

Appendixes for online publication

B Data sources and cleaning

B.1 Data sources

The *Departament d'Ensenyament* (the ministry of education of the autonomous community of Catalonia) provided enrollment records for all schools in Barcelona, from primary education to upper secondary education, from the school year 2009/2010 to 2015/2016. For each year, data include school and class attended by the students and information on promotion or retention.¹ In this paper I focus on public middle schools, because information on final evaluations assigned by teachers at the end of the year, and on time-invariant characteristics such as gender and nationality are available only for public schools. The data also allow me to identify children with special education needs.

A separate data source from the ministry of education contains the applications to middle schools in 2009 and 2010. For each applicant, I observe the application list, priority points for the top choice, and whether he or she enrolled in the preferred school.

The *Consell d'Avaluació de Catalunya* (public agency in charge of evaluating the educational system) provided me with the results of standardized tests taken by all the students in the region attending 6th grade of primary school and 4th grade of middle school. Such tests are administered in the spring since 2008/2009 for primary school and since 2011/2012 for middle school. They assess students' competence in Maths, Catalan, and Spanish.² These exams are externally designed and graded. In this paper I refer to the test scores as external evaluations, in contrast with the final evaluations given by teachers in the school, which I call internal evaluations. The tests are administered in two consecutive days in the same premises in which students typically attend lectures. Normally, every student is required to take all the tests, however children that are sick one or both days and do not show up at school are not evaluated.³

Information on the student's family background, more specifically on parental education, are collected from the Census (2002) and local register data (*Padró*). When the information can be retrieved from both sources, I impute the highest level of education, presumably the most up-to-date information.

The previous data sources have been anonymized by the Institut Català d'Estadística (IDESCAT), which provided unique identifiers to merge them. Finally, I use publicly available data from the national Tax Agency (*Agencia Tributaria*) for the average income at the postal code level. I use

¹The IT infrastructure that supports the automatic collection of data has been progressively introduced since the school year 2009/2010. By year 2010/2011 most of the schools have already adopted it, while some are missing for 2009/2010. 47 middle schools have sufficient data from year 2009. See next Subsection B.2 for more details.

²The tests are low stakes, because they do not have a direct impact on student evaluations or progress to the next grades but they are transmitted to the principal of the school, who forwards them to the teachers, families and students. More information can be found at: http://csda.gencat.cat/ca/arees_d_actuacio/avaluacions-consell (in Catalan)

³Evaluations can be missing for three reasons: 1. the student did not show up the day of the test; 2. the student did not attend primary school in Catalonia, she moved in the region only when she started middle school; 3. the student did take the test, but due to severe misspelling in the name or date of birth it was not possible to match the information with the enrollment data.

data for year 2013, the first one for which information are available.⁴

B.2 Sample creation and data cleaning

The initial sample consists of students who enroll for the first time in lower secondary education in September 2009 or September 2010 in a public middle schools in Barcelona. I drop about 2.5% of students who have special education needs. In fact, special need children may have a personalized curriculum, and follow different retention rules. Therefore it would not be appropriate to include them in the estimation of the model. Furthermore, I focus the analysis on students for whom I could retrieve the evaluations in primary school, namely about 80% of the students who enroll in a public middle school in the period under analysis.

Further data cleaning includes: 1. dropping 13 students who were younger than 12 years old or older than 13 when they enroll⁵; 2. 89 students who appear in the enrollment data only once at 12/13 years old, assuming that they moved in another region; 3. 45 students for whom I could not retrieve internal evaluations in first or second grade.

Less than 10% of the students change school during lower secondary education. About half of them stay in the public system, while the other switch to a non-public school. I observe enrollment data for the latter, but I don't have information on performance, therefore I drop them from the analysis. Conversely, students who change school within the public system are included in the final sample. They are assigned the middle school that they attend for the first year, and the peers that they would have if they staid in the same class.⁶

Finally, I restrict the analysis to the schools with enough students in each cohort, and in which those students are a large share of the children enrolled in first grade. More specifically, I focus on schools with at least 18 students in the sample in each cohort (dropping 5 schools) and with more than 40% of pupils in first grade who belong to the sample (dropping 1 school). The chosen thresholds exclude about 3% of the students in the selected sample. Thus, I exploit 47 out of 53 public middle schools that appear in the enrollment data in the time period under analysis.⁷ The final sample consists of 1540 students.

C Model: modeling choices and limitations

This appendix sketches a highly simplified static version of the model to illustrate the issues discussed in Section 3.6 of the paper. For simplicity, the student index i is omitted.

Suppose that a student's cognitive skill is given by

$$C = h + T + \alpha x \tag{A-1}$$

where $h \sim \mathcal{N}(0, \sigma)$ is cognitive ability, unobserved by both the student and the econometrician,

⁴See <https://www.agenciatributaria.es>.

⁵As explained in Section 2.1 of the paper, usually students are 12 years old when they enroll (or turn 12 before the end of the year). They are 13 years old if they repeat a year during primary education.

⁶This approach introduces some measurement error, however results are very similar if those observations are omitted from the sample.

⁷In a previous version of this paper, I imposed more restrictive thresholds and exploited data from only 44 schools. Overall, results are very similar.

$T \sim \mathcal{N}(0, \tau)$ is a permanent non-cognitive trait (e.g. motivation), observed by the student but unobserved by the econometrician, and x is an observable characteristic. Assume that $\text{Cov}(h, T) = 0$.

The student receives an evaluation

$$r = C + \epsilon = h + T + \alpha x + \epsilon, \quad (\text{A-2})$$

where $\epsilon \sim \mathcal{N}(0, \rho)$ is i.i.d. and independent of all other variables. From the econometrician's perspective, this equation can be rewritten as

$$r = h + \alpha x + e, \quad (\text{A-3})$$

where $e = T + \epsilon$. While ϵ is i.i.d., the composite error term e may be correlated with x if $\text{Cov}(T, x) \neq 0$.

Because the student observes T , they can construct the signal

$$s = r - T - \alpha x = h + \epsilon. \quad (\text{A-4})$$

Given the prior on h , the student's posterior belief about ability is

$$m = \frac{\sigma}{\sigma + \rho}(h + \epsilon), \quad \omega = \frac{\sigma \rho}{\sigma + \rho}, \quad (\text{A-5})$$

The student therefore believes skills to be

$$\mathbb{E}(C) = m + T + \alpha x = \frac{\sigma}{\sigma + \rho}(h + \epsilon) + T + \alpha x. \quad (\text{A-6})$$

The student chooses whether to continue schooling according to

$$u = \phi \mathbb{E}(C) + \beta x + \gamma T + \varepsilon \geq 0, \quad (\text{A-7})$$

where ε is a logistic error term.

The econometrician, who ignores the presence of T , observes

$$s^* = r - \alpha x = h + T + \epsilon \quad (\text{A-8})$$

and treats it as a noisy signal for ability. As a result, the econometrician (incorrectly) computes the posterior mean and variance

$$m^* = \frac{\sigma}{\sigma + \rho}(h + T + \epsilon), \quad \omega^* = \omega = \frac{\sigma \rho}{\sigma + \rho}, \quad (\text{A-9})$$

while the correct posterior distribution would be

$$m_{\text{true}}^* = \frac{\sigma}{\sigma + \tau + \rho}(h + T + \epsilon), \quad \omega_{\text{true}}^* = \frac{\sigma(\tau + \rho)}{\sigma + \tau + \rho}, \quad (\text{A-10})$$

The econometrician's inferred belief about skills is therefore

$$E^*(C) = m^* + \alpha x = \frac{\sigma}{\sigma + \rho}(h + \epsilon) + \frac{\sigma}{\sigma + \rho}T + \alpha x, \quad (\text{A-11})$$

Substituting this expression into the utility function yields

$$u = \phi E^*(C) + \beta x + \eta = \phi \left[\frac{\sigma}{\sigma + \rho}(h + T + \epsilon) + \alpha x \right] + \beta x + \eta, \quad (\text{A-12})$$

where $\eta = \left(\phi \frac{\rho}{\sigma + \rho} + \gamma \right) T + \varepsilon$. While ε is i.i.d, η is correlated with $E^*(C)$ and may be correlated with x .

This highly simplified model highlights several implications:

1. $\omega^* < \omega_{\text{true}}^*$. The econometrician is overly optimistic about the precision of the signal. Even with repeated signals, they would become overconfident about learning ability, because precision about $h + T$ is mistaken for precision about h .
2. $\text{Var}(E^*(C)|x) < \text{Var}(E(C)|x)$.⁸ Conditional on observables, the distribution of beliefs about skills inferred by the econometrician is less dispersed than the true distribution of beliefs held by students. As a result, inferred beliefs are more tightly concentrated around the prior mean than the beliefs computed by students.
3. From the econometrician's perspective, T is part of the error term in the grade equation. Thus, if $\text{Corr}(T, x) \neq 0$, the estimated coefficient $\hat{\alpha}$ on x captures part of the effect of T . Specifically, $\hat{\alpha} > \alpha$ if $\text{Corr}(T, x) > 0$ and $\hat{\alpha} < \alpha$ if $\text{Corr}(T, x) < 0$.
4. Similarly, if $\gamma \neq 0$ and $\text{Corr}(T, x) \neq 0$, the estimated coefficient $\hat{\beta}$ in the utility equation absorbs part of the effect of the omitted permanent trait.

D Estimation: additional explanations

This appendix provides more details about the estimation strategy summarized in Section 4 of the paper. I repeat the individual likelihood here for ease of reading.

Define $d_i = (d_{it})_t$ ($t \in \{1, 2, 3\}$) to be the vector of choices of student i , $\text{fail}_i = (\text{fail}_{it})_t$ to be the vector of retention/graduation events, and $o_i = (o_{it})_t$ to be the vector of evaluations observed by i .⁹ The student makes $T_d \in \{1, 2, 3\}$ decisions, receives $T_f \in \{1, 2, 3\}$ notifications of retention/graduation, and observes signals in $T_d + 1$ periods. More specifically, they receive T_d internal evaluations and $T_r \geq 1$ external evaluations. For instance, consider a student who is retained in level I, stays in school one more period and then drops out. They make two choices, i.e. $d_i = (1, 0)$. At $t = 1$ they are retained, and at $t = 2$ they are promoted to level II, such that $\text{fail}_i = (1, 0)$. They are evaluated externally at the end of primary school, and internally in level I at $t = 1$ and $t = 2$, i.e. $o_i = (r_{0,i}, g_{I,i1}, g_{I,i2})$.

⁸ $\text{Var}(E(C)|x) - \text{Var}(E^*(C)|x) = \left(1 - \left(\frac{\sigma}{\sigma + \rho} \right)^2 \right) \tau > 0$.

⁹ o_{it} is a vector containing one evaluation at $t = 0$ and in level I, and up to two evaluations in level II. Recall that I use $g_{\tau,it}$ ($r_{\tau,it}$) for internal (external) evaluation at time t in level τ . I denote g_{it} (r_{it}) as the evaluation at time t , while abstracting from the level, and $g_{\tau,i}$ ($r_{\tau,i}$) as the last evaluation in level τ , while abstracting from time.

Recall that ϕ is the pdf of ability $h \sim \mathcal{N}(0, \sigma)$. Omitting the dependence on observable characteristics for ease of notation, the individual likelihood is given by:

$$\begin{aligned} L_i = & \int L(r_{i0}|h)L(\text{fail}_{i1}|h, r_{i0})L(g_{i1}|h, r_{i0})L(d_{i1}|h, r_{i0}, g_{i1}) \cdots L(d_{iT_d}|h, o_i, d_{i1}, \dots, d_{iT_d-1}) \phi(h)dh = \\ & \left(L(d_{i1}|r_{i0}, g_{i1}) \cdots L(d_{iT_d}|o_i, d_{i1}, \dots, d_{iT_d-1}) \right) \times \left(L(\text{fail}_{i1}|r_{i0}) \cdots L(\text{fail}_{iT_f}|o_i, d_{i1}, \dots, d_{iT_d-1}) \right) \times \\ & \times \int L(o_{iT_d}|h, d_{i1}, \dots, r_{i0}, \dots) \cdots L(r_{i0}|h) \phi(h)dh \end{aligned} \quad (\text{A-13})$$

Therefore, the log-likelihood is separable in three parts that can be estimated sequentially:

$$\log L_i = \log L_{i,d} + \log L_{i,\text{fail}} + \log L_{i,o} \quad (\text{A-14})$$

Maximizing the likelihood $\log L_{i,o}$ would be computationally costly because of the integration of h . Following James (2011) and Arcidiacono et al. (2025) I use an Expectation-Maximization (EM) algorithm to overcome this issue. I summarize the implemented approach in Subsection D.1. Once coefficients of $\log L_{i,o}$ have been estimated, they can be used in the other components. In particular beliefs on cognitive skills can be computed for each student at any point in time and used as regressors. Then one has to estimate a logit model for the probabilities of failure, and a model of dynamic choices with logistic errors. More details are provided in subsections D.2 and D.3.

D.1 Cognitive skills

Let ζ be the vector of all the parameters that enter the grades equations (including variances of the idiosyncratic errors). Recall that $\phi(h)$ is the density function of the unobserved ability, which follow normal distribution $\mathcal{N}(0, \sigma)$, and $\psi_i(h) = \psi(h|o_i; \zeta, \sigma)$ is the conditional density of h for individual i given her evaluations and the parameters.

For each individual i the likelihood $L_{i,o} = L(o_i; \zeta, \sigma)$ is the joint density function of the evaluations. To estimate the parameters (ζ, σ) one has to find

$$\arg \max_{\zeta, \sigma} \sum_i \log L(o_i; \zeta, \sigma) = \arg \max_{\zeta, \sigma} \sum_i \log \int L(o_i; \zeta, \sigma | h) \phi(h) dh \quad (\text{A-15})$$

The main point behind this application of the EM algorithm is that if $\hat{\zeta}$ is a maximizer for (A-15), then it also solves

$$\arg \max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma | h) \psi_i(h) dh \quad (\text{A-16})$$

Therefore for a given value of σ , $\hat{\zeta}$ can be retrieved using (A-16) rather than (A-15). $(\hat{\zeta}, \hat{\sigma})$ can be estimated using an iterative algorithm: at each iteration k , first (E-step) posterior distributions $\psi_i^k(h)$ are computed for all individuals using previous iteration estimates ζ^{k-1} . Then (M-step) estimates of parameters ζ^k are computed as solution of (A-16).

Appendix (H) provides a more detailed theoretical motivation. Next paragraphs describe the estimation procedure.

D.1.1 E-step

At step k , posterior distribution $\psi_i^k(h)$ is computed for every students using all the observed evaluations and the parameters $(\zeta^{k-1}, \sigma^{k-1})$ estimated in the previous iteration. Let $E_i^k(h)$ be the individual posterior belief for h at iteration k , and ω_i^k the posterior variance. Moreover, at the end of E-step, the estimate for the population variance is updated; this new σ^k is used at the beginning of next step $k + 1$. The updating formula for σ^k is retrieved using the law of total variance:¹⁰

$$\sigma = E(\omega_i + E_i(h)E_i(h)'), \quad (\text{A-17})$$

The sample equivalent at step k is computed as

$$\hat{\sigma}^k = \frac{1}{N} \left(\omega_i^k + E_i^k(h)^2 \right) \quad (\text{A-18})$$

D.1.2 M-step

Given the individual posterior density functions ψ_i^k obtained in the E-step,

$$\arg \max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma^{k-1} | h) \psi_i^k(h) dh \quad (\text{A-19})$$

can be solved to obtain an updated estimate ζ^k for the parameters in the evaluations equations. More specifically:

$$\sum_i \int \log L(o_i; \zeta, \sigma^{k-1} | h) \psi_i^k(h) dh = \quad (\text{A-20})$$

$$\begin{aligned} &= \sum_i \left(\sum_t \int \log L(r_{it}; \zeta, \sigma^{k-1} | h) \psi_i^k(h) dh + \sum_t \int \log L(g_{it}; \zeta, \sigma^{k-1} | h) \psi_i^k(h) dh \right) = \\ &= \sum_i \int \log L(r_{0,i}; \zeta, \sigma^{k-1} | h) \psi_i^k(h) dh + \sum_{it} \int \log L(g_{I,it}; \zeta, \sigma^{k-1} | h) \psi_i^k(h) dh + \\ &+ \sum_{it} \int \log L(g_{II,it}; \zeta, \sigma^{k-1} | h) \psi_i^k(h) dh + \sum_{it} \int \log L(r_{II,it}; \zeta, \sigma^{k-1} | h) \psi_i^k(h) dh \end{aligned} \quad (\text{A-21})$$

where each sum is taken only on the relevant individuals and times. Given that the errors of the evaluations are normally distributed and the posterior distribution $\psi_i^k(h)$ is known, the above

¹⁰Let f be the vector of signals. Give that $E(h) = 0$ and applying law of iterated expectations:

$$\text{Var}(h) = E(h \cdot h') - E(h) \cdot E(h') = E(h \cdot h') = E(E(h \cdot h' | f)) = E(\text{Var}(h \cdot h' | f) + E(h | f) \cdot E(h | f))$$

expression can be derived as follow:¹¹

$$\begin{aligned}
& - \sum_i \log L_i = \sum_i \frac{1}{2} \log(2\pi\rho_{r_0}) + \frac{1}{2\rho_{r_0}} \left(\omega_i^k + \left(r_{0,i} - (\mathbf{E}_i^k(h) + z'_{i0}\beta_0) \right)^2 \right) + \\
& + \sum_{it} \frac{1}{2} \log(2\pi\rho_{g_I}) + \frac{1}{2\rho_{g_I}} \left(\omega_i^k + \left(g_{I,it} - (\nu_I + \mu_I \mathbf{E}_i^k(h) + z'_{it}(\beta_I + \gamma) + \kappa_I I_{0,i}) \right)^2 \right) + \\
& + \sum_{it} \frac{1}{2} \log(2\pi\rho_{r_{II}}) + \frac{1}{2\rho_{r_{II}}} \left(\omega_i^k + \left(r_{II,it} - (o_{II} + \lambda_{II} \mathbf{E}_i^k(h) + z'_{it}\beta_{II} + \kappa_{II} I_{I,i}) \right)^2 \right) + \\
& + \sum_{it} \frac{1}{2} \log(2\pi\rho_{g_{II}}) + \frac{1}{2\rho_{g_{II}}} \left(\omega_i^k + \left(g_{II,it} - (\nu_{II} + \mu \lambda_{II} \mathbf{E}_i^k(h) + \mu(z'_{it}\beta_{II} + \kappa_{II} I_{I,i}) + z'_{it}\gamma) \right)^2 \right)
\end{aligned} \tag{A-22}$$

The total likelihood in (A-22) is the sum of four parts, one for each type of evaluations. Some students may contribute twice to the likelihood of an evaluation if they are retained, or they may not have some of them (if they drop out or do not take the final external evaluation).

If all the regressors were time invariant (i.e. if $I_{\tau,i}$ were not in the equations), the joint estimation of (A-22) would be completely equivalent to separately estimate the coefficients for r_0 , then for g_I , and finally jointly estimates coefficients of r_{II} and g_{II} . Conversely the presence of time varying regressors makes all the four parts interdependent because past regressors have an indirect effect on evaluations in the following periods. Therefore a joint estimation is the most efficient. In practice, I found the following two-step MLE to be a good compromise between efficiency and speediness of the computations:

1. *Parameters for external evaluation at the end of primary school.*

- Perform OLS regressions of $r_{0,i} - h_i$ over z_{i0} . This provides us with updated estimates β_0^k , and allows the computation of $I_{i0}^k = s_{i0}\beta_{s,0}^k$, which is used in next step.
- Update variances $\rho_{r,0}^k$, using the sample equivalent of $E_\epsilon(E_h(\epsilon_{r_{0,i}}|r_{0,i}))$:

$$\text{Var}(\epsilon_{r_{0,i}}) = E(\epsilon_{r_{0,i}}^2) = E(E((r_{0,i} - h_i - z'_{i0}\beta_0|r_{0,i})^2)) = \tag{A-23}$$

$$= E \left(\int (r_{0,i} - h - z'_{i0}\beta_0)^2 \psi_i(h) dh \right) = \tag{A-24}$$

$$= E [\text{Var}_i(h) + ((r_{0,i} - E_i(h) - z'_{i0}\beta_0)^2)] \tag{A-25}$$

¹¹It is easy to see how to derive (A-22) from (A-21). For instance the contribution of the first region-wide test is given by:

$$\begin{aligned}
& \int \log L(r_{0,i}; \zeta, \sigma^{k-1}|\eta) \psi_i^k(h) dh = \int \log \left(\frac{1}{\sqrt{2\pi\rho_{r_0}}} \exp \left(-\frac{(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2}{2\rho_{r_0}} \right) \right) \psi_i^k(h) dh = \\
& = \int \left(-\frac{1}{2} \log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}} (r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2 \right) \psi_i^k(h) dh = \\
& = -\frac{1}{2} \log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}} E_i^k((r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2) = \\
& = -\frac{1}{2} \log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}} \left(\text{Var}_i^k(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0}) + (E_i^k(r_{0,i} - h - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0}))^2 \right) = \\
& = -\frac{1}{2} \log(2\pi\rho_{r_0}) - \frac{1}{2\rho_{r_0}} \left(\omega_i^k + (r_{0,i} - E_i^k(h) - x'_i\beta_{x,0} - p'_{i,0}\beta_{p,0})^2 \right)
\end{aligned}$$

which, in each iteration k , can be estimated from the sample as

$$\rho_{r,0}^k = \frac{\sum_i (\omega_i^k + (r_{0,i} - E_i^k(h) - z'_{i0}\beta_0^k)^2)}{N} \quad (\text{A-26})$$

2. *Parameters for the other evaluations.* Maximize the joint likelihood of g_I, g_{II}, r_{II} using $I_{i,0}^k$ as a regressor.

D.2 Retention and graduation probabilities

I estimate the parameters of the logit model described in Section 3.3 of the paper. Regressors include beliefs on cognitive skills and school leniency effects, which are estimated using results from the previous stage. More specifically, I use the estimated parameters $\hat{\zeta}$ and individual posterior distributions of ability $\hat{\psi}_i$ to compute beliefs about cognitive skills at each point in time. Moreover, the estimated school effects $\hat{\mathcal{J}}$ are used as a measure of school leniency.

The estimated coefficients can then be used to compute the probability of repeating level I at time $t = 1$ and the probability of graduation in the following time periods. The individual probability of failure in level II enters the student's maximization problem, while the probability of retention in level I does not, given that it happens before any decision has to be taken. However the estimation of the latter is necessary for simulations and counterfactual analyses.

D.3 Dynamic choices

We use the parameters of evaluations and probabilities to estimate the last piece of the model: the likelihood of the students' choices. It is important to recall that students use their *beliefs* on ability $E_{i,t}(h)$ when they take a decision, not their true ability h_i ; thus, when they compute their expected utility they anticipate that they will receive new signals and modify their beliefs. Therefore the computation of their expected utility for a given choice at time t requires to integrate over all the signals that they may receive from $t + 1$ on. Their distribution is a multivariate normal, obtained through the usual bayesian updating. Let $\mathcal{N}(\hat{h}_{it}, \omega_{it})$ be the (estimated) posterior distribution of h_i at t . Then $s(r_{\tau, it'})$, with $t' > t$, has posterior distribution $\mathcal{N}(\hat{h}_{it}, \omega_{it} + \lambda_{\tau}^{-2} \rho_{r_{\tau}})$ and similarly $s(g_{\tau, it'})$ has posterior distribution $\mathcal{N}(\hat{h}_{it}, \omega_{it} + \mu_{\tau}^{-2} \rho_{g_{\tau}})$. Moreover, the posterior covariance of two signals is ω_{it} .

From now on, I will use $\hat{\psi}_{it}(\mathbf{s})$ for the joint density function at t of a vector $\mathbf{s}$ of future signals. The updated belief $\hat{h}_{it}$ is a linear combination of prior belief $\hat{h}_{it-1}$, and contemporaneous signals $\mathbf{s}_t$; in other words there exists a vector of coefficients $\mathbf{c}_t$ such that $\hat{h}_{it} = (\hat{h}_{it-1}, \mathbf{s}_t') \mathbf{c}_t$. The elements of $\mathbf{c}_t$ are functions of the elements of the covariance matrix and therefore are known to the agent. I will use this notation in the rest of this section to simplify the formulas.

By assumption, error terms ε_{it} are standard logistic, and uncorrelated with regressors and over time. It is well known that, under these assumptions, the value of u_{it} just before observing the random shock to preferences ε_{it} (but knowing everything else) is

$$\begin{aligned} E_{\varepsilon}(u_{it} | v_{it}^A) &= \log(\exp(v_{it}^A) + 1) = \\ &= \log(\exp(\phi_{A,r} E_{i,t}(C_{II}) + y'_{it} \theta_A) + 1) \end{aligned} \quad (\text{A-27})$$

Recall that $E_{i,t}(C_{II}) = E_{i,t}(h) + z'_{i,t}\beta_{II} + k_{II}I_{t-1}$. Given that in level II a student receives two signals, $\mathbf{s}_i = (s_{g,it}, s_{r,it})$, using the notation just introduced:

$$E_{i,t}(h) = (\hat{h}_{it-1}, \mathbf{s}'_t)\mathbf{c}_t = c_{0,t}\hat{h}_{it-1} + c_{1,t}s_{r,t} + c_{2,t}s_{g,t} \quad (\text{A-28})$$

Therefore, the ex-ante value in the previous period $t-1$ is

$$\begin{aligned} E_{i,t-1}(u_{it}|\text{grad}_{it} = 1) &= \int \log(\exp(v_{i,A}) + 1) \cdot \hat{\psi}_{it-1}(s_{g,t}, s_{r,t}) d\mathbf{s}_t = \\ &= \int \log \left[\exp(\phi_{A,r}(k_{II}I_{t-1} + z'_{it}\beta_{II}) + y'_{it}\theta_A) \cdot \exp(\phi_{A,r}c_{0,t}\hat{h}_{it-1}) \cdot \right. \\ &\quad \left. \exp(\phi_{A,r}(c_{1,t}s_{r,t} + c_{2,t}s_{g,t})) + 1 \right] \cdot \hat{\psi}_{it-1}(s_{g,t}, s_{r,t}) d\mathbf{s}_t \end{aligned} \quad (\text{A-29})$$

Moreover, the individual in period $t-1$ can compute $\widehat{\text{Pr}}(\text{grad}_{it} = 1)$ (the probability of graduating next period) using the estimated parameters for the probability of graduation and retention. This gives us a closed formula for v_{i2}^M :

$$v_{i2}^M(\hat{h}_{i2}, z_{i2}) + \varepsilon_{i2} = U_{i2}^M + \delta \widehat{\text{Pr}}(\text{grad}_{i3} = 1) \int \log(\exp(v_{i3}^A) + 1) \cdot \hat{\psi}_{i2}(s_{g,3}, s_{r,3}) d\mathbf{s}_3 \quad (\text{A-30})$$

Similarly, we are able to compute $\widehat{\text{Pr}}(\text{grad}_{i2} = 1)E_{i,1}(u_{i2}|\text{grad}_{i2} = 1)$. To conclude, we need to derive an expression for $E_{i,1}(u_{i2}|\text{grad}_{i2} = 0)$.

Again, thanks to the fact that errors are logistic, $E_\varepsilon(u_{i2}|v_{i2}^M) = \log(\exp(v_{i2}^M) + 1)$ and $E_{i,1}(u_{i2}|\text{grad}_{i2} = 0) = \int \log(\exp(v_{i2}^M) + 1) \cdot \hat{\psi}_{i,1}(\mathbf{s}_1) d\mathbf{s}_1$. Finally, we can compute values for the first period:

$$v_{i,1}^M|(\text{fail} = 1) + \varepsilon_{i,1} = U_{i,1}^M + \delta E_{i,1}(u_{i2}|\text{grad}_{i2} = 0) \quad (\text{A-31})$$

$$\begin{aligned} v_{i,1}^M|(\text{fail} = 0) + \varepsilon_{i,1} &= U_{i,1}^M + \delta \left(\text{Pr}(\text{grad}_{i2} = 1)E_{i,1}(u_{i2}|\text{grad}_{i2} = 1) + \right. \\ &\quad \left. + (1 - \text{Pr}(\text{grad}_{i2} = 1))E_{i,1}(u_{i2}|\text{grad}_{i2} = 0) \right) \end{aligned} \quad (\text{A-32})$$

It is important to stress that from the point of view of a student in first period, $\hat{h}_{i2} = (\hat{h}_{i1}, \mathbf{s}'_2)\mathbf{c}_2$ is a random variable, and therefore it should be integrated out to compute the expectation. In v_{i2}^M , it appears in the flow utility, in the continuation value from graduation, and in the probability of getting the diploma.

Finally, the likelihood of the individual choices can be easily retrieved computing probabilities with the usual formula for binary choices with logistically distributed preference shifters:

$$p_{it}(d_{it} = 1|d_{it-1} = 1) = \frac{\exp(v_{it})}{1 + \exp(v_{it})} \quad (\text{A-33})$$

I maximize the total loglikelihood to estimate the parameters, following Rust (1987).¹²

¹²The integration over the signals is the most computationally costly part of the maximum likelihood estimation, it is performed using Gauss-Hermit quadrature.

D.4 Missing external evaluations

About 6% of the students in the sample who attain last grade did not undertake the external evaluation; this can happen if students are absent from school the day of the test. This possibility entails a small complication for my model: most students receive two signals in second level, but some only observe internal evaluations; they will therefore update their posterior beliefs differently. Moreover, when students (and the econometrician) compute expected utility they should take in account that with probability p they will observe two signals in last period, while with probability $1 - p$ they will observe only one signals. In practice I calibrate $\hat{p}$ using the sample, and I allow for the two different scenarios in the computation of expected utility.

E Results: additional explanations

E.1 Cognitive skills and evaluations

Table A-10: Variance of unobserved ability

	$\mu_l^2 \hat{\sigma}$	$\text{Var}(C_l)$	%	$\text{Var}(E_l(h))$	$\text{Var}(E_l(C_l))$
$l = 0$	0.280 (0.013)	0.572 (0.019)	49.0 (1.5)	0.112 (0.010)	0.406 (0.017)
$l = \text{I}$	0.576 (0.018)	1.202 (0.068)	47.9 (2.4)	0.443 (0.018)	1.069 (0.084)
$l = \text{II}$	0.678 (0.023)	1.164 (0.039)	58.2 (1.6)	0.545 (0.021)	0.962 (0.036)

Panel A: The first column contains the variance of cognitive skills due to unobserved ability (by row: before starting middle school, in level I, and in level II). The second column contains the total variance of cognitive skills, and the third column the share of total variance due to unobserved ability, i.e. $\mu_l^2 \hat{\sigma} / \text{Var}(C_l)$. Similarly, the fourth column contains the variance of beliefs about unobserved ability, and the fifth column the variance of beliefs about cognitive skills. Bootstrap standard errors in parentheses.

Time	Signals received from 0 to t	Posterior Variance
$t = 0$	r_0	0.1685 (0.0040)
$t = 1$	$r_0, g_{\text{I},1}$	0.0650 (0.0040)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{I},2}$	0.0402 (0.0029)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{II},2}$	0.0392 (0.0028)
$t = 2$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}$	0.0321 (0.0019)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{I},2}, g_{\text{II},3}$	0.0286 (0.0022)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{I},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0246 (0.0016)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, g_{\text{II},3}$	0.0281 (0.0021)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}, g_{\text{II},3}$	0.0243 (0.0016)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0243 (0.0016)
$t = 3$	$r_0, g_{\text{I},1}, g_{\text{II},2}, r_{\text{II},2}, g_{\text{II},3}, r_{\text{II},3}$	0.0213 (0.0013)

Panel B: Posterior variance for each set of signals at a given time. Bootstrap standard errors in parentheses.

The first two columns of Table A-10 (Panel A) show, for each school level, the estimated variance of the individual's (unknown) ability h and the estimated total variance of cognitive skills. Given the assumption that h is uncorrelated with the regressors, it is possible to decompose it into variance due to observable characteristics, and variance due to ability. The third column of the table shows the share of total variance due to ability. This is the most interesting statistic because, as explained in Section 3 of the paper, the variance of cognitive skills in a given level does not have any direct interpretation. In all levels, about half of the variance in cognitive skills is due to the

variance in ability h . More specifically, the variance is about 49% before starting middle school and about 58% at the end of middle school.¹³

Panel B presents the posterior variance in each period for every possible set of signals. Evaluations are informative, and indeed the posterior variance is about 0.17 at time 0, it reduces to 0.06 at time $t = 1$, and to less than 0.04 at $t = 2$. It is about 0.02 for retained students who remain in middle school up to $t = 3$. Posterior variance is only slightly larger for students who did not receive an external evaluation in level II. In sum, when students make their first choice at time $t = 1$, they generally have relatively accurate beliefs about their ability and therefore about their cognitive skills. This suggests that students' choices in the later stages would likely remain unchanged even if they were to directly observe their exact skill level rather than relying on their beliefs. However, uncertainty may still influence decisions in earlier periods, when beliefs are less precise.

Table A-11 presents the estimated parameters governing evaluations, and in turn cognitive skills, before starting middle school (β_0) and in level I and II (β_I and β_{II} , respectively). The last column of the table presents the parameters that capture differences between internal and external evaluations (γ). Since the evaluations have been standardized, the coefficients of the individual characteristics can be interpreted as standard deviation changes.

The results indicate that parental background is a fundamental determinant of cognitive skills. Having better educated parents is associated with large improvements in performance, with similar effect for internal and external evaluations. For instance, in level II having a mother with tertiary rather than primary education increases performance by more than 0.6 s.d., while having a highly educated father increases performance by more than 0.3 s.d. Thus, a student with two highly educated parents is expected to score almost 1 s.d. higher on internal and external evaluations than an identical classmate whose parents have only a basic education. Students whose parents completed high school (i.e. they have an “average” education level) are in an intermediate position: everything else equal they perform about 0.5 s.d. better than students with low-educated parents.

Immigrant students perform somewhat worse than natives. In contrast, there is no difference between cognitive skill accumulation of boys and girls. However, girls score higher on internal evaluations (by almost 0.4 s.d.). Being younger on entering primary education is a disadvantage, although the gap diminishes over time. Indeed, being born at the beginning of January rather than at the end of December is associated with an increase of about 0.25 s.d. in skills at the beginning of middle school and with an increase of 0.14 s.d. at the end.¹⁴ Starting middle school with a one-year delay is associated with lower performance (by up to -0.7 s.d. at the end of middle school). The quality of the student's neighborhood has a small positive effect on performance: an increase of 1 s.d. in neighborhood quality is associated with an improvement of at most 0.08 s.d.¹⁵

The set of covariates include a dummy for students not attending the school that was their first

¹³For the sake of comparison, the fourth and fifth columns of the table present the variance of beliefs about ability and about total cognitive skills in each level. In levels I and II, the variance of beliefs is similar to the previous estimates, although slightly smaller. This is consistent with the finding in the next panel, i.e. that students have accurate beliefs, and therefore the distribution of beliefs is similar to the distribution of actual ability.

¹⁴Calsamiglia and Loviglio (2020) document the disadvantage of being younger at school entry in Catalonia, an effect that persists throughout compulsory education.

¹⁵The indexes of neighborhood quality and peer quality are described in Section 2.2 of the paper. I also estimated the model with separate variables for a neighborhood's income level and education level, and the results are very similar. Therefore, I opted for the most parsimonious specification. Furthermore, I estimated the model with a vector of variables for peer characteristics, results are aligned although less precisely estimated.

Table A-11: Estimates of evaluations parameters.

	Skills			γ
	β_0	β_I	β_{II}	
Female	0.034 (0.026)	-0.025 (0.126)	0.100 (0.099)	0.396 (0.047)
Immigrant	-0.256 (0.031)	-0.324 (0.183)	-0.296 (0.136)	0.111 (0.064)
Mother education average	0.241 (0.032)	0.345 (0.131)	0.364 (0.133)	-0.098 (0.057)
Mother education high	0.470 (0.033)	0.686 (0.170)	0.661 (0.141)	-0.102 (0.052)
Father education average	0.231 (0.033)	0.213 (0.124)	0.258 (0.118)	-0.047 (0.045)
Father education high	0.368 (0.038)	0.346 (0.117)	0.375 (0.121)	-0.041 (0.060)
Day of birth	0.247 (0.050)	0.144 (0.064)	0.138 (0.056)	-0.010 (0.047)
Retained in primary school	-0.551 (0.052)	-0.850 (0.091)	-0.738 (0.103)	0.520 (0.070)
Neighborhood SES	0.089 (0.018)	0.056 (0.031)	0.081 (0.027)	0.021 (0.024)
School not first choice	-0.004 (0.046)	0.051 (0.075)	0.040 (0.068)	0.019 (0.056)
Repeat level		0.174 (0.281)	0.557 (0.165)	-0.189 (0.258)
Peer quality		0.288 (0.069)	0.158 (0.046)	-0.270 (0.049)
Share female		0.041 (0.024)	0.005 (0.016)	-0.046 (0.020)
School input ($\mathcal{A}$ or $\mathcal{J}$): p75 - p25		0.367 (0.087)	0.364 (0.091)	0.408 (0.070)
School input: p80 - p20		0.444 (0.099)	0.441 (0.100)	0.530 (0.075)
School input: st. dev.		0.275 (0.058)	0.273 (0.060)	0.298 (0.039)
Female X school input		-0.035 (0.054)	-0.051 (0.039)	-0.002 (0.041)
Immigrant X school input		0.005 (0.084)	-0.002 (0.060)	-0.025 (0.049)
Mother edu. avg. X school input		-0.031 (0.050)	-0.044 (0.053)	-0.030 (0.033)
Mother edu. high X school input		-0.022 (0.068)	-0.002 (0.059)	-0.009 (0.039)
Father edu. avg X school input		0.030 (0.052)	0.002 (0.049)	0.023 (0.034)
Father edu. high X school input		0.045 (0.052)	0.029 (0.055)	0.012 (0.045)
Previous time-varying regressors		-0.014 (0.051)	0.211 (0.317)	
Unobserved ability	1	1.434 (0.040)	1.555 (0.046)	
$\rho_{g\tau}$		0.217 (0.010)	0.240 (0.012)	
$\rho_{r\tau}$	0.423 (0.014)		0.280 (0.011)	

Note. The estimation includes cohort dummies effects, two dummy variables that take value one if information on mother or father is missing, and a vector of dummies for primary schools attended. For each peer variable a cubic polynomial is used; the table reports the effect on the average student in the sample of having peers at the mean rather than 1 s.d. below the mean. Moreover, school effects are estimated: the table reports the interquantile range, the difference between the 80 and the 20 percentiles, and the standard deviations (computed weighting the school effect by size of the school). The estimation includes interaction between school effects and the dummies for gender, nationality, parental education. The table reports the change for a given characteristics of an increase of 1 standard deviations in the school effect.

Bootstrap standard errors in parentheses.

choice. Coefficients are always insignificant and negligible in size. This provides some evidence that students with differing preferences for schools do not have a different development of their cognitive skills, everything else equal.

Repeating a level for a second time has a positive and sizable effect on cognitive skills. In particular, repeating level II is associated with a large improvement of about 0.5 s.d.

Peer quality has a positive effect on cognitive skills, while the proportion of females does not appear to have any relevant effect. In order to capture any non-linear peer effect, I use a polynomial of degree 3 for each peer regressor. To facilitate the interpretation of the results, Table A-11 shows the effect associated with having peers at the mean versus 1 s.d. below the mean, which is associated with an improvement of almost 0.3 s.d. in level I. Figure A-13 in Appendix A.2 of the paper plots peer effects along the distribution of the peer variables. As can be seen, better peers always increases evaluations in level I, although the effect is more pronounced in the tails of the distribution. The pattern is similar in level II, except that the effect is smaller in magnitude, though still significant. On the other hand, the positive effect is offset in internal evaluations. This is in line with the finding in Calsamiglia and Loviglio (2019) that in Catalonia teachers’ evaluations are deflated in the presence of better-quality peers, i.e. teachers are “grading on a curve” to some extent. The proportion of female peers has no effect on skills at any point in the distribution.

With respect to the estimated school inputs $\mathcal{A}$, Table A-11 presents the difference between the 75th percentile and the 25th percentile (interquantile range), between the 80th and 20th percentile, and the standard deviation. $\mathcal{A}$ varies considerably across schools. For instance, the interquantile range is almost 0.4 s.d. in both level I and II, which is comparable to previous results in the literature. For example, replacing the lowest-ranked Boston public school with an average school would improve performance of the affected students by 0.37 s.d. (Angrist et al., 2017). Similarly, attending an oversubscribed charter school in Boston improves performance by approximately 0.4 s.d. in math and 0.2 s.d. in English (Abdulkadiroğlu et al., 2011).

Table A-11 presents similar statistics for school input $\mathcal{J}$, which captures school grading policies. Its variation across schools is also sizable. Finally, school inputs $\mathcal{J}$ and $\mathcal{A}$ are negatively correlated (correlation of about -0.7). In other words, schools that improve cognitive skills the most tend to have a stricter grading policy for internal evaluations.¹⁶

The empirical specification includes interactions between the vectors of school inputs $\mathcal{A}$ and $\mathcal{J}$, and the dummies for gender, immigrant status, and parental education, which is intended to account for variation in school effects across socioeconomic groups. If, for instance, the school’s effect on cognitive skills is larger for girls than for boys, then this would be captured by the coefficient of the interaction term. To ease the interpretation of the results, Table A-11 reports the change due to an increase of 1 s.d. in the school effect for each individual characteristic. The estimated interaction effects are small and insignificant. Therefore, the effect of school environment on cognitive skills development is about the same for all students enrolled in a given class, regardless of gender, nationality or parental education.

¹⁶This is consistent with the evidence presented in Calsamiglia and Loviglio (2019) that schools in Catalonia with higher average external evaluation scores have stricter grading policies.

E.2 Retention and graduation

Table A-12 presents the estimated parameters governing the logit model for retention.¹⁷ Unsurprisingly, higher cognitive skills decreases the probability of retention. Girls have a lower probability of retention regardless of skill level. Parental education appears to be less relevant, although having highly educated parents (especially a father) is associated with a negative effect.¹⁸

As shown by Figure A-13, having higher quality peers increases the probability of retention at any given level of skills. This is consistent with the negative effect of peer quality on internal evaluations discussed in the previous subsection. Teachers may be more prone to fail a student who lags behind his peers than an identical student who is close to the median of his class. Furthermore, students in more lenient schools (larger $\hat{J}$) are less likely to be retained and the effect is sizable. For instance, increasing school leniency J by 1 s.d. has about the same effect on retention as increasing the school input $\mathcal{A}$ by 1 s.d. (thereby improving skills by almost 0.3 s.d.).¹⁹

E.3 Educational choices

Table A-13 presents the estimates of the flow utility parameters. Beliefs about cognitive skills have a large positive effect on both the choice of not to drop out and the choice to enroll in high school. Girls are less likely to drop out and more likely to enroll in high school after graduation. Students with a highly educated father are more likely to choose to enroll in high school.

Retention during lower secondary education has a negative effect on flow utility from the choice of high school. The interactions of cognitive skills with dummies for repeating level I or level II also have negative coefficients, and therefore the total coefficient of cognitive skills for retained students is smaller than for students who are progressing regularly.²⁰ In other words, retention decreases the weight that individuals attribute to cognitive skills when making their choices. Figure A-9 of the paper helps to illustrate this point by showing that, at any level of skills, retained students have a lower payoff, although the gap is wider at higher levels of cognitive skills. Thus, an increase in skill levels has less of a positive effect on the utility for retained students than for non-retained students.²¹ As discussed in Subsection E.2, retention likelihood is decreasing in cognitive skill level, but it is still high for average or just below average skill level, especially among male students. Retention may in particular discourage these students from pursuing higher education. The effect of retention on the choice to stay in school goes in a similar direction, although the coefficients are generally

¹⁷Table A-12 has the same structure of Table A-11.

¹⁸A student's behavior in class may vary with parental background, and it may affect the retention decision. Moreover, teachers communicate with parents during the school year if their child is at risk of retention, and to some extent they take into account the parents' preferences. This may explain the results.

¹⁹Results are similar when the leniency of the school's retention policy is estimated directly using school dummies. The correlation between the estimated vector of school effects and J is more than 0.8. Given that rules for grading and for retention are most likely designed jointly at the school level, I prefer to estimate only one parameter for each school.

²⁰The coefficient of the interaction with the dummy for repeating level I is significant at 1%, while the other coefficient is slightly smaller and only marginally significant.

²¹It is important to recall that the dummies for retention also capture any difference in the value of the outside option for retained and regular students. If retained students have better outside options (for instance, because being older it is easier for them to find a job), then the negative gap between regular and retained students would be captured by the coefficient of the dummies. Thus, the fact that the lines for retained students lie below the one for regular students does not mean that they like high school less in absolute terms, but rather that they value it less compared to their outside options.

Table A-12: Estimates of retention parameters

	Repeat grade
Belief cognitive skills	-2.104 (0.089)
in II at $t = 2$	-1.106 (0.104)
In II after repeating I	-0.516 (0.200)
Second time in II	-0.349 (0.293)
$\widehat{C} \times$ in II at $t = 2$	-0.384 (0.126)
$\widehat{C} \times$ in II after repeating I	0.498 (0.194)
$\widehat{C} \times$ second time in II	1.162 (0.419)
Female	-0.796 (0.097)
Immigrant	-0.177 (0.139)
Mother education average	0.089 (0.100)
Mother education high	-0.161 (0.138)
Father education average	-0.007 (0.120)
Father education high	-0.262 (0.157)
Day of birth	-0.122 (0.163)
Retained in primary school	-1.198 (0.292)
Neighborhood SES	-0.030 (0.064)
School not first choice	0.182 (0.180)
Peer quality	0.513 (0.124)
Share female	0.032 (0.053)
J	-2.071 (0.183)
School input: p75 - p25	-0.846 (0.142)
School input: p80 - p20	-1.097 (0.141)
School input: st. dev.	0.617 (0.065)

Note. The estimation includes cohort dummies and two dummy variables that take value one if information on mother or father is missing. For each peer regressor a cubic polynomial is used; the table reports the effect on the average student in the sample of having peers at the mean rather than 1 s.d. below the mean.

Bootstrap standard errors in parentheses.

smaller in size and not precisely estimated.

For a given level of cognitive skills, students with below-average peers are more likely to choose to stay in middle school than those with above average peers. This may be related to ranking concerns: students at the margin of dropping out may have a further reason to leave school if they dislike being at the bottom of the class.²² Figure A-13 shows that the effect is non-linear: it flattens for values of peer quality above the average. No relevant peer effect is detected for the choice to enroll in high school.

Total school inputs $\mathcal{T}_M$, which affect the choice to stay in middle school, and total school inputs $\mathcal{T}_A$, which affect the choice to enroll in high school are sizable. In both cases, school input at the 75th percentile rather than the 25th percentile increases the flow utility by as much as improving cognitive skills by 0.5 s.d. Both $\mathcal{J}$ and the orthogonal inputs $\mathcal{M}_M$ and $\mathcal{M}_A$ explain a large portion of the aggregate school inputs on choices. About 50% of the variance of $\mathcal{T}_M$ is due to $\mathcal{M}_M$, which increases to 75% for $\mathcal{T}_A$ and $\mathcal{M}_A$. The direct effect of school inputs on the flow utility from high school is comparable to the indirect effect through cognitive skills. Indeed, raising either $\mathcal{M}_A$ or $\mathcal{A}$ by 1 s.d. increases the flow utility by about 0.5 s.d.²³

It is important to stress that although school inputs have a linear effect on flow utility, the school environment may differentially affect classmates according to their individual characteristics or cognitive skills. This is due to the non-linearity of the probability functions, such that the effect on the probability of a change in one of the regressors depends on the initial level of the underlying utility.²⁴ Thus, the school environment may differentially affect the educational decisions of students with differing background even if they are in the same class.

²²The findings in Pop-Eleches and Urquiola (2013) corroborate this interpretation. They find that having better peers has a positive effect on achievement, but children who make it into more selective schools realize they are relatively weak and feel marginalized.

²³The effect of increasing $\mathcal{A}$ is the product of the increase in skills (as in Table A-11) and the effect of skills on flow utility (as in Table A-13).

²⁴As explained in Section D.3, given the assumption that shocks to preferences follow a logistic distribution, the probability of making choice d_t at time t is $\Lambda(v_t(z)) = \frac{\exp(v_t(z))}{1 + \exp(v_t(z))}$, where z is the vector of individual and school variables and v is expected utility (i.e. the flow utility at time t before observing the shock to preferences plus future expected). The marginal effect of a change in the regressor z_k is

$$\frac{\partial \Lambda(v_t(z))}{\partial z_k} = \frac{1}{1 + \exp(v_t(z))} \frac{\exp(v_t(z))}{1 + \exp(v_t(z))} \times \frac{\partial v_t(z)}{\partial z_k},$$

where $\frac{\exp(v_t(z))}{(1 + \exp(v_t(z)))^2}$ reach its maximum at $v_t(z) = 0$ (which corresponds to a probability of 0.5), while it goes to 0 for $v_t(z) \rightarrow \pm\infty$. Thus, the marginal effect is greatest when the probability is near 0.5 and smallest when it is near 0 or near 1. Thus, if, for example, $v_t(z)$ is very large, not only it is likely that the student makes the choice, but also that probability will be almost unaffected by small changes in the variables.

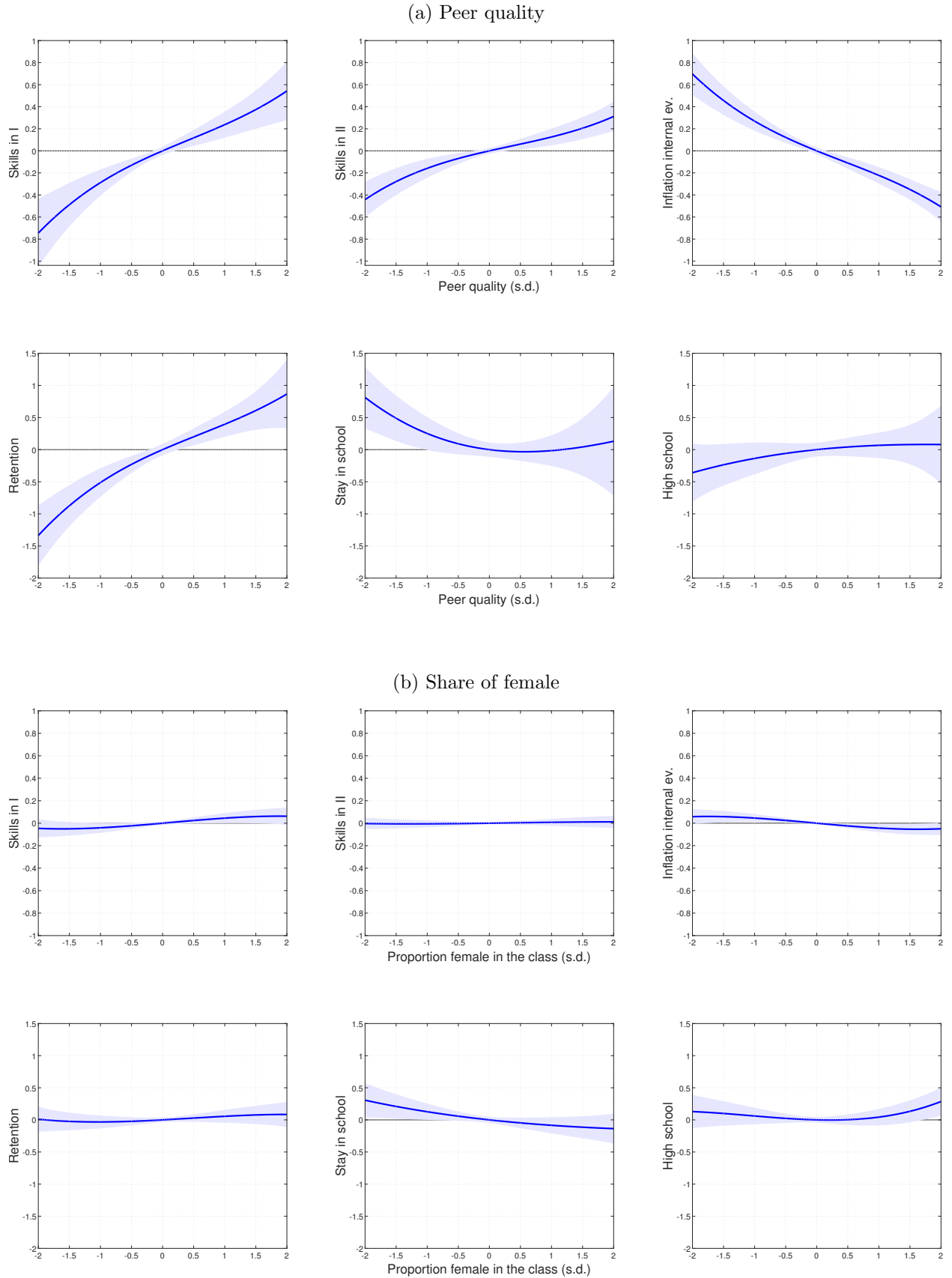
Table A-13: Estimates of choices parameters

	Stay in middle school	Enroll in high school
$\hat{C}$ (belief cognitive skills)	1.310 (0.220)	1.800 (0.122)
$\hat{C} \times$ second time in I	-0.480 (0.217)	
$\hat{C} \times$ in II after repeating I	-0.128 (0.233)	-0.678 (0.204)
$\hat{C} \times$ second time in II	-0.284 (0.392)	-0.555 (0.410)
Female	0.285 (0.096)	0.691 (0.095)
Immigrant	-0.099 (0.109)	0.588 (0.172)
Mother education average	-0.069 (0.108)	-0.023 (0.117)
Mother education high	0.081 (0.166)	-0.119 (0.161)
Father education average	-0.205 (0.111)	0.248 (0.115)
Father education high	-0.372 (0.181)	0.563 (0.169)
Day of birth	-0.331 (0.155)	-0.165 (0.177)
Retained in primary school	0.350 (0.135)	0.987 (0.205)
Neighborhood SES	-0.130 (0.071)	0.102 (0.084)
School not first choice	-0.022 (0.169)	0.063 (0.178)
Second time in I	-1.250 (0.236)	
Second time in II	0.168 (0.479)	-1.730 (0.313)
In II after repeating I	-0.616 (0.415)	-1.773 (0.176)
Peer quality	-0.251 (0.136)	0.137 (0.136)
Share female	-0.127 (0.073)	-0.061 (0.073)
School input ($\mathcal{T}_M$ or $\mathcal{T}_A$): p75 - p25	0.684 (0.140)	0.904 (0.149)
School input: p80 - p20	0.827 (0.149)	0.996 (0.157)
School input: st. dev.	0.495 (0.059)	0.610 (0.077)
$\mathcal{J}$	1.137 (0.177)	1.009 (0.254)
School input $\mathcal{M}_M$ or $\mathcal{M}_A$: p75 - p25	0.383 (0.119)	0.717 (0.140)
School input $\mathcal{M}_M$ or $\mathcal{M}_A$: p80 - p20	0.604 (0.137)	0.933 (0.185)
School input $\mathcal{M}_M$ or $\mathcal{M}_A$: st. dev.	0.357 (0.054)	0.531 (0.073)

Note. The estimation includes cohort dummies and two dummy variables that take value one if information on mother or father is missing. For each peer regressor a cubic polynomial is used; the table reports the effect on the average student in the sample of having peers at the mean rather than 1 s.d. below the mean. School effects are estimated: the table reports the interquantile range, the difference between the 80 and the 20 percentiles, and the standard deviation (computed weighting the school effect by size of the school) of the estimated school effects.

Bootstrap standard errors in parentheses.

Figure A-13: Effects of peer variables on outcomes



Note. The figures plot the effects of peer quality (panel (a)) and share of female (panel (b)). Shaded areas are 95% confidence intervals.

F Simulations: additional analysis and tables

F.1 Aggregate patterns in school contributions

This appendix details the analyses summarized in Subsection 6.1.1 of the paper.

F.1.1 Student-specific school ranking

Using the estimated parameters of the model, I simulate the educational outcomes that every student in the sample would have in each middle school in Barcelona. Specifically, student i 's outcomes in school s are simulated using i 's individual characteristics and test scores at time $t = 0$, along with their estimated unobserved ability. Peer characteristics are set at the average value for each school and level. For each student i and school s , I draw 1000 sets of shocks to evaluations, preferences, and retention events using the estimated distributions, and simulate i 's choices and outcomes. Averaging the results, I then compute i 's expected outcomes in school s . More specifically, I focus on three main outcomes of interest: 1. cognitive skills C_{II} at the end of lower secondary education; 2. the ex ante (at time $t = 0$) probability of graduation; and 3. the ex ante probability of enrolling in academic upper secondary education.²⁵ The first outcome is a measure of performance, while the other two are measures of attainment (for brevity, I will refer to them as “graduation” and “high school enrollment”, respectively). Analogous aggregate measures are computed by averaging the observed outcomes in the data at the school level: 1. average external evaluations at the end of lower secondary education; 2. the proportion of graduates; and 3. the proportion of students who enroll in high school.

Three school rankings are computed for each student based on the three simulated outcomes of interest, and individual rankings are compared to aggregate rankings from observational data. This allows for an assessment of whether aggregate rankings are informative about individual prospects. Furthermore, the correlation between rankings based on different outcomes is analyzed to determine if schools that improve performance also improve attainment.

I first look at rankings separately by outcomes. For each student, I compute the correlation between each individual ranking and its aggregate counterpart from the observational data. Mean correlations are shown in the first line of Table A-14. Moreover, I assess how school rankings for a given outcome vary across students by calculating for each school the standard deviation of the school ranking across students.²⁶ Average results are shown in the second line of the table. On average, the correlation between individual and aggregate rankings by performance is high (0.89). Moreover, individual rankings have a low variance: on average, the standard deviation of the position of a given school across individual rankings is about 1, which corresponds to about 2 percentiles on a 0-100 scale. Conversely, the table shows that the correlation between individual and aggregate rankings is positive but lower for attainment than for performance (0.55 for graduation and 0.59 for high school enrollment, respectively) and there is sizable variation across individual

²⁵Skills at the end of lower secondary education are computed using the occurrences in which the student reaches level II. This occurs in at least some iterations for all students in all schools.

²⁶If a school is ranked the same for all students in the sample, the standard deviation is 0. Conversely, if the school's position follows a discrete uniform distribution (i.e. it can take any position with equal probability) its standard deviation is $\sqrt{(\#schools^2 - 1)/12}$, which is equal to almost 14 when there are 47 schools.

Table A-14: School Rankings

	C_{II}	graduate	enroll in high school
avg rank corr(data, simulation _i)	0.891 (3.7e-04)	0.552 (0.002)	0.594 (0.001)
avg sd(rank _s)	1.060 (0.099)	6.088 (0.293)	4.518 (0.329)
	Corr(C_{II} , grad.)	Corr(C_{II} , high sch.)	Corr(grad., high sch.)
rank corr in the data	0.646	0.778	0.897
avg rank corr in the simulation	0.150 (0.002)	0.233 (0.002)	0.692 (0.002)

Note. The first line of the top panel plots the average correlation between individual rankings simulated through the model and observed data for the three educational outcomes displayed in the columns (skills at the end of lower secondary education, graduation, enrollment in high school). The second line plots the average standard deviation of the school position across individuals. The second panel of the tables plot correlations between rankings based on different outcomes; the first line shows the empirical correlations, the second line the average correlations between individual rankings. Standard error of the mean in parentheses.

rankings. On average, the standard deviation of the school ranking is 6.1 (13 percentiles) for graduation and 4.5 (10 percentiles) for high school enrollment.

Of interest is also the cross-correlation between school rankings according to different outcomes. As discussed in Section 2.3 of the paper, average performance and average attainment at the school level are positively correlated in the data. More specifically the rank correlation between observed performance and graduation rate is 0.65 and that between performance and high school enrollment rate is 0.78. These relatively large correlations in the data may be driven by the characteristics of the students. For instance, schools with a high proportion of students with a high socio-economic background typically show both high test scores and high attainment. Indeed, results in the second panel of Table A-14 show that the correlation between performance and attainment plummets when individual rankings are used. It drops to 0.15 for graduation and to 0.23 for high school enrollment.

The results up to this point suggest that a simple school ranking based on average test scores can provide useful information about a school’s contribution to skill development, and therefore performance – provided that the student does not drop out. On the other hand, it does not convey much information about the student’s attainment prospects. This is because the correlation observed in the raw data at the school level between performance and attainment is driven mainly by student characteristics, rather than a school’s impact on attainment. In particular, it is not informative about how a given school’s environment would contribute to the attainment of students who are not typical of their current student population.

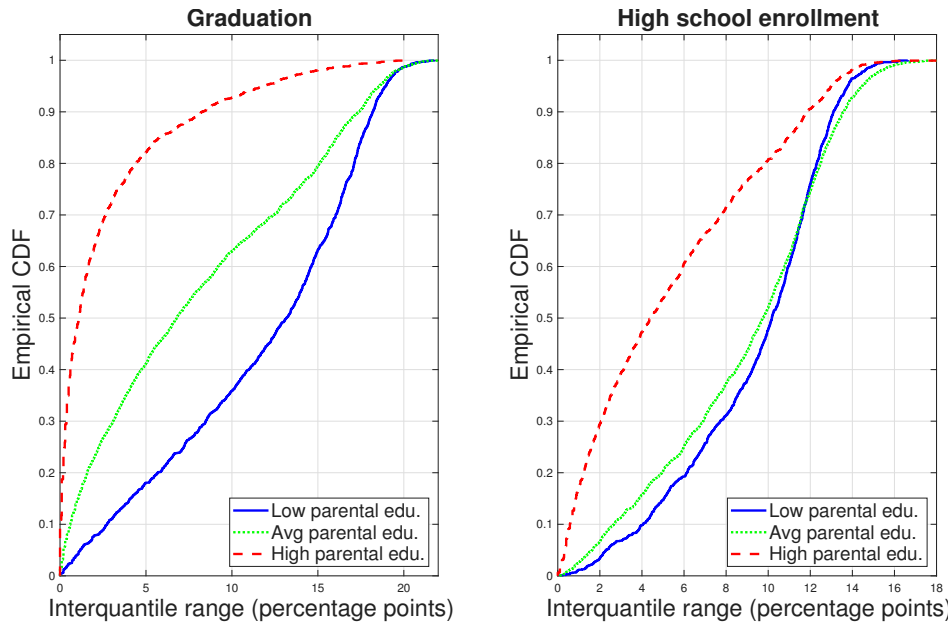
F.1.2 Differences in outcomes by student characteristics

Ignoring attainment is particularly problematic in the case of students whose probability of graduation or high school enrollment vary significantly across schools. While the previous Subsection F.1.1 focused on ordinal rankings, in this subsection I assess the magnitude of a school’s contribution to the educational outcomes of students according to parental education.

More specifically, for each student I measure the variation of simulated outcomes across schools using the interquantile range (IQR), namely the difference between attending the school at the 75th

percentile of the outcome distribution and attending the school at the 25th percentile. I find the interpretation of the IQR to be particularly intuitive, nonetheless I also repeat the analysis using other measures such as standard deviation. Figure A-14 plots the empirical cumulative distribution of the IQR of graduation probability (left panel) and the IQR of high school enrollment probability (right panel) for different levels of parental education. Table A-15 presents the average IQR and standard deviation in the entire sample and by parental education.

Figure A-14: Variation of outcomes across schools. CDF by parental education



Note. The figure plots the empirical CDF of the interquartile range of the graduation probability (left) and high school enrollment probability (right) by parental education (low, average, and high). Statistics are computed using simulated student outcomes in each school. For each school, the graduation (or high school enrollment) probability is first computed using the simulated outcomes. Schools are then ranked according to this probability, and the interquartile range is defined as the difference between the outcome in the school at the 75th percentile and the outcome in the school at the 25th percentile of the distribution.

The results of the simulation make it possible to quantify the total effects of the various channels discussed in Section 5 of the paper. Skill levels at the end of lower secondary education vary considerably across schools (average IQR of about 0.40 s.d.). Given that the production function of skills is almost linear in its inputs, the contribution of the school environment to skill development is about the same for every student enrolled in a given school with a given set of peers. The variation across schools is also sizable for graduation and high school enrollment: the average IQRs are about 7 p.p. and 8 p.p., respectively. However, it is evident from Figure A-14 that there are major differences across groups. In fact, students with lower parental education improve their skills less over the years, therefore they are more likely to repeat a level and drop out. Thus, on average they have a lower probability of graduating and enrolling in high school than students with higher parental education (because of both their skill level and the direct negative effect of retention on educational choices). Given that they are at the margin between staying and leaving, the school environment has a larger effect on their probability of pursuing further education.

The empirical CDF in the graph shows that the median IQR for graduation is about 13 p.p. for students with low-educated parents, while it declines to about 1 p.p. for students with highly educated parents. Moreover, the IQR is more than 9 p.p. for 70% of the students with low-educated parents and 16 p.p. or more for 30% of them. Conversely, it is less than 3 p.p. for 70% of the students with highly educated parents, and less than 0.5 p.p. for 30% of them. The empirical CDF for students with average parental background lies approximately midway between those two groups. The situation is comparable for high school enrollment, though with two main differences. First, the CDF for average parental background is closer to the one for low parental background. Second, the variation across schools is non-negligible for many students with high parental education as well, such that the median IQR value is almost 5 p.p.

Table A-15: Variation of outcomes across schools by parental education

	C_{II}	graduate	enroll in high school
<i>Interquantile range:</i>			
low parental education	0.410 (5.2e-04)	0.117 (1.6e-03)	0.093 (9.8e-04)
average parental education	0.391 (7.2e-04)	0.081 (1.8e-03)	0.089 (1.1e-03)
high parental education	0.417 (5.3e-04)	0.027 (9.0e-04)	0.054 (1.0e-03)
all students	0.405 (3.6e-04)	0.071 (9.0e-04)	0.078 (6.1e-04)
<i>Standard deviation:</i>			
low parental education	0.344 (4.7e-04)	0.077 (1.0e-03)	0.076 (7.5e-04)
average parental education	0.329 (6.3e-04)	0.054 (1.1e-03)	0.071 (8.5e-04)
high parental education	0.352 (5.0e-04)	0.020 (6.0e-04)	0.044 (7.7e-04)
all students	0.341 (3.2e-04)	0.048 (5.7e-04)	0.063 (4.6e-04)

Note. The table plots the average interquantile range (upper panel) and the average standard deviation (bottom panel) of the simulated individual outcomes in each school of the sample. Standard error of the mean in parentheses.

Table A-16 repeats the analysis by gender and immigrant status. The variance of attainment outcomes across schools is somewhat larger for males than females (for example, the average IQR for graduation are 8.2 p.p. for males and 6 p.p. for females). This is consistent with the results in the previous section which showed that, everything else equal, males are more at risk of failing a level and dropping out. Therefore, the contribution of school environment is more likely to make a difference for their choices. Finally, attainment varies more across schools for immigrant students than for natives.

F.2 School contributions by student type

This appendix provides additional details about the simulations discussed in Subsection 6.1.2 of the paper.

Prevalence of type L and type H in the sample. In the simulation *Type L* has low-educated parents, and more specifically both parents attained at most lower secondary education. *Type H* has highly educated parents, and more specifically both parents completed tertiary education. Overall, 18% of students in the sample have two parents with at most lower secondary education. In the case of the others in the group with low-educated parents, one parent has basic education and the

Table A-16: Variation of outcomes across schools by gender and immigrant status

	C_{II}	graduate	enroll in high school
<i>Interquantile range:</i>			
male	0.422 (4.0e-04)	0.081 (1.2e-03)	0.084 (7.9e-04)
female	0.387 (3.1e-04)	0.060 (1.3e-03)	0.071 (9.1e-04)
immigrant	0.401 (8.6e-04)	0.124 (2.1e-03)	0.099 (1.4e-03)
native	0.406 (3.9e-04)	0.062 (9.2e-04)	0.074 (6.6e-04)
<i>Standard deviation:</i>			
male	0.357 (3.3e-04)	0.056 (8.0e-04)	0.068 (6.0e-04)
female	0.324 (3.0e-04)	0.039 (7.9e-04)	0.057 (6.9e-04)
immigrant	0.336 (7.8e-04)	0.080 (1.3e-03)	0.077 (1.0e-03)
native	0.342 (3.5e-04)	0.042 (5.9e-04)	0.060 (5.1e-04)

Note. The table plots the average interquantile range (upper panel) and the average standard deviation (bottom panel) of the simulated individual outcomes in each school of the sample. Standard error of the mean in parentheses.

other education level is unknown. 16% of the students have two parents with tertiary education. In the case of the others with highly educated parents, one parent has tertiary education and the other has upper secondary education or it is unknown.

Individual characteristics used in the simulation. The following are individual characteristics shared by both types: male, native Spanish, born in the middle of the year (July 1st), began lower secondary education at 12 years of age, and attends the first-choice middle school. Primary school effect, cohort effect, and neighborhood quality are set to their mean value. I repeated the analyses with other sets of characteristics, particularly for female and immigrant students. Overall, the results exhibit similar patterns, and the differences go in the expected direction. For instance, females are less likely to be retained, and they more often choose to pursue further education, while immigrants drop out more.

Table with statistics for L_0 and H_0 . Table A-17 focuses on two students with average ability (referred to as L_0 and H_0) and shows the variation across schools of skill levels in level I and level II, retention, drop out, graduation and high school enrollment (conditional on graduation and otherwise). Figure 6 in the main text shows the main outcomes of interest for each school in the sample.

F.3 Uncertainty and retention for students with low educated parents

I perform an analysis similar to that described in Section 6.1.3 of the paper, focusing on the representative student L_0 , who has low-educated parents and average ability $h = 0$.²⁷ As in the analysis for the overall sample, I compare the outcomes of student L_0 in each school in the sample with two counterfactual scenarios. In the first scenario, ability is perfectly known to the student but not to school personnel; therefore, uncertainty affects retention decisions but not the student's choices. In the second scenario, ability is observed by both students and school personnel, so

²⁷Other characteristics are as described in Appendix F.2.

Table A-17: Variation of outcomes across schools: comparing average ability students with low and high parental background

	Student L_0				Student H_0			
	mean	p80-p20	p75-p25	s.d.	mean	p80-p20	p75-p25	s.d.
graduate	0.788	0.144	0.120	0.082	0.986	0.012	0.009	0.008
enrol in high school	0.408	0.170	0.117	0.101	0.900	0.058	0.044	0.037
$C_{I,1}$	-0.309	0.543	0.442	0.367	0.632	0.543	0.442	0.367
$C_{II grad.}$	-0.285	0.459	0.400	0.331	0.619	0.481	0.394	0.336
dropout at $t = 1$	0.102	0.066	0.060	0.043	0.011	0.010	0.009	0.007
drop at $t = 2 stay$	0.177	0.090	0.081	0.051	0.047	0.038	0.031	0.018
retained at $t=1$	0.280	0.129	0.109	0.083	0.029	0.016	0.013	0.012
retained at $t=2 stay$	0.175	0.097	0.078	0.057	0.013	0.009	0.007	0.006
graduate stay	0.943	0.059	0.049	0.033	0.999	0.001	0.001	0.001
enrol in h.s. grad.	0.515	0.163	0.124	0.101	0.913	0.049	0.039	0.032

Note. Frequencies are constructed using 10000 simulations of the structural model for each type and every school of the sample. Type L_0 has parents with primary education; type H_0 has parents with tertiary education. They both have average ability $h = 0$, and are male, Spanish, born on July 1, they began middle school at 12 years old. Primary school effect, cohort effect, and neighborhood quality are set at their average value. The average peer characteristics in each school are used. The first column of each part of the table contains the mean outcome (weighted by school size), the other columns contain difference between quantiles and the standard deviation.

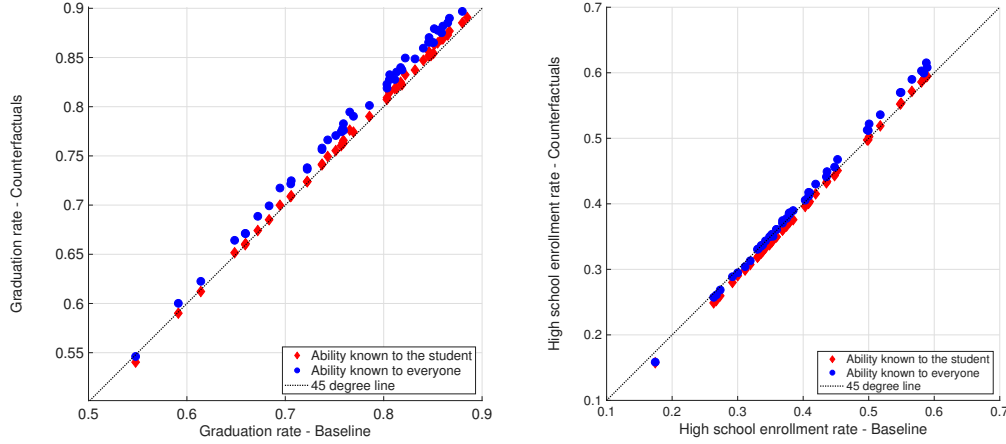
uncertainty is completely eliminated. Figure A-15 compares graduation rates (left panel) and high school enrollment rates (right panel) across schools in the baseline and counterfactual scenarios. In almost all schools, eliminating uncertainty increases the graduation rate, although the gains are modest, typically around 1–2 percentage points. Changes in the probability of enrolling in high school are small. They tend to be slightly negative in schools with lower baseline enrollment rates and slightly positive in schools with higher enrollment rates. When ability is known only to students, outcomes are generally closer to the baseline scenario.

The interplay between retention and uncertainty is further explored in an additional simulation in which peer variables and school inputs are set at their average values among students with low-educated parents. This exercise illustrates more explicitly how uncertainty interacts with retention and beliefs to shape the educational trajectory of the representative student L_0 . Table A-18 reports simulated educational outcomes, beliefs, and true cognitive skills at each point in time by retention status. At times $t = 1$ and $t = 2$, retained students tend to hold beliefs about their ability that are slightly lower than their true ability, although the gap is modest. As a result, even abstracting from any direct disutility of retention, these students are somewhat more likely to drop out. By contrast, among students who remain enrolled and graduate at time $t = 3$, both perceived and actual ability are higher than those observed at time $t = 2$. Consequently, the large differences in high school enrollment between students graduating at different times are primarily driven by differences in preferences rather than by beliefs.²⁸

To isolate the role of uncertainty, I repeat the analysis under the counterfactual scenario in which ability h is perfectly observed (Table A-19). Eliminating uncertainty increases L_0 's probability of graduating by about 2 percentage points and his probability of enrolling in high school by

²⁸I replicate the analysis for student H_0 . The qualitative patterns are similar, although retention is much less frequent for this type (about 3% in the first level and 1% in the second level).

Figure A-15: Outcomes with and without uncertainty - type L_0 student



Note. The left figure contrasts the graduation rate in each school of the sample simulated using the baseline model (x -axis) and a counterfactual model without uncertainty (y -axis). Frequencies are constructed using 100000 simulations for the representative student L_0 in each school in the sample. The right figure has the same structure, but the outcome plotted is the enrollment rate in high school. Student L_0 has parents with primary education and ability at the average value of $h = 0$.

about 0.5 percentage points, with the largest gains occurring when the student is retained. These improvements are primarily driven by a lower probability of dropping out following retention at times $t = 1$ and $t = 2$. Overall, uncertainty modestly penalizes student L_0 , especially when he is retained, although most differences in choices and attainment continue to be driven by preferences rather than by beliefs.

Finally, it is worth noting that in both scenarios student L_0 is very likely to graduate conditional on remaining enrolled. While his overall probability of not completing lower secondary education is about 23%, only about 4 percentage points are due to failure at time $t = 3$, with the remaining 19 percentage points driven by dropout decisions at times $t = 1$ and $t = 2$.

G Students' allocation to classes

While students' allocation to schools is centrally regulated (as discussed in Section 2.1 of the paper), there aren't strict rules about students' allocation to classes within a given school. Anecdotal evidence suggests that in basic education students are typically assigned to classes randomly, in some cases conditionally on their gender. However, it is not possible to rule out that some principals may follow a different approach. Sorting of students across classes might challenge the identification; in particular, the effect of peer quality on performance or choices might reflect unobserved characteristics of the student rather than a proper peer effect.

Following Ammermueller and Pischke (2009), I perform a series of Pearson χ^2 tests to test if students' characteristics and the class students are assigned to are statistically independent. Table A-20 presents the p-values for the main individual characteristics used in the model: gender, immigrant status, mother education and father education. Tests are performed both by cohort and pooling all the observations together. For those variables the null hypothesis of independence is

Table A-18: Educational outcomes by retention status: student L_0 in his typical school environment. Baseline simulation

(a) *Ex ante* probabilities

	overall	promoted at t=1	retained at t=1
graduate	0.773	0.884	0.526
enrol in high school	0.396	0.508	0.145
retained at t=1	0.309	0.000	1.000
retained at t=2	0.100	0.145	0.000
drop. at t=1 or t=2	0.185	0.104	0.367
stay but fail at t=3	0.041	0.012	0.107

(b) Conditional probabilities

	overall	promoted	ret. at t=1	ret. at t=2
dropout at t=1	0.112	0.084	0.174	-
drop. at t=2 stay & do not grad.	0.206	-	0.234	0.137
graduate stay	0.949	0.987	0.831	0.903
enrol in h.s. graduate	0.512	0.604	0.275	0.378

(c) Actual and perceived skills

	overall	promoted	ret. at t=1	ret. at t=2
true $C_{I,1}$	-0.416	-0.416	-0.416	-
perceived $\hat{C}_{I,1}$	-0.418	-0.390	-0.479	-
true $C_{\tau,2}, \tau \in \{I, II\}$	-	-0.428	-0.257	-0.428
perceived $\hat{C}_{\tau,2}, \tau \in \{I, II\}$	-	-0.387	-0.278	-0.508
true C_{II}	-0.365	-0.428	-0.394	0.089
perceived $\hat{C}_{II}$	-0.343	-0.387	-0.393	0.044

Note. Frequencies are computed using 10000 simulations of the estimated structural model for the representative student L_0 . Type L_0 has parents with primary education, average ability $h = 0$, and is male, Spanish, born on July 1, they began middle school at 12 years old. Primary school effect, cohort effect, and neighborhood quality are set at their average value. Peers characteristics and school inputs take the average values among students with low parental background.

Table A-19: Educational outcomes by retention status: student L_0 in his typical school environment. Counterfactual simulation with observed h

(a) <i>Ex ante</i> probabilities				
	overall	promoted at t=1	retained at t=1	
graduate	0.795	0.890	0.557	
enrol in high school	0.401	0.503	0.151	
retained at t=1	0.287	0.000	1.000	
retained at t=2	0.094	0.131	0.000	
drop. at t=1 or t=2	0.168	0.099	0.339	
stay but fail at t=3	0.037	0.010	0.104	

(b) Conditional probabilities				
	overall	promoted	ret. at t=1	ret. at t=2
dropout at t=1	0.103	0.084	0.150	-
drop. at t=2 stay & do not grad.	0.192	-	0.222	0.116
graduate stay	0.955	0.989	0.843	0.913
enrol in h.s. graduate	0.505	0.588	0.270	0.386

(c) Skills				
	overall	promoted	ret. at t=1	ret. at t=2
true $C_{I,1}$	-0.416	-0.416	-0.416	-
true $C_{\tau,2}, \quad \tau \in \{I, II\}$	-	-0.428	-0.257	-0.428
true C_{II}	-0.369	-0.428	-0.394	0.089

Note. As in Table A-18, frequencies are computed using 100000 simulations for the representative student L_0 . However, a counterfactual model with known ability is used. Student L_0 has parents with primary education, average ability $h = 0$, and is male, Spanish, born on July 1, they began middle school at 12 years old. Primary school effect, cohort effect, and neighborhood quality are set at their average value. Peers characteristics and school inputs take the average values among students with low parental background.

never rejected. On the other hand, the null hypothesis is rejected for the last variable of the table: a dummy for students with above average primary school test score.²⁹ More specifically, the null hypothesis is rejected in 4 schools for the cohort 2009 and in 6 schools for the cohort 2010. In only 1 school it is rejected for both cohorts. Overall, results do not allow us to fully rule out that in some cases students are assigned to classes taking into account their past performances, although it does not seem that schools systematically sort students.

I estimate the model restricting the sample to the 36 schools for which the null hypothesis of independence is never rejected. Overall, results are qualitatively and quantitatively very similar. The estimated peer effects on choices and retention are close to the baseline specifications. The estimated peer effect on cognitive skills is similar at the tails of the distribution but flatter around the average. This suggests that, if anything, the main specification might slightly overestimate the positive effect on cognitive skills of an improvement of peer quality when the baseline value is close to the average. The estimated school inputs and the coefficients of individual characteristics are quite close to the baseline ones.

Table A-20: Tests for independence of peer variables and class assignment

	female	immigrant	mother edu.	father edu.	test score
Cohort 2009	0.856	0.546	0.358	0.235	0.000
Cohort 2010	0.991	0.137	0.559	0.355	0.022
All	0.924	0.340	0.458	0.295	0.011

Note. The table reports p-values for Pearson χ^2 tests of independence between the student characteristics and classroom assignment within each school using the individual-level data.

H EM algorithm: theoretical framework

Let ζ be the vector of all the parameters that enter the grades equations (including variances of the errors); recall that σ is the variance of the ability h . The likelihood $L(o_i; \zeta, \sigma)$ is the joint density function of the outcomes. As discussed in previous section

$$\log L(o_i; \zeta, \sigma) = \log \int L(o_i; \zeta, \sigma | h) \phi(h) dh \quad (\text{A-34})$$

$$L(o_i; \zeta, \sigma | h) = L(r_{i,0}; \zeta, \sigma | h) L(g_{i,1}; \zeta, \sigma | h) \dots L(o_{i,T_d}; \zeta, \sigma | h) \quad (\text{A-35})$$

where the likelihood of each evaluation conditional on h is a normal density function. For instance:

$$L(r_{i,0}; \zeta, \sigma | h) = \frac{1}{\sqrt{2\pi\rho_{r_0}}} \exp \left(-\frac{(r_{i,0} - h - z'_{i,0}\beta_0)^2}{2\rho_{r_0}} \right) \quad (\text{A-36})$$

Taking the log of (A-35) would simplify the expression and allow an easy estimation through maximum likelihood. Unfortunately the integral over h prevent us from doing so. The proposed

²⁹To account for the continuous nature of the variable, I also performed Kruskal-Wallis tests in each school using the primary school test score. Results at the school level are extremely similar, but the aggregation in one statistic is less straightforward than with the Pearson χ^2 .

approach aims at overcoming this issue.

The FOC of the sum of individual log-likelihoods are as follow:

$$\frac{\partial}{\partial \zeta} \sum_i \log L(o_i; \zeta, \sigma) = \sum_i \frac{1}{L(o_i; \zeta, \sigma)} \int \frac{\partial L(o_i; \zeta, \sigma|h)}{\partial \zeta} \phi(h) dh = 0 \quad (\text{A-37})$$

$\psi_i(h) = \psi(h|o_i; \zeta, \sigma)$ is the conditional density of h for individual i given her outcomes and the parameters. By definition of conditional density

$$\psi_i(h) = \frac{L(o_i; \zeta, \sigma|h)\phi(h)}{L(o_i; \zeta, \sigma)} \quad (\text{A-38})$$

Now, moving $L(o_i; \zeta, \sigma)$ under the integral and multiplying by $1 = \frac{L(o_i; \zeta, \sigma|h)}{L(o_i; \zeta, \sigma|h)}$, equation (A-37) can be rewritten as

$$\sum_i \int \frac{L(o_i; \zeta, \sigma|h)\phi(h)}{L(o_i; \zeta, \sigma)} \frac{1}{L(o_i; \zeta, \sigma|h)} \frac{\partial L(o_i; \zeta, \sigma|h)}{\partial \zeta} dh = \quad (\text{A-39})$$

$$= \sum_i \int \frac{1}{L(o_i; \zeta, \sigma|h)} \frac{\partial L(o_i; \zeta, \sigma|h)}{\partial \zeta} \psi_i(h) dh = \sum_i \int \frac{\partial}{\partial \zeta} (\log L(o_i; \zeta, \sigma|h)) \psi_i(h) dh = \quad (\text{A-40})$$

$$= \frac{\partial}{\partial \zeta} \left[\sum_i \int \log L(o_i; \zeta, \sigma|h) \psi_i(h) dh \right] = 0 \quad (\text{A-41})$$

Thus if $\hat{\zeta}$ solves equation (A-37) it solves also equation (A-41) and vice-versa. The advantage of the second object is that it allows to work with $\log L(o_i; \zeta, \sigma|h)$ and the individual posterior distributions. In next section I will give an explicit formulation for it.

Parameters can be estimated using an iterative algorithm which is a taylored application of the EM algorithm. In a nutshell, at each iteration k , first (E-step) posterior distributions $\psi_i^k(h)$ are estimated for all individuals using previous iteration estimates ζ^{k-1} . Then (M-step) estimates of parameters ζ^k are computed as solution of

$$\zeta^k = \arg \max_{\zeta} \sum_i \int \log L(o_i; \zeta, \sigma^k|h) \psi_i^k(h) dh \quad (\text{A-42})$$

The general theory ensures convergence of the algorithm.³⁰

³⁰Dempster, Laird, and Rubin (1977)

References

- Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011). Accountability and flexibility in public schools: Evidence from boston’s charters and pilots. *The Quarterly Journal of Economics* 126(2), 699.
- Ammermueller, A. and J. Pischke (2009). Peer effects in European primary schools: Evidence from the progress in International Reading Literacy Study. *Journal of Labor Economics* 27(3), 315–348.
- Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2017). Leveraging lotteries for school value-added: Testing and estimation*. *The Quarterly Journal of Economics* 132(2), 871–919.
- Arcidiacono, P., E. Aucejo, A. Maurel, and T. Ransom (2025). College attrition and the dynamics of information revelation. *Journal of Political Economy* 133(1), 53–110.
- Calsamiglia, C. and A. Loviglio (2019). Grading on a curve: When having good peers is not good. *Economics of Education Review* 73, 101916.
- Calsamiglia, C. and A. Loviglio (2020, March). Maturity and school outcomes in an inflexible system: evidence from catalonia. *SERIEs* 11(1), 1–49.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1), 1–38.
- James, J. (2011, October). Ability matching and occupational choice. Working Paper 11-25.
- Pop-Eleches, C. and M. Urquiola (2013, June). Going to a better school: Effects and behavioral responses. *American Economic Review* 103(4), 1289–1324.
- Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica* 55(5), 999–1033.