distribution of male students versus the distribution of female students. In the alternative decomposition shown in Equation (5), the learning effect is computed by from the distribution of female students. The sorting effect is calculated using male students' value-added scores and by comparing them across the distribution of male and female students.

These decompositions are useful in determining whether different channels explain the importance of between-gender differential contributions to the gender gap. A strongly positive (negative) sorting effect denotes that females are under(over)-represented in schools with high value-added outcomes for women. A large positive (negative) learning effect implies that females are receiving fewer (more) value-added resources from schools than males.

2.5 Gender-specific returns to school characteristics in value-added

We finally investigate whether school characteristics have different returns depending on gender in terms of school value-added. To this end, we regress by OLS the estimated value-added for a vector of school characteristics z_i:

$$\hat{\psi}^F_j = z_j\delta^F + \omega_j \tag{6}$$

$$\hat{\psi}_j^M = z_j \delta^M + \phi_j \tag{7}$$

From these regressions, we can relate school characteristics to an estimated school value-added measure that is controlled for time-invariant student characteristics, school/student trends and matching effects to avoid confounding factors shaping estimated returns. All regressions are weighted by school size (students-year) and include municipality dummies, and errors $\omega_j and \phi_j$ are clustered at the municipality level.

Note that we are estimating an equation for female value-added and another for male value-added. In using this approach, we determine whether the same school characteristic yields a different return to value-added depending on gender. To further illustrate related differences, we run the following regression:

$$\hat{\psi}_j^M - \hat{\psi}_j^F = z_j \times \gamma + e_{it}$$

which corresponds to the same regression as that shown above for school characteristics but with a different value-added measure (i.e., $\psi_j^M - \psi_j^F$) used as the dependent variable. It is straightforward to show that $\gamma = \delta^M - \delta^F$. This regression allows for one to easily compute differences in returns for each characteristic in z_j and to also determine if these differences are statistically significant.

3 Data

To conduct the analysis described in the above section, we use data from the System for Measuring the Quality of Education (SIMCE) test obtained from Agency for Education Quality datasets, which are available upon request (MINEDUC, 2013). The SIMCE test was standardized in 1999 to follow up on school performance and is based on the item response theory (IRT) methodology used for the PISA test. As the SIMCE is applied every year to several students, we can use its yearly datasets to build a representative panel, from which we observe 108,555 female students and 105,415 male students who belong to approximately 6900 mixed-gender schools and who were tested in 2007, 2011 and 2013 or in the 4th, 8th and 10th grades.

Besides using representative data on individual test scores, our dataset also includes individual identifiers, unique school identifiers and the genders of students. By using unique school identifiers, we use school information such as types of dependence (i.e., if they are private or public) in our heterogeneity analysis. These and other school and teacher characteristics are measured in surveys applied to teachers and parents of students who had participated in the SIMCE test. We use this information to explore the mechanisms that may explain our results.

Figure 1 shows the mean scores of female and male students with different school grades included in our sample. We also include in the graph confidence intervals used for the mean score at the 95% level. A consistent and significant difference is observed between the girls' and boys' scores that persists through the observed period. Figure 2 shows the share of girls achieving each percentile of standardized SIMCE scores. The fraction of females drops substantially at higher SIMCE percentiles. In 10th grade, the declining pattern is even more pronounced. These patterns are typical of other student cohorts for Chile (Bharadwaj et al., 2016) and for other countries (Mead, 2006; Fryer and Levitt, 2010).

Table 1 shows some basic descriptive statistics for our panel. In Column (1), we report sample mean test scores, standard deviations, the number of observations, and the number of students and schools included in the sample for males and females. Our "overall sample" (the one summarized in Column (1) of Table 1) consists of students enrolled in mixed-gender schools who appear at least two times in the SIMCE score database. We observe a (cross-sectional) gender gap of 8.35 points from the SIMCE test results that corresponds to 0.17 of a standard deviation[3]. As noted by AKM and above, the school effects in a two-way fixed effects model such as (1) are only identified within a connected set of schools linked by student mobility. We use the algorithm developed in Correia's (2016) work to obtain this subset of schools and show the same descriptive statistics in Column (2). This subset, known as the largest connected set of schools, corresponds for both males and females to 99.9% of the students and to 99.7% of the schools in the original samples. Additionally, sample means and standard deviations do not considerably

[3]We standardized the results to the observed mean and standard deviation for each year

change between samples. Our restriction to the largest connected set does not importantly affect our data because high school mobility is observed in our sample. In total, 34% of the students observed in 2007 changed schools between 4th and 8th grade, and 73% of the students observed in 2013 changed schools between 8th and 10th grade[4] This pattern of mobility for 8th grade and the overall high level of mobility observed have been documented for Chile (Larroulet, 2011).

As we use specific fixed effects for males and females estimated for each school simultaneously in the analysis depicted in the previous section, for consistency and following Card et al. (2016), we also use the "dual connected set" of schools. The dual connected set of schools includes schools belonging to the largest connected set for both male and female students. We can see from Column (3) of Table 1 that the dual connected set represents 95,3% of students and 94,9% of schools for the female sample and 97,3% of students and 95,6% of schools for the male sample. As is found for the largest connected set of both males and females, sample means and standard deviations of scores of the dual connected set do not largely deviate from overall sample statistics.

4 Results

In this section we show the estimation results of our model in Equation (1). Columns (1) and (2) of Table 2 summarize parameter estimates and the fit of the model for female and male students for the largest connected set of each gender. The model includes a student fixed effect and a gender-specific school fixed effect, and elements of $x_{it}'\beta$ are time dummies. Due to the dimensionality problems that estimating student and school fixed effects pose, we use the numerical method developed by Correia (2016) to make our OLS estimates.

[4]The mobility level of our sample is higher than the mobility level of firms examined in other AKM studies. For example, Abowd et al. (1999) find a mobility level of 27,3% in their sample, and Alvarez et al. (2018) observe mobility levels of 25% to 40% in their sample, reflecting the largest connected set corresponding to 98% of their full sample. Therefore, is not surprising that our largest connected set represents the whole sample almost entirely.

We first show standard deviations and the number of observations for the largest connected set of female and male students. We then show the number of effects estimated for students and schools, the root mean squared error (RMSE) and adjusted R^2 statistics of the model, standard deviations for $(x_{it}\hat{\beta})$, the estimated school and student effects $(\hat{\alpha}_i, \hat{\psi_j^M}, \hat{\psi_j^F})$ and the correlation between the estimated student and school fixed effects by gender. Note that we obtain a high adjusted R^2 statistic, showing that we have captured much of the variance in test scores with our AKM model of student, school and time effects for both female and male students. For both genders, the standard deviations of student effects are more than twice the standard deviations of school effects, implying that inequality/variance in test scores for both genders is mostly attributed to student characteristics. The correlation found between student and school effects is positive for both genders, indicating that male and female students with a strong individual component that impacts their test scores are disproportionately represented in schools that impact students' scores more. This is equivalent to saying that positive associative matching occurs between higher value-added schools and students with better skills or a social background that enables them to achieve better scores.

The central panel of Table 2 shows fit statistics for a generalized model that includes a dummy for every student-school match. The model relaxes the additive structure of (1), and if match effects are important in predicting test scores, we expect to find a substantial improvement in fit by estimating it. The RMSE of the match effects model is 1,9% lower, and the adjusted R^2 is just 1% higher than that of the AKM model for female students. For male students, the RMSE of the match effects model is 1% lower, and the adjusted R^2 is less than 1% higher than that of the AKM model. With this exercise, we confirm that our value-added measures are consistent and not confounded with matching effects.

We now analyze the distribution of the obtained estimates. As shown in Section 2.2, we can use the obtained distribution of $(\hat{\psi}_j^F, \hat{\psi}_j^M)$ for simple variance decomposition and to study what explains the variance in test scores observed for both female and male students. The last panel of Table 2 shows the main components of (fitted) variance decomposition illustrated in Equation (2). For both female and male students, individual effects on scores account for approximately 67,5% of overall test score variations and for

9,6% of school effects. These similarities are also evident when we graph the distributions of school effects as shown by (weighted by gender) kernel density estimates presented in Appendix Figure A1. For the other components of test score variance, we find that for female students, the correlation between school and individual effects accounts for 3,4% of the variance in scores, that returns to time trends and associated correlations account for just 2.3%, and that the share of residuals accounts for approximately 15% of the score variance. For male students, the correlation between school and individual effects accounts for 2,9% of the variance, returns to time trends and associated correlations account for 2.9%, and the share of residuals accounts for 15,5% of the total variance.

These results indicate that heterogeneity or inequality in test scores is mainly driven by heterogeneity in students' characteristics. Regarding the gender gap, we do not find differences between the share of variance explained by school effects in female scores and the share of variance explained by school effects in male scores. Therefore, we find no evidence that schools make, in aggregate, female scores more unequal relative to those of male students. This finding serves as a preview of our next set of results, which show that schools do not seem to have a significant impact on test scores overall.

4.1 Value-added differentials and the math gender gap

We next use the estimated school effects to obtain a between-gender differential $\hat{\psi}_j^M - \hat{\psi}_j^F$ for every school. Recalling the framework of Equation (3), this term of expectation represents the average contribution of a school to the math gender gap. As noted above, we use the dual connected sample for this part of the analysis, as we are considering both female and male school effects simultaneously. We can only estimate (ψ_j^F, ψ_j^M) from OLS for a school or group of schools and interpret means of the estimate effects better, and thus, we normalize them with respect to schools belonging to the lowest decile of average test scores. The top row of Table 3 shows terms involved in the school effects component of the average gender gap, which is the sample analog of $\mathbb{E}[\psi_j^M|male] - \mathbb{E}[\psi_j^F|female]$.

Column (1) of Table 3 shows that the gender gap of the dual connected set is valued at roughly 0.17 standard deviations, which is similar to the gap observed for the overall sample. Columns (2) and (3) show mean normalized estimated school value-added/school effects for female and male students. Column (4) shows the difference observed between the two, which is the contribution of between-gender differences observed in schools to the math gender gap. We find that the average differential is almost negligible with schools making a negative contribution to the gap of less than 0.01 standard deviations. This result is consistent with previous evidence showing that estimations of the gender gap do not change when estimating the gender gap for schools with respect to the simple average gender gap[5] (Bharadwaj et al., 2016; Carrell et al., 2010). Our approach is advantageous in that we estimate gender-specific school effects while at the same time controlling for a student fixed effect. Therefore, we can directly determine if schools impact students' test scores differently depending on gender. We confirm these previous results by showing that schools on average and while controlling for time-invariant student characteristics do not impact the test scores of girls and boys differently because of gender.

We now focus on the impact of different kinds of schools on the gender gap. By conditioning the average gender gap only to private, voucher and public schools, we can check for heterogeneity in the impact of the between-gender differential produced at schools. Private, voucher and public schools characterize schools of the Chilean system, and they differ mostly in terms of resources and management characteristics. The last three rows of Table 3 show the same decomposition based on Equation (3) but while separating the sample and calculating terms according to the type of school considered. Note that we find a sizeable average value-added result for private schools for both genders and higher value-added results for voucher schools relative to those of public schools. The magnitudes found are consistent with the literature that addresses differences in types of schools found in Chile[6]. We find evidence of heterogeneity in the impact of the between-gender differential. We find that for public and voucher schools, the gender gap is similar to that

[5] Bharadwaj et al. (2016) and Carrell et al. (2010) estimations of the cross-sectional gap do not change when controlling/adjusting for school fixed effects

[6] See Paredes and Drago (2011) for a review of voucher school correlations with test scores and Mizala and Romaguera (2000a) for evidence of the large correlation between private schools and test scores

observed for the full sample, and the impact of the between-gender differential on the gap decreases in magnitude relative to that of the full sample. Specifically, we find that for public schools, the contribution of the between-gender differential favors females by less than 0.01 standard deviations. For voucher schools, the effect is favorable for males but also within less than 0.01 standard deviations. For private schools, we find that the gap is 38,6% smaller than that of the full sample, and we find that the between-gender differential is mostly favorable to female students, providing 0.07 standard deviations more in value-added. In the following sections, we investigate mechanisms that may shape this heterogeneity and whether school characteristics yield different returns depending on gender.

4.2 Mechanisms: Decomposition of the between-gender differential

To address the mechanisms of the above findings, we now turn to the results of our Oaxaca-style decomposition of the between-gender differential into sorting and "gender-specific value-added" effects by showing the fitted version of Equations (4) and (5). For the US, there is evidence of gender sorting across schools and within school types (Long and Conger, 2013b,a) that explains gender differences in college enrollment levels. We test whether a similar mechanism of gender sorting across schools/among school types translates into differences in overall value-added obtained by females relative to males. The part of the value-added differential that is attributable to a sorting channel is calculated by taking the difference of male value-added weighted by shares of female versus male students or the difference between female value-added weighting by shares of female versus male students. Table 4, Columns (2) and (3) show that the overall sorting effect is small: when using female effects to estimate sorting, we find that it explains a reduction in the gap of less than 0.01 standard deviations. When using male effects, this corresponds to a reduction in the gap of 0.01 standard deviations. The obtained result indicates that potential gender sorting at schools is not translating into better or worse value-added outcomes obtained by girls. This is consistent with existing evidence for the US showing that even if non-random gender sorting does occur in schools, parents value school attributes similarly in relation to their sons and daughters (Long and Conger, 2013b).

The gender-specific value-added effect is the average difference found between value-added results that girls receive relative to those of males at all schools. This is computed by taking the average differential across the distribution of girls/boys. Columns (4) and (5) of Table 4 show the "gender-specific value-added" effect, and we find that the overall impact of this effect is also small. We find that the difference in $\hat{\psi}_j^M - \hat{\psi}_j^F$ across the distribution of (weighted) females contributes positively to the gap within less than 0.01 standard deviations. However, when observed across the distribution of males, it helps close the gap by a similar magnitude. As sorting and learning effects observed are small, we find that compensation does not explain the small impact of the school between-gender differential observed between the two effects.

From the same heterogeneity exercise as that described above, when we restrict our analysis to private, voucher and public schools, we can check for heterogeneity in sorting and gender-specific effects as well. The lower panel of Table 4 shows the same decomposition of the between-gender differential but while restricting the sample to the school of the indicated type using the estimated school effects. Public schools seem to reduce the gap on average but by a small magnitude (-0.01 standard deviations), while voucher schools expand the gap on average but also by a small magnitude (0.01 SDs). In terms of sorting effects, we find the same pattern as that of the overall analysis. Rather, when observing female effects, we find a small effect of less than 0.01 standard deviations in magnitude, while when observing male effects, we find that sorting effects close a small portion of the gap by roughly 0.01 standard deviations. When observing the learning effects of public and voucher schools, in all cases, they are positive but small in magnitude (less than 0.01 standard deviations).

The results for private schools show that sorting effects are again small at approximately 0.01 standard deviations when considering male effects and of less than 0.01 standard deviations when considering female effects. However, the gender-specific effect is substantial with a magnitude of 0.06 standard deviations found when considering the female distribution and of 0.05 standard deviations when considering the male distribution. This result means that the reduction in the gender gap caused by private schools is mainly due to girls obtaining more value-added than boys enrolled at these schools (i.e., $\hat{\psi}_j^M - \hat{\psi}_j^F$ is negative and significant as expected for a given private school conditional on either gender). This result is useful for understanding the previous results on the math gender gap of elite schools. Ellison and Swanson (2010) finds that high-performing girls come from a small set of elite schools. This fact could be the result of simple sorting in school choice: girls being over-represented in high value-added schools. What we find for the elite schools of our sample (private ones[7]) is the opposite result; girls are achieving more value-added than boys. This means that private schools could be helping girls develop their skills while in school. In the following section, we further our analysis by correlating school and average teacher characteristics to potentially identify factors that shape gender-driven differences in school value-added.

4.3 Mechanisms: Gender-specific returns to school characteristics

As a final exercise to address mechanisms that shape the impact of gender-specific value-added on the math gender gap, we correlate the estimated value-added results with school characteristics. We in turn determine whether characteristics of schools yield different returns to value-added depending on gender. Using unique school identifiers included in our database, we can obtain teacher and school characteristics and then use them as an independent variable in the simple regressions depicted in Equations (6) and (7). The advantage of the teachers' survey lies in the fact that we can identify teachers who have actually interacted in classrooms with the students of our sample for the observed period. Table 5 shows the weighted (by school size[8]) OLS results of the estimation. The estimated coefficients shown in Table 5 are given in standard deviations, as this is the

[7]Private schools are considered elite schools, as they are much more expensive than voucher schools (with co-payment), and they represent less than 10% off enrollment in the several cohorts (Mizala and Romaguera, 2000b,a)

[8]We weight by school size to precisely estimate and test the statistical significance of the difference in returns as shown in Column (3) of Table 5. Appendix Table A1 shows the results derived when we weight the regressions by the quantity of girls and boys in each regression accordingly, noting that they are not qualitatively different from the results shown in Table 5.

measure of school value-added. Regressions include municipality fixed effects, and the error is clustered at the municipality level. Column (1) shows the regression applied when the dependent variable is female value-added, and Column (2) shows the regression conducted on male value-added.

Column (3) shows our regression of the differential applied on the same set of characteristics, therefore obtaining $\hat{\delta}^F - \hat{\delta}^M$. This allows for us to easily calculate the difference in returns to value-added for girls relative to boys on school/teacher characteristics. It also allows for us to conduct a formal test of the statistical significance of the mentioned difference (t-test of the regression). A negative coefficient shown in Column (3) of Table 5 denotes that the characteristics favor female students over males, and vice-versa.

We find that the share of female teachers in a school correlates with a higher value-added for female students consistent with the previous literature for Chile (Paredes, 2014; Bharadwaj et al., 2016) and other countries (Carrell et al., 2010; Dee, 2007), showing positive effects of gender-matching for female students. The difference in estimated returns found is not statistically significant, but we find that the effect on each gender's value-added is strong. Additionally, teachers' average levels of education have an important effect school value-added as the literature finds (Clotfelter et al., 2010), but no large/significant differences are found in estimated returns for male and female students. We also control for the share of teachers who would leave a school if offered similar conditions, and we find no difference in the return to this teacher's satisfaction measure in value-added: the effect is negative and similar in magnitude for both genders.

The impact of teachers' expectations regarding their students has been studied in the literature (Gershenson et al., 2016b) in addition to its long-term effects on math scores (Hinnant et al., 2009). In particular, it has been shown that expectations have different effects depending on racial characteristics involved (Figlio, 2005; Ferguson, 2003) due to biases or self-fulfilled prophecies related to student characteristics (Hinnant et al., 2009; Gershenson et al., 2016a). We find that average expectations impact school value-added differently depending on gender. The share of teachers with low expectations (those who believe that the majority of their students will complete only high school) has a negative impact on value-added but with a higher magnitude for female students. Additionally, high expectations (the share of teachers who believe that their students will complete graduate studies) have a positive impact but are much more important for male value-added than for female value-added. This difference in returns is also statistically significant.

We control for the religious orientations of schools by using a dummy to indicate whether a school is Catholic to see if there is a differential return to this characteristic depending on gender. This effect of religious orientation could be attributed to the cultural particularities of such schools (e.g., second-generation immigrants present a larger gender gap due to cultural restrictions and traits (Nollenberger et al., 2016)). We find no statistically significant differences in the returns of Catholic schools depending on gender when controlling for all other characteristics.

Finally, we add dummies that indicate whether a school is a private or voucher school, and public schools are omitted. Even when controlling for religious orientations, characteristics, expectations of teachers and municipality fixed effects, we find that private schools yield a (statistically) significant difference in value-added that favors female students. Private schools could be hiring better teachers in the sense that they could be addressing certain difficulties that girls may face when taking standardized tests such as the SIMCE test studied here. In Niederle and Vesterlund (2010), it is noted that girls respond differently to competitive environments. Additionally, on average, girls make worse self-assessments than boys regarding their math abilities, and this correlates with their poorer test scores (Bharadwaj et al., 2016). Private schools could be providing a better environment for girls to perform better on the test. This could be the case because private schools, due to having access to more resources and better using them, may pay more for better teachers and for more qualified principals who are aware of the differences described above. In this way, private schools may be changing poorer self-assessments from girls or girls' responses to competitive environments. Further evidence is needed to address these alternative explanations and their relative importance, but we offer these explanations as cultural/religious orientations, the average gender matching of math teachers, teachers' expectations regarding students and school conditions for teachers do not seem to fully explain the impact of private schools on the gender gap.

4.4 Validity of the Exogenous Mobility Assumption

We now analyze the validity of the exogenous mobility assumption. To this end, we adapt an indirect test used in Card et al. (2013) and Card et al. (2016) to our educational focus. For this analysis, we first select male and female students who changed schools between 8th and 10th grade. We then compute the mean classmate test scores of every student that changed schools for each year and assign each the quartile of his or her mean classmate test scores for each year. In Figures 3 and 4, we plot the mean test scores measured before and after female and male students changed schools for the first (lowest) quartile of classmate scores and for the fourth (higher) quartile. The figures show that male and female students who move from schools with low classmate scores to schools with high classmate scores experience substantial average gains in their scores. The opposite occurs for students who move in the opposite direction. As noted above, this symmetry provides suggestive evidence showing that mobility is not correlated with matching premiums η_{ij}. Additionally, note that no important downward trends in scores are observed prior to a move, suggesting that mobility is not associated with adverse temporary shocks in schools. Symmetry found in score changes after a move also shows that positive score shocks are not a source of incentive to move. This means that we find no evidence of a correlation between mobility and the temporary school component of scores ξ_{jt}.

In Figure 5, we plot the average test scores of students who did not change schools in the sample period and who belong to schools of the fourth and first quartiles of average classmate scores. We note that they follow the same patterns of school movers who stay enrolled at schools of the same quartile. This indicates that even if trends are present, they are not correlated with mobility. If this were true, we should expect to find a different pattern for students who remain at the same school in the observed period. Rather, if matching effects η_{ij} shaped the mobility of students who remain at schools of the highest quartile, then “stayers” should present more horizontal movement in scores. If an upward trend incentivizes parents to have their children change schools, indicating the presence of a correlation between mobility and the ϵ_{it} of these students, we should again find more modest changes in the scores of those who do not change schools. However, what we observe in Figure 5 is that students who do not change schools follow the same patterns as those who do. The same occurs for students who move between schools with the lowest quartiles of classmate scores. They follow the same pattern of scores as those remaining at schools of the lowest quartile as shown in Figure 5. If only students who perform poorly were moving to the worst schools, we would expect to find those not changing schools to show no change in scores.

Appendix Tables A2 (females) and A3 (males) show the mean test scores of all mobility groups constructed from the 4 quartiles of mean classmate test scores. We find that patterns reflecting symmetry and trends not correlated with mobility also appear when observed the other mobility groups (those who change schools originating from schools of the 2nd and 3rd quartiles of mean classmate scores). The tables also show that for any mobility group, at least 1000 students move between different kinds of schools, ensuring the consistency of the estimated fixed effect for schools of different quantiles.

We note in Figures 3 and 4 and also in Appendix Tables A2 and A3 that the levels and quantities of students of different mobility groups are different. This is likely the case due to a relation between mobility and socio-economic or socio-demographic characteristics, which are contained in α_i (e.g., not being able to pay a school fee, which is positively correlated with test scores on average Mizala and Romaguera (2000b,a), or transportation costs related to moving to a better school (Gallego and Hernando, 2009)). Parents may also be sorting themselves in schools based on other socio-economic factors. For example, parents of a higher income group may value test scores more (Gallego and Hernando, 2009). Again, these household characteristics are contained in the time invariant component of scores α_i. As we control for student effects α_i, these sorting sources do not bias our results.

5 Concluding Remarks

In this paper, we study the contribution of schools to the math gender gap. From rich and representative micro-level Chilean data on test scores, we estimate a two-way fixed effects model (school and student) following the econometric framework developed by (Abowd et al., 1999). The AKM approach is advantageous in that it allows for us to estimate value-added results while controlling for time-invariant student characteristics. It also allows for us to test for different types of transitory shocks on school mobility that could bias our estimates. We use the estimates to explore the impact of schools on the math gender gap and mechanisms that shape these effects.

We find that on average, gender differences in value-added do not explain a substantial portion of the gender gap. However, this effect is found to be heterogeneous. Rather, private schools present a between-gender differential in value-added favoring women, explaining a substantial reduction observed in the gender gap. Public and voucher schools, in contrast, have a much smaller impact on the gap.

We then analyze the mechanisms that explain the above results. First, following the work of Card et al. (2016), we conduct a Oaxaca-style decomposition to separate the differential into sorting and gender-specific channels. We find that sorting effects are small across school types, suggesting that female students are not under- or over-represented at high-performing schools. In private schools, the gender-specific effect dominates. Rather, girls on average obtain more value-added than boys at these schools. We find that the share of female teachers yields a higher return for female students. We also find that average teacher expectations, whether positive or negative, favor male students. Finally, when controlling for several school and teacher characteristics, private schools correlate with higher value-added results for female students than for males.

Our results contribute new insights to the literature on the math gender gap. First, we find that controlling for student characteristics and ruling out biases resulting from mobility and schools' contributions to the gender gap seem to be negligible. Second, we find that this effect is heterogeneous with private schools significantly helping to close the gap by 38,6%. Second, we rule out the possibility that a pure sorting channel explains the reduction in the gender gap that private schools cause. It is a gender-specific effect that is behind the results found for private schools, which seems to be helping girls develop their skills. Third, we analyze several mechanisms that shape our results both within and beyond the scope of existing literature. That is, we find the differential returns of several school and teacher characteristics to value-added depending on gender.

References

Abdulkadiroglu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2017). Do parents value school effectiveness? Working Paper 23912, National Bureau of Economic Research.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Altonji, J. G. and Blank, R. (1999). Race and gender in the labor market. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, chapter vol. 3c, pages 3144–3259. Eslevier Science, Amsterdam.

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics*, 4(1):185–223.

Alvarez, J., Benguria, F., Engbom, N., and Moser, C. (2018). Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics*, 10(1):149–89.

Angrist, J. D., Hull, P. D., Pathak, P. A., and Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics*, 132(2):871–919.

Bharadwaj, P., Giorgi, G. D., Hansen, D., and Neilson, C. A. (2016). The Gender Gap in Mathematics: Evidence from Chile. *Economic Development and Cultural Change*, 65(1):141–166.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, 131(2):633–686.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, 45(3):655–681.

Correia, S. (2016). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Technical report. Working Paper.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, 42(3).

Ellison, G. and Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: Evidence from the american mathematics competitions. *Journal of Economic Perspectives*, 24(2):109–28.

Ferguson, R. (2003). Teachers' perceptions and expectations and the black-white test score gap. *Urban Economics*, 38(4):460–507.

Figlio, D. N. (2005). Names, expectations and the black-white test score gap. Working Paper 11195, National Bureau of Economic Research.

Fryer, R. G. and Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2):210–40.

Gallego, F. and Hernando, A. (2009). School choice in chile: Looking at the demand side. Working Paper 356, Instituto de Economía, Pontificia Universidad Católica de Chile.

Gershenson, S., Holt, S. B., and Papageorge, N. W. (2016a). Who believes in me? the effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, 52:209 – 224.

Gershenson, S., Kang, K., and Papageorge, N. W. (2016b). Teacher expectations matter. Working Paper 10165.

Hinnant, J. B., O'Brien, M., , and Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101(3):662–670.

Larroulet, C. (2011). Análisis de la movilidad escolar en chile. Master's thesis, Instituto de Economía, Pontificia Universidad Católica de Chile.

Long, M. C. and Conger, D. (2013a). Gender gaps in college enrollment: The role of gender sorting across public high schools. *Educational Researcher*, 42(7):371–380.

Long, M. C. and Conger, D. (2013b). Gender sorting across k–12 schools in the united states. *American Journal of Education*, 119(3):349–372.

Mead, S. (2006). The truth about boys and girls. Working paper, Education Sector.

MINEDUC (2013). Base de Datos de la Agencia de Calidad de la Educación 2007-2013. Santiago, Chile.

Mizala, A. and Romaguera, P. (2000a). Determinacion de factores explicativos de los resultados escolares en educacion media en chile. Working Paper 85, Centro de Economía Aplicada. Universidad de Chile.

Mizala, A. and Romaguera, P. (2000b). School performance and choice: The chilean experience. *Journal of Human Resources*, 35(2):392–417.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., and Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19(4):547–568.

Murnane, R. J., Willett, J. B., and Levy, F. (1995). The Growing Importance of Cognitive Skills in Wage Determination. *The Review of Economics and Statistics*, 77(2):251–266.

Niederle, M. and Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24(2):129–44.

Nollenberger, N., Rodríguez-Planas, N., and Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5):257–61.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

OECD (2017). *The Pursuit of Gender Equality*. OECD Publishing.

Paglin, M. and Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics*, 8(1):123–44.

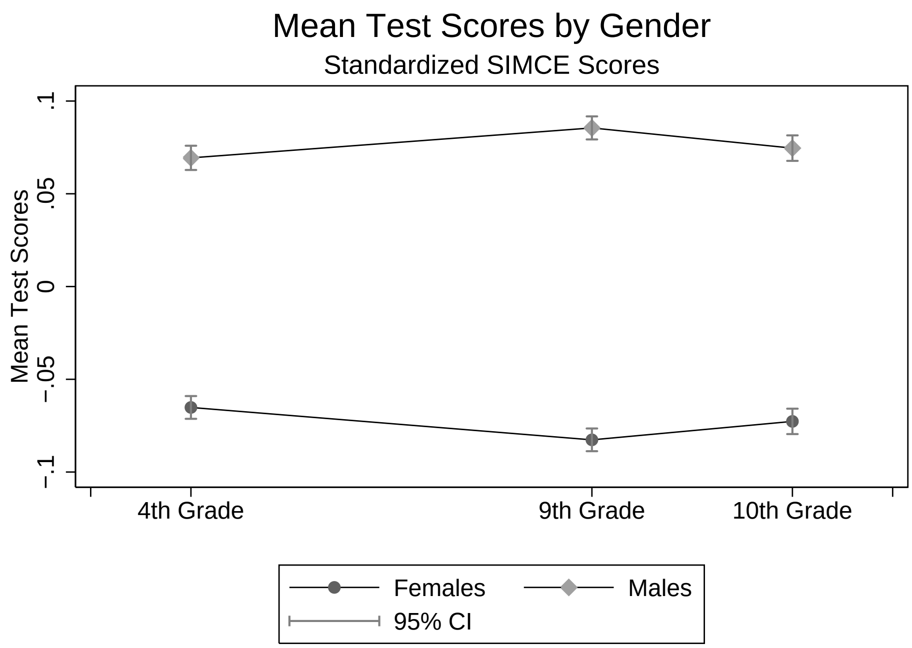
Paredes, R. D. and Drago, J. L. (2011). The quality gap in chile's education system. *CEPAL Review*, 104.

Paredes, V. (2014). A teacher like me or a student like me? role model versus teacher bias effect. *Economics of Education Review*, 39:38–49.

Weinberger, C. J. (1999). Mathematical college majors and the gender gap in wages. *Industrial Relations*, 38(July):407–413.

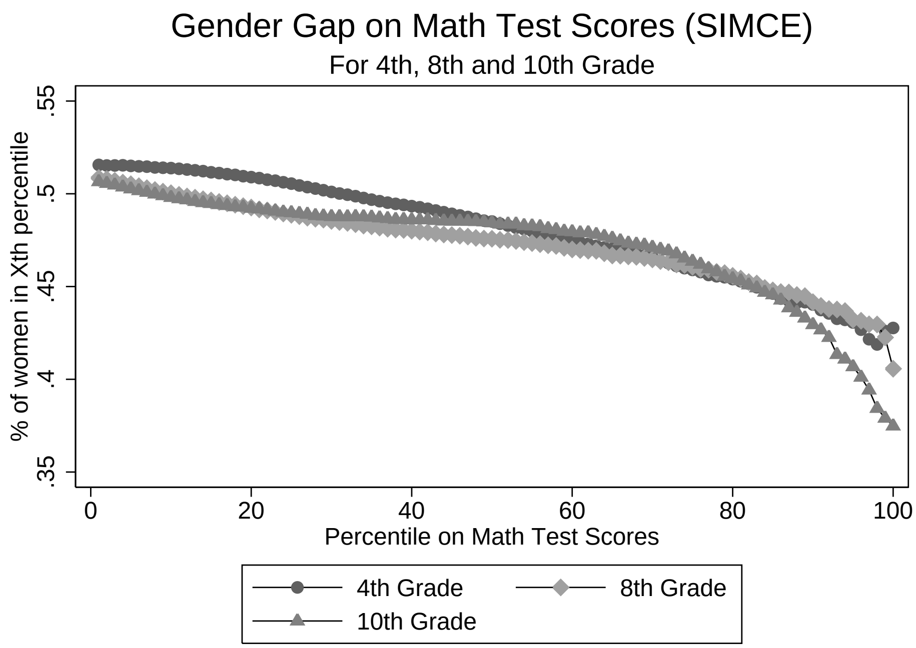
Weinberger, C. J. (2001). Is teaching more girls more math the key to higher wages? In King, M. C., editor, *Squaring Up: Policy Strategies to Raise Women's Incomes in the United States*, chapter 11. University of Michigan Press, Michigan.

Figure 1: Mean Test Scores of Male and Female Students



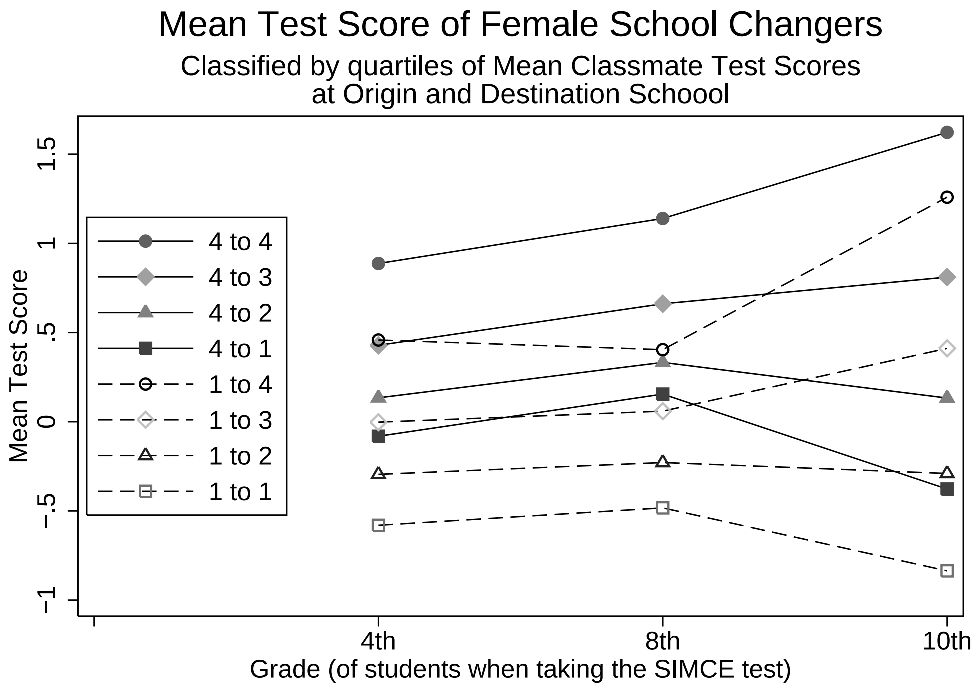
Notes: Mean standardized SIMCE score obtained by female and male students in the sample while attending school in different grades. 95% Confidence intervals are also shown.

Figure 2: Math Gender Gap on SIMCE Percentiles



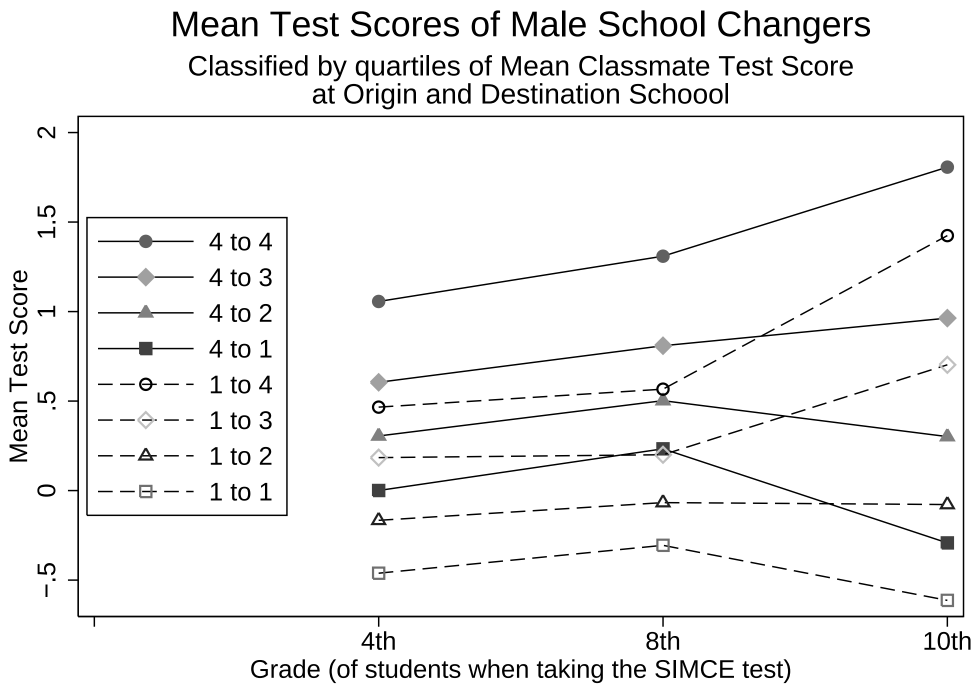
Notes: Every point shows the share of females present in each percentile of the SIMCE distribution.

Figure 3: Mean Test Scores of Female School Changers



Notes: Mean test scores of female students who changed schools between 8th and 10th grade in the sample period. Each school is classified into quartiles based on the mean test scores of classmates in 8th and 10th grade. See the text for additional information.

Figure 4: Mean Test Scores of Male Students Changing Schools

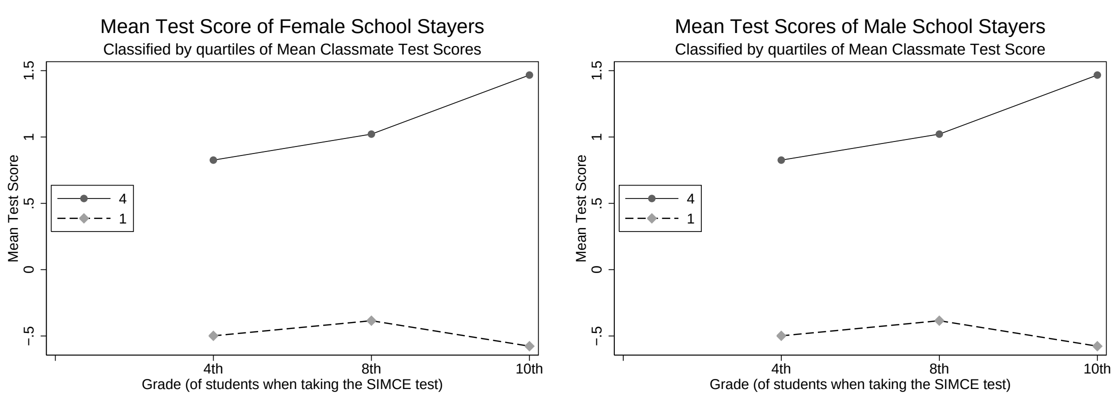


Note: Mean test scores of male students who changed schools between 8th and 10th grade in the sample period. Each school is classified into quartiles based on the mean test scores of classmates in 8th and 10th grade. See the text for additional information.

Figure 5: Mean Test Scores of Female and Male Students Not Changing Schools in the 1st and 4th Quartiles

(a) Female stayers

(b) Male stayers



Panel (a) shows the mean test scores of female students who did not change schools in the sample period. Panel (b) shows the same means but for male students only. Each school is classified into quartiles based on the mean test scores of schools for every year. We only show test score means for students who stayed at schools belonging to the first and fourth quartiles. See the text for additional information.

Table 1: Descriptive Statistics for Different Samples by Gender

	Overall Sample (1)	Largest Connected Set (2)	Dual Connected Set (3)
Female Math Test Scores			
Sample Mean Score	258.61	258.60	256.98
Std. Dev.	55.55	55.55	55.31
Student-Year Observations	282,568	282,455	253,642
Number of students	108,555	108,496	103,415
Number of schools	6,954	6,936	6,596
Male Math Test Scores			
Sample Mean Score	266.96	266.96	265.30
Std. Dev.	56.07	56.06	55.58
Student-Year Observations	271,065	270,993	255,043
Number of students	105,415	105,377	102,612
Number of schools	6,867	6,848	6,596
Raw Gender Gap	8.35	8.34	8.32
Standardized Gender Gap	0.17	0.17	0.17

Column (1) shows summary statistics for the sample of female and male students enrolled at mixed gender schools tested in 2007, 2009 and 2013. Columns (2) and (3) show the same statistics for the largest connected set and dual connected set (see the text for definitions) for the sake of comparison.

Table 2: Summary of AKM Estimations by Gender

	All Females (1)	All Males (2)
Std Dev of Test Scores	1.11	1.12
Student-Year Obs	282455	270993
Summary of Parameter Estimates		
Number of Student Effects	108496	105377
Number of School Effects	6936	6848
RMSE of Estimation	0.56	0.58
Adjusted R2	0.75	0.74
Std Dev of Student Effects	0.91	0.92
Std Dev of School Effects	0.34	0.35
Std Dev of Xb	0.17	0.19
Corr of Student-School Effs	0.13	0.11
Match Effects Model		
RMSE of Estimation	0.55	0.57
Adjusted R2	0.76	0.74
Inequality Decomposition		
Share of variance due to:		
Student Effs	67.47	67.43
School Effs	9.56	9.60
Corr of School-Student Effs	3.41	2.90
Share of Xb	2.27	2.87
Share of Residuals	14.88	15.51

The table shows a summary of AKM estimations, match effects models and variance decompositions of test scores. Estimated models include student fixed effects, school fixed effects and time dummies. Match effects models include a dummy for each student-school match and time dummies. Samples only include observations made of the largest connected set.

Table 3: Contributions of Between-gender Differentials in School Effects to the Math Gender Gap

	Test Score Gender Gap (1)	Mean Female value-added (2)	Mean Male value-added (3)	Contributions of Schools to the Gap (4)
All Schools	0.17	0.19	0.18	-0.01 (-3.16%)
By Type of School:				
Public	0.18	0.12	0.12	-0.00 (-1.90%)
Voucher	0.17	0.19	0.19	0.00 (1.91%)
Private	0.11	0.54	0.47	-0.07 (-38.55%)

The sample includes female and male students of the dual connected set of schools (see Table 1, Column 3). Column (1) shows the difference in the mean test scores of males and females estimated across all students of the subset of schools indicated by row name. Estimated school effects derive from the models described in Table 2 and are normalized with respect to schools of the lower decile of test scores. Column (4) shows contributions of the between-gender differential in school effects to the gap, showing in parentheses the ratio of corresponding entry over the average gender gap for all schools (0.17 SDs).

Table 4: Sorting and "Learning" Effects in Between-gender Differentials of School value-added

		Sorting Effect		"Gender-specific" Effect	
	Contributions of Schools to the Gap	Using Female Effects	Using Male Effects	Female Distribution	Male Distribution
	(1)	(2)	(3)	(4)	(5)
All Schools	-0.01	-0.00	-0.01	0.00	-0.00
By Type of School:					
Public	-0.00	-0.00	-0.01	0.00	0.00
Voucher	0.00	0.00	-0.01	0.00	0.00
Private	-0.07	-0.01	-0.00	-0.06	-0.05

A decomposition of the between-gender differential of school effects estimates as described in the text (see Section 2.4) is shown. Column (1) shows the between-gender differential. Columns (2) and (3) show the sorting effect of the differential, and Columns (4) and (5) show the gender-specific value-added effect of the differential.

Table 5: Correlation between School Characteristics and School value-added

	(1) Female value-added	(2) Male value-added	(3) Difference
Share of:			
Female Math Teachers	0.057***	0.046***	-0.010
	(0.017)	(0.015)	(0.011)
Teachers with Graduate Studies	0.093***	0.095***	0.002
	(0.017)	(0.015)	(0.009)
Teachers who disagree with School conditions	-0.050***	-0.063***	-0.013
	(0.019)	(0.018)	(0.014)
Teachers who expect Students to complete only High School	-0.148***	-0.086**	0.063**
	(0.036)	(0.036)	(0.027)
Teachers who expect Students to complete Graduate Studies	0.142*	0.219***	0.078
	(0.073)	(0.078)	(0.053)
Other School Characteristics:			
School is Private	0.359***	0.314***	-0.045*
	(0.032)	(0.025)	(0.027)
School is Subsidized/Voucher	0.080***	0.097***	0.017*
	(0.018)	(0.015)	(0.010)
School is Catholic	0.084***	0.086***	0.002
	(0.017)	(0.014)	(0.011)
N	5798	5798	5798
R^2	0.218	0.220	0.179

Estimated parameters derived from regressions of gender-specific value-added with school characteristics used as explanatory variables are shown. Column (3) shows the difference between male and female value-added as the dependent variable. Regressions are weighted by school size and include municipality fixed effects. Standard errors clustered at the municipality level are shown in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix

Figure A1: Kernel Density Estimate of Male and Female value-added

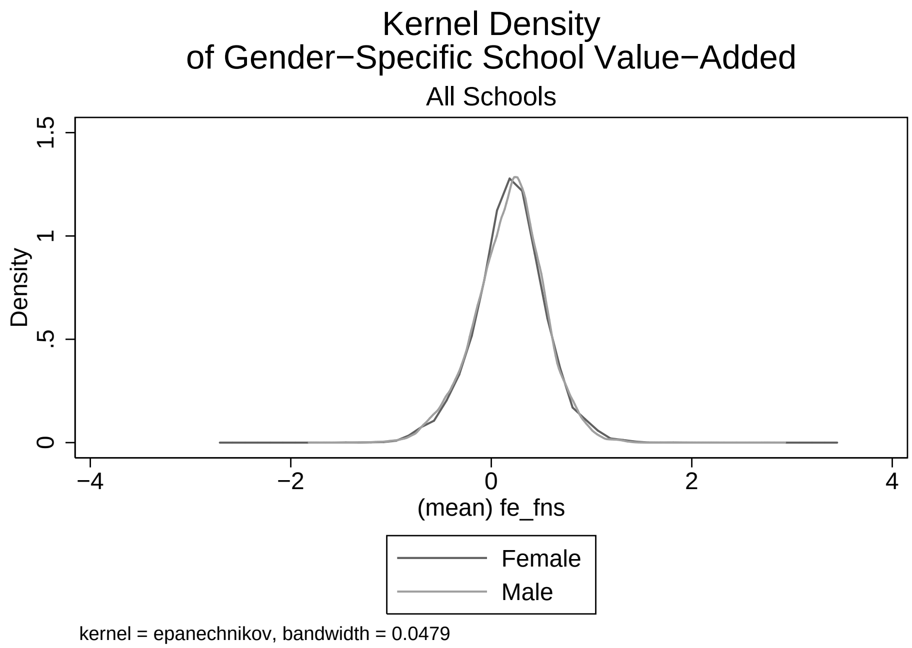


Table A1: Correlation between School Characteristics and School value-added - Gender Weighted

	(1) Female value-added	(2) Male value-added
Female Math Teachers	0.037***	0.024*
	(0.014)	(0.014)
Teachers with Graduate Studies	0.058***	0.077***
	(0.014)	(0.013)
Teachers who disagree with School conditions	-0.053***	-0.064***
	(0.016)	(0.016)
Teachers who expect Students to complete only High School	-0.096***	-0.051
	(0.028)	(0.032)
Teachers who expect Students to complete Graduate Studies	0.148**	0.229***
	(0.068)	(0.069)
School is Private	0.342***	0.282***
	(0.030)	(0.023)
School is Voucher	0.063***	0.076***
	(0.014)	(0.013)
School is Catholic	0.058***	0.070***
	(0.014)	(0.012)
N	5798	5798
R^2	0.234	0.225

Estimated parameters from (weighted by gender) regressions of gender-specific value-added with school characteristics as explanatory variables are shown. Column (3) shows results derived for the difference between male and female value-added as a dependent variable. Regressions include municipality fixed effects. Standard errors clustered at the municipality level are shown in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Number and Scores of Females Changing and Not Changing Schools of Different Groups

Group	N (1)	Score in 4th (2)	Score in 8th (3)	Score in 10th (4)
1 stayers	4882	-0.499	-0.385	-0.577
1 to 1	7508	-0.581	-0.483	-0.836
1 to 2	8109	-0.295	-0.229	-0.290
1 to 3	3332	-0.002	0.059	0.411
1 to 4	1359	0.458	0.404	1.259
2 stayers	12046	-0.090	-0.002	0.054
2 to 1	7451	-0.459	-0.320	-0.707
2 to 2	10712	-0.198	-0.097	-0.178
2 to 3	5259	0.171	0.211	0.517
2 to 4	2583	0.498	0.554	1.313
3 stayers	25183	0.335	0.430	0.711
3 to 1	5566	-0.338	-0.148	-0.556
3 to 2	9245	-0.067	0.076	-0.059
3 to 3	6674	0.249	0.403	0.625
3 to 4	3892	0.667	0.823	1.447
4 stayers	35155	0.826	1.022	1.467
4 to 1	2326	-0.081	0.155	-0.377
4 to 2	4192	0.134	0.333	0.133
4 to 3	4174	0.429	0.661	0.811
4 to 4	4338	0.887	1.140	1.623

The table shows the average test scores (columns 2 to 4) and numbers (column 1) of female students moving from/to schools representing different quartiles of average classmate scores.

Table A3: Number and Scores of Males Changing and Not Changing Schools of Different Groups

Group	N (1)	Score in 4th (2)	Score in 8th (3)	Score in 10th (4)
1 stayers	4744	-0.361	-0.218	-0.392
1 to 1	7202	-0.462	-0.306	-0.613
1 to 2	7276	-0.166	-0.067	-0.078
1 to 3	4219	0.184	0.200	0.703
1 to 4	1003	0.466	0.567	1.424
2 stayers	11451	0.071	0.193	0.233
2 to 1	6723	-0.297	-0.106	-0.468
2 to 2	9059	-0.005	0.083	0.064
2 to 3	6222	0.308	0.394	0.790
2 to 4	1843	0.707	0.774	1.580
3 stayers	22942	0.501	0.621	0.908
3 to 1	5280	-0.154	0.032	-0.352
3 to 2	7848	0.069	0.248	0.128
3 to 3	7146	0.369	0.536	0.819
3 to 4	3247	0.774	0.957	1.614
4 stayers	34037	0.993	1.195	1.688
4 to 1	2233	0.001	0.233	-0.292
4 to 2	3761	0.305	0.502	0.302
4 to 3	4487	0.605	0.809	0.964
4 to 4	3605	1.057	1.309	1.807

The table shows the average test scores (columns 2 to 4) and numbers (column 1) of female students moving from/to schools representing different quartiles of average classmate scores.