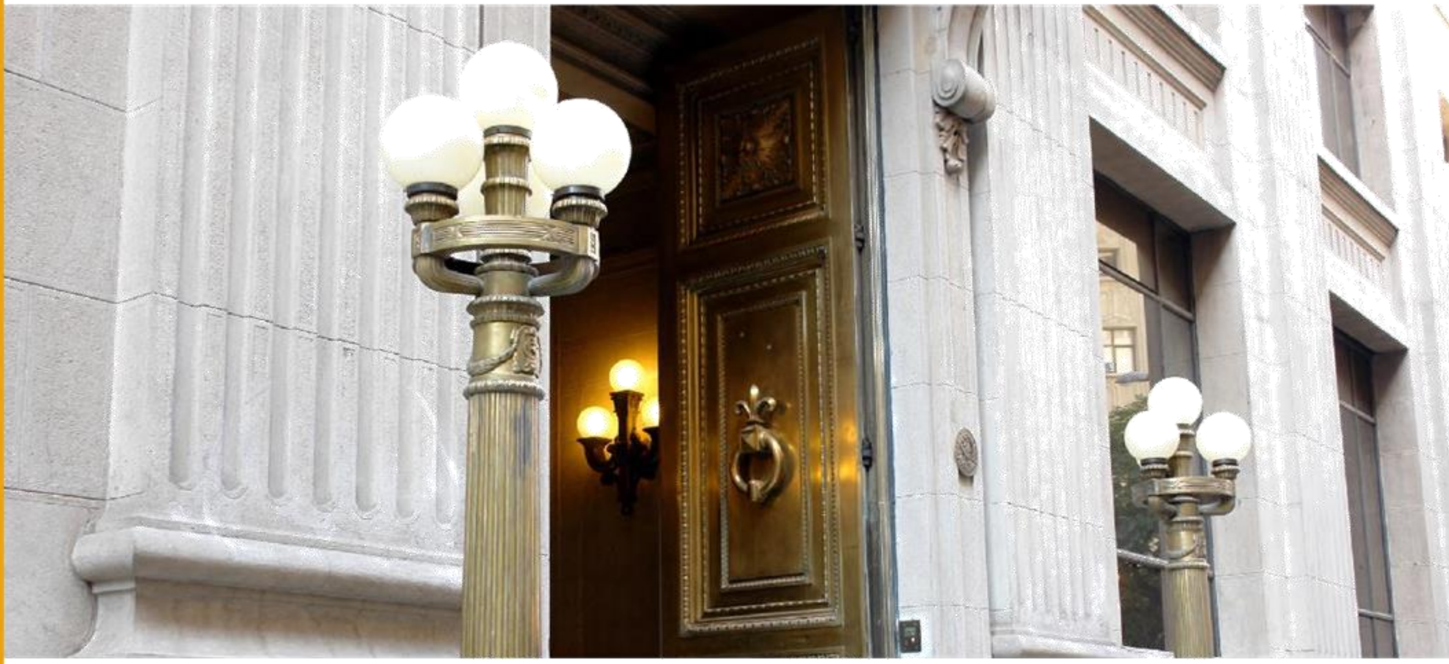
DOCUMENTOS DE TRABAJO

Learning Your Own Ability

Carlos Madeira

N° 919 Agosto 2021



Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180, Santiago, Chile
Teléfono: (56-2) 3882475; Fax: (56-2) 3882231

Documento de Trabajo
N° 919

Working Paper
N° 919

Learning your own ability[*]

Carlos Madeira[†]
Central Bank of Chile

Abstract
Families' human capital investments depend on beliefs about their children's performance. I build a dynamic model of expectation formation to show how agents use both observable and unobservable information to predict their school scores. The model shows parents and students have substantial knowledge of unobservable factors affecting their performance, especially in middle and high school. Families are overconfident towards expecting higher grades and expectation formation differs by race. Families' ability to predict future scores improved substantially during middle school due to several factors: lower bias and variance of the prediction errors, and a better use of past scores as predictors.

Resumen
La inversión de capital humano de las familias depende de creencias al respecto del desempeño académico de los estudiantes. Este trabajo presenta un modelo dinámico de formación de expectativas para estudiar como los agentes utilizan información observable y no observable para predecir sus clasificaciones escolares. El modelo muestra que padres y estudiantes tienen un conocimiento sustancial de factores no observables que impactan al desempeño académico, sobretodo en la escuela media y secundaria. Las familias tienen exceso de confianza en relación a expectativas de calificaciones elevadas y la formación de expectativas difiere por etnia. La calidad de la predicción de los resultados escolares futuros mejoró sustancialmente para las familias durante la escuela media debido a varios factores: menor sesgo y varianza de los errores de predicción, y mejor uso de los resultados escolares pasados como indicadores del desempeño futuro.

[*] Central Bank of Chile, Agustinas 1180, Santiago, Chile. I gratefully acknowledge the comments and suggestions from the editor James Bullard and one anonymous reviewer. I would like to express my enormous debt to Orazio Attanasio, Sergio Urzua, Elie Tamer, Chris Taber, and Chuck Manski for their guidance and comments from seminar participants at University of Essex, Northwestern University, Universitat Autonoma de Barcelona, Universidad de Alicante, University of Exeter, Fundacao Getulio Vargas, the Shanghai Econometric Society World Congress, ITAU, NERA, and the Central Bank of Chile. I thank Romina Oses for research assistance. Financial support from Fundação Calouste Gulbenkian and Fundação para a Ciência e Tecnologia is gratefully acknowledged. Comments are welcome at carlosmadeira2009@u.northwestern.edu. All errors are my own. This research used the The Beginning School Study, 1982-2002 (Log# 01293) dataset (made accessible in 2004, numeric data …les). These data were collected by Karl L. Alexander and Doris R. Entwisle and are available through the Henry A. Murray Research Archive of the Institute for Quantitative SocialScience at Harvard University, Cambridge, Massachusetts (Producer and Distributor).

[†] The views and conclusions presented here are exclusively those of the authors and do not necessarily reflect the position of the Central Bank of Chile or its Board members. Email: cmadeira@bcentral.cl

1 Introduction

Predicting individual future performance is important in decisions with uncertain outcomes, such as starting a firm, paying loans (Madeira 2019), raising children (Jervis 2017), choosing a work career, applying to college or saving for retirement (Delavande and Rohwedder 2011). Expectations are particularly relevant for academic choices, since many human capital decisions are made early in life (Cunha et al. 2005, Delalibera and Ferreira 2019). Economic models of education choice usually assume that agents are able to predict their academic performance and the expected returns of each option. This is a strong assumption, since a student's performance may change substantially when arriving to a new grade or school level. It can be difficult for families to forecast their children's academic achievement for several reasons: educators may favor different teaching methods, schools provide different social environments and curriculum materials may change between years.

This paper uses a publicly-available dataset, the Beginning School Study (BSS), to study how families use available information to form expectations of their academic achievement. The BSS elicited point predictions of academic outcomes for a panel of 825 parents and students from the children's first-grade until their adult years. Previous work with the BSS data shows that families are systematically overoptimistic about their children's school performance, although their predictions improved as the students aged (Madeira 2018). To evaluate how respondents use available information I specify a model of student achievement and agents' expectations, where respondents forecast their future scores based both on observable information (demographic characteristics and past performance) and private information of the agents[1] that is unobservable to the econometrician. Respondents know the cut-off values used by the teachers to assign grades based on the students' achievement. The model is distinct for parents and children, since the respondents within the same family do not necessarily share the same private information (Giustinelli and Manski 2018): it is possible that the parents and children' private information sets have a common component and are therefore correlated, but it is not necessary to specify and estimate this correlation because the dynamic model analyzes parents and children separately. Furthermore,

[1] The private information of the agents can include any sort of relevant factors that were not measured by the BSS survey, such as the school's quality, the teacher's grading standards, the student's study habits, how much grandparents help with the child, neighborhood problems, parents' attention to school matters, permanent and transitory income of the household, or even health issues.

in each period the respondents update both their observable (students' past grades) and private information. The distribution of the respondents' private information can be identified by the econometric model from two sources: 1) the heterogeneity of beliefs among agents with the same observable information, 2) the correlation of beliefs with the actual outcomes. I then compare the respondents' predictions to those of a counterfactual rational agent that has access only to observable information and which therefore knows less than the respondents (which also have private information). The model can then be used to examine several sources of the agents' prediction errors: 1) overconfidence (defined as a positive bias between the expected achievement and the respondents' expectations), 2) inefficient use of the information observed in the student's previous academic scores (defined as the average respondent reacting too little to the information in past grades relative to a rational agent's predictions), or 3) noisy use of private information (defined as the excess variance in respondents' prediction errors relative to those of a rational agent).

Using this model of expectation formation I find that in elementary school parents and students presented both a large bias and variance for their predictions, indicating that respondents are both overconfident and use noisy private information. The results suggest the overconfidence bias is smaller for families with higher education levels, older parents, and parents of girls. Respondents in elementary school react to differences in past academic achievement, but update their expectations more slowly than a rational agent would. Black families are more optimistic than average, but had prediction errors with a similar variance as white families. Over the years, however, black families made similar gains in their ability to predict performance as their white counterparts.

The model also shows that the role of parents and students' private information about unobserved factors affecting school performance increased substantially during middle school and high school. Students in particular have a much higher degree of private information than their parents during high school. In middle school families' bias and prediction variance show a strong decrease, in particular for students. In addition, I find that both parents and students made better use of the information available in past marks, updating recent information about school performance more quickly. Respondents' ability to predict future scores improved substantially during middle school and high school. This improvement can be decomposed in terms of several components: lower bias and variance of the prediction errors, and a better use of past scores as predictors. Economic models of human capital decisions, therefore, should take into account higher degrees of private information as students grow older and that the unobserved information known by teenage students increases at quicker rates than for their parents. One potential implication for overcoming the information loss of parents in middle and high school is to increase the quality and frequency of the information sent to parents as children age (Bergman 2020, Bergman et al. 2018, Berlinski et al. 2016). Other studies are also taking into account that academic decisions are often a joint choice process made by parents and children (Attanasio and Kaufmann 2014). One potential policy impact of considering parents and children's joint roles in the education process is that incentives can be more efficiently targeted, with more capable parents managing better the resources for their children, while higher performing children can receive direct rewards for their achievement (Berry 2015). Although there is a lack of quality survey data on parents and children's expectations and their roles on the final academic outcomes (Giustinelli and Manski 2018), some studies are dealing with how expectations of parents and children shape their schooling choices (Giustinelli 2016). However, panel data on the joint expectations of parents and children is often still missing in this literature (Giustinelli 2016, Giustinelli and Manski 2018), therefore the BSS dataset and the dynamic learning model estimated for its respondents is a good complement to past studies.

Incorrect beliefs about achievement may lead families to make inefficient investments. For instance, overconfident students may put less effort in school. I find black parents and students are more optimistic about their academic scores even after several years, which may explain why they stay longer as students in school (Rivkin 1995, Lang and Manove 2011), despite having similar returns to education (Lang and Rudd 1986, Lang and Manove 2011).

This work is related to a large body of literature testing rational expectations, models of learning and updating beliefs (Manski 2018, Attanasio et al. 2020, Cornand and Hubert 2020), both for continuous variables (Madeira and Zafar 2015) and qualitative outcomes (Madeira 2018). It is also related to the education process and expectations of disadvantaged families (Alexander et al. 1988, Oyserman 2015). Other work finds that incorrect beliefs about academic performance may explain inefficient education choices, such as college dropout decisions (Stinebrickner and Stinebrickner 2012, 2014). Many empirical studies show that agents tend to overestimate their ability and their estimates do not improve significantly with feedback on the past performance (Hoelzl and Rustichini 2005). Some laboratory experiments studied how agents update their beliefs with new information (Houser et al. 2004). However, lab studies may fail to replicate how agents learn over longer periods or in less standardized environments. Because few datasets follow the same respondents over many years, little is known about how agents adjust their beliefs (Manski 2018, van der Klaauw 2012, Conlon, Pilossoph, Wiswall and Zafar 2018). This work fills some of that gap, since the extended time panel of the BSS dataset allows the researcher to observe how families change their beliefs as they age and learn more information (Conlon, Pilossoph, Wiswall and Zafar 2018).

The paper is organized as follows. Section 2 describes the BSS data. Section 3 describes the structural model of expectation formation. Section 4 presents the estimation results of the structural model and explains the main changes in parents and students' predictions of their academic performance. Finally, Section 5 presents a summary of the main results.

2 Data description

2.1 Sample design of the survey

The Beginning School Study (BSS) consists of 838 children that were randomly selected from the first-grade rosters in 1982 of a set of 20 Baltimore public elementary schools. First, a set of 20 schools stratified by racial attendance was selected: 6 schools predominantly black, 6 schools predominantly white and 8 mixed schools. Afterwards, within each school a random sample of students were selected from each first-grade classroom[2]. Parental permission for participation in the study was obtained from 825 parents. These families make up the initial sample size in 1982. The families in the BSS survey come mainly from a disadvantaged background and are quite close to federal poverty lines. Around 63% of the BSS students participated in the federal program for subsidized school lunches. Most parents had low education, with more than 30% of both white and black parents having less than a complete high school information. Also, less than 35% of the parents went to college and only 10% actually completed a four year college degree. Furthermore, 28% of the mothers in the BSS sample were teenagers at time of the birth of their children[3].

[2] There are no families with multiple children in the BSS dataset. In principle, it could be possible for a family to have two children in first-grade, either because the students could be twins or one sibling could be repeating first-grade. But due to the random-sampling no such pair of same family students were selected.

[3] Unfortunately, the parental questionnaire does not have much information about the parents besides their age, education, and occupation. It would be interesting, for instance, to know what knowledge parents have about the

Table 1.1: Identity of the parents replying to the survey

Parents in the first survey	Fraction of the sample: All sample	Fraction of the sample: For white / black	Fraction of the sample: For parents of boys / girls	Parents always remaining the same respondent: All sample	Parents always remaining the same respondent: For white / black	Parents always remaining the same respondent: For parents of boys / girls
Mother	86.6%	86.9% / 85.1%	85.0% / 88.2%	78.1%	77.6% / 78.6%	76.5% / 79.7%
Father	8.3%	10.2% / 6.8%	8.7% / 8.0%	13.0%	18.4% / 6.5%	16.7% / 9.1%
Other[a)]	5.1%	3.5% / 6.4%	6.3% / 3.9%	23.8%	23.1% / 24.1%	30.8% / 12.5%
Total	100%	45% / 55%	50% / 50%	70.0%	69.6% / 70.2%	68.4% / 71.5%

Statistics are a percentage of the observed sample (829 parents that responded to the BSS).

a) Some survey years do not identify who the "Other relative" is, therefore for some surveys it could a grandparent or step-parent.

Analyzing which parents replied to the survey and how many responded to all the future parental surveys, Table 1.1 shows that most of the parents that answered the first survey were the mothers of the students (87% of the sample). In fact, mothers were the major respondent (above 85%) for students of either race (white or black) or gender (boys and girls). It is not always the same parent responding to the survey. However, the first responding parent remains the same respondent for all the BSS surveys in 70% of the cases. The rate of respondents that always remained the same is high for mothers (78% of the cases), but it is actually low for fathers (13%) and other relatives (24%). It is also noticeable that black fathers reply less to the survey (6.8% vs 10.2% for white families) and that the black fathers are less likely to persist as the same respondent for the future surveys (6.5% vs 18.4% for white families). Perhaps this pattern of higher absence of the black fathers in the BSS survey could be due to high rates of imprisonment and family separation among black families in Baltimore (Entwisle, Alexander and Olson 1997). It is also relevant to note that despite the young age of the BSS mothers, the great majority of the BSS students have siblings (Table 1.2). Around 74% of the BSS students have some siblings, with 51% having older siblings and 40% having younger siblings. Only 25% of the students born when their mother was a teenager have older siblings, but 50% of them already have younger siblings. Children born when their mother

school system. Most of the parental questionnaire is related to the parents' beliefs about their child's performance in school (whether in terms of academic marks, conduct, motivation to learn or relations to peers), their aspirations and expectations for the child as an adult, and some activities done at home (reading, watching TV, going to museums).

was in their thirties and in their forties already have on average 2 and 5 older siblings, respectively.

Table 1.2: Age of the mother at birth of the BSS student and number of siblings

	Fraction of the BSS sample				Average number of siblings$^{a)}$		
Mom's age	All sample	With siblings	With older siblings	With younger siblings	All siblings	Older siblings	Younger siblings
10-20	35.5%	63.1%	24.7%	49.8%	1.0	0.3	0.6
21-30	53.7%	81.1%	65.0%	38.0%	1.5	1.0	0.5
31-40	10.0%	81.3%	76.0%	21.3%	2.3	2.0	0.2
41-50	0.8%	80.0%	80.0%	20.0%	5.0	4.8	0.2
Total	100%	73.7%	51.4%	40.2%	1.4	0.9	0.5

Statistics are a percentage of the observed sample (825 parents).

a) The average number includes the BSS students with zero siblings.

All children attended schools with the same basic curriculum and with teachers on the same salary scale. Students' grades were assigned on a letter-mark basis for an academic year with 4 quarters. Public schools marks consist of letters in the same percentage scale: Excellent (90-100), Good (80-89), Satisfactory (70-79) and Unsatisfactory (0-69). Marks in middle and high school were also reported on a known interval scale with + and - for each letter. Standardized California Achievement Tests (CATs) were administered in all Baltimore public schools in October and May of each year until 1990. The CAT scores were made available not only to parents and children, but also to teachers and school officials. Information on grade years, school marks, CAT scores and subsidized-lunch status of the students were collected directly from the school records. Most students in the BSS have poor performance, with more than 40% of the students having repeated one grade or more before starting high school (Table 1.3). Grade retention in the Baltimore Public School System has a certain degree of subjective choice on the part of the teachers and the school principal (Alexander et al. 2003). In particular, grade retention can be recommended by the teacher if the student is performing with "Unsatisfactory" at either Math or English/Reading at the end of the year. However, it is possible that the teacher decides to pass the student to the next grade if the teacher sees it as the most appropriate option: some teachers can recommend compensatory schooling during the summer as an alternative to grade retention. After the teacher recommends a grade retention, the final decision must be approved by the school's principal and parents can dispute such a decision by appealing to the Board of Education of Baltimore County (parents are not obliged to dispute the decision, since parents may also be persuaded that repeating the grade can be the best educational path for their child). Around 60% of the students had marks of Satisfactory or Unsatisfactory in Math during their elementary and middle school periods (Table 1.4). Therefore a large proportion of the BSS students were at risk of suffering grade retention.

Table 1.3: Nr of past failed grades by academic year

Nr of Failed Grades	1983/84	1984/85	1987/88	1989/90
One grade ahead	0.4%	0.7%	1.1%	0.4%
Never failed a grade	81.3%	72.8%	61.4%	55.2%
Failed one grade	17.6%	23.7%	31.0%	35.5%
Failed two grades or more	0.4%	2.8%	6.5%	8.8%
Missing	7.0%	11.5%	9.6%	19.0%

All statistics are a percentage of the observed sample, except for the missing sample values which are a percentage of the whole population.

Table 1.4 - Distribution of Students marks:

Letter mark (Math, 1982-89)	1982-89 Q1	1982-89 Q2	1982-89 Q3	1982-89 Q4
1 - Unsatisfactory	19.2%	19.2%	18.5%	16.5%
2 - Satisfactory	47.0%	43.2%	42.5%	42.4%
3 - Good	26.8%	28.9%	29.0%	28.8%
4 - Excellent	7.0%	8.7%	10.0%	12.4%
Missing	30.3%	29.9%	29.1%	30.5%

All statistics are a percentage of the observed sample, except for the missing sample values which are a percentage of the whole population.

Throughout the paper my results are based on a missing conditionally at random (MCAR) assumption. The BSS had very low rates of item nonresponse to particular questions (item nonresponse tends to be inferior to 1% in all survey years) and collected retroactive data on students' school scores (Alexander and Entwisle 2004); therefore marks are observed even for years with missing interviews. Madeira (2018) presents a comprehensive analysis of the missing survey data of the BSS, showing that in general the MCAR assumption is reasonable. Parents and students with missing interviews had similar characteristics in terms of parental education and test scores relative to the families that participated in all the survey years (Madeira 2018). Furthermore, linear regressions of parents and students expectations for respondent with or without missing data have similar coefficients for the updating of Math and English/Reading expectations (Madeira 2018), confirming the validity of the MCAR assumption for the BSS sample.

It is also relevant to note that performing all the exercises of this article with just the sample of households that participated in all the panel surveys does not change the results significantly.

2.2 Questions about the expectations of future marks

The time-line of the survey in each academic year started with the parent questionnaire in the summer or at the beginning of the fall quarter. BSS families are often fragile and with absent parents, therefore it is not always necessarily the same parent that replies to the questionnaire: it can be the mom, dad, or even a step-parent. CAT scores were implemented in the fall and spring quarter before the student surveys. Student surveys were then implemented in the fall and spring quarters before the quarterly marks were given. From 1982 to 1994 the survey asked the following expectation questions of students and parents:

Students: What mark do you think you are going to get in (Math / Reading) - Excellent (90-100), Good (80-89), Satisfactory (70-79), Unsatisfactory (0-69)?

Parents: Please guess the marks your child will get in Reading and Mathematics on the first report card this fall: Excellent (90-100), Good (80-89), Satisfactory (70-79), Unsatisfactory (0-69).

The wording of the questions changes slightly from year to year, but remains consistent in asking for the respondents' best guess. In the case of this article, I interpret the respondents' best guess as their subjective median interval. This assumption is credible because the letter marks correspond to ordinal outcomes and therefore the median interval comes out as the best guess as long as the loss function of the respondents attaches larger losses to predictions farther away from the outcome (Madeira 2018). However, it is also possible to study the expectations of parents and children under a mode outcome, such as the non-parametric test for rational expectations suggested by Madeira (2018).

During the whole time period of the survey, school grades in Baltimore were awarded in an Excelent, Good, Satisfactory, and Unsatisfactory category scale (Entwisle, Alexander and Olson 1997), which translated into an equivalent numerical scale of 90-100, 80-89, 70-79, and 0-69. This educational scale system of the Baltimore school system was different from the one applied in the majority of the United States, which is based on A-F letters. However, it was a system that was universally applied in Baltimore during the whole time period of the BSS survey (1982-1994). Families also answered questions about their predictions of the education level and jobs they would have as adults. Parents indicated their age, education, occupation, employment status and weekly hours of work if employed. Other questions included whether they had stress, problems at work or within the family. Teacher surveys were also implemented at the end of the school year, eliciting their best guesses of the students' marks in the following fall quarter and their future educational attainment.

Table 1.5 - Respondents' expectations	**Parents**		**Students**	
Best guess for the Math (Fall) mark	**1982**	**1982-90**	**1982**	**1982-90**
1 - Unsatisfactory	3.2%	2.5%	2.5%	2.6%
2 - Satisfactory	35.6%	30.2%	12.5%	16.8%
3 - Good	48.7%	47.2%	37.4%	44.7%
4 - Excellent	12.5%	20.2%	47.6%	35.9%
Nr of observations	786	3,379	824	3,335

Some work on expectations has been done before with the BSS data. Alexander, Entwisle and Thompson (1988) show that parents' expectations are significant for predicting first grade outcomes, even after conditioning on previous academic results and family background. The authors also show that parental expectations are affected by race and education. However, no previous work studied how expectations were revised over time based on observable and private information signals.

In Table 1.5 I show that around 60% of the parents believed their children would attain a Good or Excellent mark at Math in the next fall quarter, although only 33% of students actually obtained those marks (Table 1.4). Furthermore, a significant share of the parents and students can fail their prediction for the marks by more than one category. Table 1.6 shows that in first-grade (1982), around 11.6% of the parents and 38.6% of the children provided predictions 2 or more categories above the mark finally obtained. For the period of 1982-1990, around 16.4% of the parents and

29.2% of the children provided predictions 2 or more categories above the mark finally obtained. However, the number of parents or students providing predictions that are 2 or more categories below the marks obtained was lower than 2%.

Table 1.6 - Difference between respondents'	**Parents**		**Students**	
predictions and the mark obtained (Math, Fall)	**1982**	**1982-90**	**1982**	**1982-90**
-3			0.3%	0.1%
-2	1.1%	0.7%	1.5%	0.6%
-1	10.9%	8.5%	5.9%	5.8%
0	39.1%	35.2%	21.7%	26.8%
1	37.2%	39.1%	32.2%	37.5%
2	11.3%	14.9%	31.7%	24.2%
3	0.3%	1.5%	6.9%	5.0%
Number of observations	723	2,852	752	2,962

Note: positive values correspond to optimism/overconfidence error.

Madeira (2018) presents a set of nonparametric tests that rejected the rationality of both parents and students' expectations, even after controlling for a wide range of assumptions about the agents' information sets, although their predictions improved as the children aged. Both parents and students were found to have optimistic expectations, with guesses statistically being higher than the obtained marks. There is a strong persistence of families with correct predictions over time, but the BSS data does show that respondents with incorrect guesses have a significant probability of providing accurate predictions in future years, which is evidence that respondents take their predictions seriously (Madeira 2018).

3 A joint model of student achievement and expectations

3.1 A simple model of achievement and expectations

Families can depart from rationality in several ways: 1) overconfidence[4] (defined as a positive bias between the expected achievement and the respondents' expectations), 2) inefficient use of

[4] In this article, overconfidence and overoptimism or excessive optimism are mentioned as the same concept.

the observable information available (defined as the average respondent reacting differently to the information in past grades relative to what a rational agent would[5]), and 3) agents give predictions that have too much randomness (defined as the excess variance in respondents' prediction errors relative to those of a rational agent). Here I specify a model of student achievement and agents' expectation formation that measures the sources of respondents' forecast errors.

Suppose that at the end of each period t teachers have a measurement of student achievement[6], $s_{i,t}$ in the continuous scale of $[0-100]$ and teachers assign a mark, $Y_{i,t}=j\in\{1,...,J\}$ to students who fall in the interval $s_{i,t}\in(v_{j-1},v_j)$. In this model the student achievement measured by teachers to assign the letter marks is given by an increasing monotone transformation $y^*_{i,t}=f(s_{i,t})$:

1) $y^*_{i,t}=f(s_{i,t})=m(x_i,A_{i,t-1})+\varepsilon_{i,t}=m(z_{i,t})+\varepsilon_{i,t}$,

which can be expressed as the sum of a predictable component $m(x_i,A_{i,t-1})$ based on known time-invariant information x_i (such as mother's age and education) and a vector of past achievement measures $A_{i,t-1}$ (which can include the number of repeated grade levels, plus some of the past marks, such as the last mark $Y_{i,t-1}$, the average marks over the past academic life of the student $\frac{1}{t-1}\sum_{h=1}^{t-1}Y_{i,t-h}$ or the average marks in the previous academic year) plus an idiosyncratic term $\varepsilon_{i,t}$ observed by the teachers[7] but unknown to the econometrician (but that can be partially known by the respondents in terms of their private information). Let z denote the vector of observable information by the econometrician: $z_{i,t}=\{x_i,A_{i,t-1}\}$[8]. It is assumed that $\varepsilon_{i,t}$ can be

[5] In the case of the academic expectations of the application in this article, the respondents (parents or children) are reacting too little to the content of past grades. However, for other applications it could be possible for agents to react too much. For instance, perhaps for expectations of crime it is possible that agents react too much to recent events and then expect too much criminal activity in the future (Manski 2018).

[6] In the case of this article the definition of student achievement corresponds to the measurement obtained by the teachers for the purpose of assigning the academic marks. It is possible that students have other achievements and knowledge that are not observed by the teacher, but the model refers specifically to the achievement measured in school by the teacher.

[7] The idiosyncratic term $\varepsilon_{i,t}$ that is observed by teachers can include any unobservable that could impact the student's mark, such as their participation in class or the correctness of their answers during tests. Teachers use their measurement to define the student's achievement, therefore the concept can be understood as the teacher's private information or an unobservable shock to the student's achievement measured by the teacher.

[8] The vector of observable information is presumed to include variables that could matter for both achievement and the respondents' expectations. However, it does not necessarily imply that these variables do matter for either

correlated with all the other unobservable terms, $\varepsilon_{i,h} \forall h$. This happens through both a permanent random-effect (α_i, that may be understood as a permanent private information that is unobserved to the econometrician, for instance, knowledge about the parents and home environment of the student) plus a time-varying term ($v_{i,t}$, that may be understood as time varying private information such as the child's motivation or relationship with peers). Furthermore, the time-varying term $v_{i,t}$ can be decomposed as a function of private information terms that affected the student's achievement only in a specific period (that is, $v_{i,t} = f(u_{i,t}, u_{i,t-1}, ..., u_{i,1})$). I assume that $\varepsilon_{i,t}$ and $\varepsilon_{i,t-h}$ are independent conditional on the permanent and period specific private information terms (that is, $\varepsilon_{i,t} \perp \varepsilon_{i,h} \mid \alpha_i, u_{i,h}, u_{i,h-1} ..., u_{i,1}$ for any $t > h$). This implies that $y^*_{i,t}$ and $y^*_{i,t-h}$ are independent conditional on the vector $z_{i,t}, \alpha_i, u_{i,h}, u_{i,h-1} ..., u_{i,1}$ for any $t > h$.

The probability of student i receiving a mark below j is:

2) $\Pr(Y_{i,t} \leq j \mid z_{i,t}) = \Pr(y^*_{i,t} = m(z_{i,t}) + \varepsilon_{i,t} \leq f(v_j) \mid z_{i,t}) = F_{\varepsilon|z}(f(v_j) - m(z_{i,t}))$,

where I assume $F_{\varepsilon|z}(.)$ is a distribution known by the econometrician up to a parameter vector[9].

It is not easy to establish a link between the expected interval of the agent and his expectation of the latent variable. For example, if agents express their mode interval, then their mode interval does not necessarily contain the mode of the latent variable. However, a set of interval forecasts allows to evaluate the direction of the prediction losses, therefore a median or $\alpha-$quantile interval is a plausible interpretation for the agents' interval expectations. Note that the agents' categorical prediction, $P_{i,t} = \arg\min_p(\Pr(Y_{i,t} \leq p) \geq \alpha)$, is also the interval that contains their subjective $\alpha-$quantile for the latent variable, $y^*_{i,t}$, therefore the α-quantile($y^*_{i,t}$) belongs to the interval $[f(v_{P-1}), f(v_P)]$. Assuming agents use an absolute loss criterion with $\alpha = 0.5$ to form discrete predictions, $P_{i,t}$, then $P_{i,t}$ corresponds to the interval that contains their subjective median for the continuous outcome. This result is a consequence of the property of invariance of quantiles in relation to monotone transformations. The median interval prediction is an option that is easier to interpret if one

achievement or for both respondents (parents and children). For instance, the econometrician could include the number of books at home. Perhaps this variable has an impact on parents' expectations and yet it could have no impact on student achievement or on the children's expectation. In such a case, the variable would have a positive coefficient for one respondent type (the parents) and a zero coefficient for achievement and children's expectations.

[9]Note that the teachers do not need to know $F_{\varepsilon|z}(.)$. Teachers are required to know only the variable $y^*_{i,t} = f(s_{i,t})$. It is possible that the teachers ignore how $\varepsilon_{i,t}$ is distributed among the other agents.

assumes that respondents penalize more the predictions that are further away from the outcome, which is reasonable for ordinal data such as academic marks[10].

Now I specify the agents' predictions process (which can denote in a general way either parents or children):

3) $p^*_{i,t} = mp(x_i, A_{i,t-1}) + pi_{i,t} = mp(z_{i,t}) + pi_{i,t}$,

as the sum of a systematic component, $mp(x_i, A_{i,t-1})$, and a private information factor, $pi_{i,t}$, known to the agent but not to the econometrician. $mp(.)$ denotes the mean prediction made by each family based on observable information and $pi_{i,t}$ denotes the private information possessed by each agent and not observed by the econometrician. Note that $p^*_{i,t}$ can differ from the measurement of achievement made by teachers $y^*_{i,t}$. Also, the function of observables for the respondents' predictions $mp(z_{i,t})$ can differ from the teachers' $m(z_{i,t})$, because the respondents may attribute different coefficients to the same observable variables (for instance, parents may believe that if the child is repeating the grade, then his marks should improve, while the teachers may view it differently). Different agents such as parents and students may share some private information, but not necessarily all (which means that the private information of parents and children is not necessarily the same). Furthermore, $pi_{i,t}$ may be correlated with $\varepsilon_{i,t}$, although it does not happen necessarily that the respondents know all the factors affecting the achievement measured by teachers, which implies that the correlation of both terms is less than one. In the same way, it is also possible that the respondents' private information is correlated with the private information of other respondents (such as the child's parent or the other students in the same school), but this common component in the private information of parents and students is not modelled or estimated, since it is not necessary for the decomposition of the respondents' prediction errors implemented in this article. An example of $pi_{i,t}$ could be information that parents and students know about their homework or how much the teacher likes the student[11]. All families in Baltimore school system

[10] However, the mode interval could be an option for respondents that only care about predicting the right outcome and that care in the same way for all the incorrect predictions. Another possibility is that agents could provide a different interval quantile besides the median or that there is heterogeneity, with some agents providing the median, while others provide another subjective quantile or the mode. Such type of heterogeneity in the prediction rules of the agents is not considered in this article due to its complexity for the analysis.

[11] Note that the information of the agents does not have to be perfect - the parents may believe from their

know the numerical intervals assigned to each grade (Alexander and Entwisle 2004), therefore I assume the cutoffs for each grade, v_j, are known by the agents[12]. This implies that the probability of the agent i giving a discrete prediction $P_{i,t}$ below value j is:

4) $\Pr(P_{i,t} \leq j \mid z) = \Pr(p^*_{i,t} \leq f(v_j) \mid z) = F_{pi|z}(f(v_j) - mp(z))$,

where I assume that $F_{pi|z}(.)$ is a distribution known by the econometrician up to a parameter vector[13]. The case of the mode interval is harder for parametric identification of the model, because one can only write a set of inequalities for the probability of the mode interval: $\Pr(P_{i,t} = j \mid z) = F_{pi|z}(f(v_j) - mp(z)) - F_{pi|z}(f(v_{j-1}) - mp(z)) > \Pr(P_{i,t} = h \mid z) \forall h \neq j$, with $f(v_0) = -\infty$ and $f(v_1) = 0$. Such a set of inequalities would give a partial identification interval for the parameters (Bugni 2010), but the interval for the parameters and the confidence intervals are easier to obtain and interpret if it is just a single parameter (such as testing rational expectations or not, which is equivalent to zero bias, as in Madeira 2018). For the case of a model with multiple parameters then estimation takes much longer to converge due to the lack of derivatives and the resulting intervals are much harder to interpret (Bugni 2010).

The process of joint predictions and student achievement can then be summarized as:

observations that the student is not doing any homework, but the student could in fact be doing homework when the parent is not watching. Also, it does not necessarily imply that students know less about their effort than teachers do. It just means that students know less about the unobservable factors that impact the achievement measured by the teachers' marks. The student may think that a certain amount of effort will be good for the tests, but then the tests end up measuring the class materials that the student did not include in his self study tasks at home.

[12] This assumption can be easily relaxed by estimating different cutoff parameters for the parents and students, but it eases the interpretation of the results. However, imposing this assumption does not change the final results much. Note that the assumption that the cutoffs are known is reasonable, since such cutoffs were universally applied by all schools in Baltimore. It is possible that individual teachers have some degree of arbitrariness in evaluating the students' achievement, which is included in the unobservable teacher's information set $\varepsilon_{i,t}$, but the school system is responsible for defining the cutoffs, not the teachers. One could take into account that some teachers could decide the marks of individual students based on the marks of other students. Unfortunately, due to the small sample size of the BSS there are very few students being evaluated by the same teacher, therefore such approach is not easy to implement. Accounting for teacher specific factors could show important aspects of learning, but it cannot be easily implemented in the BSS, since it is a panel of children/parents and not a panel of teachers.

[13] Note that the respondents do not need to know $F_{pi|z}$. Respondents are required to know only the variable $p^*_{i,t}$. It is possible that the parents and children ignore how $pi_{i,t}$ is distributed among the other agents.

5) $\Pr(Y_{i,t}=l, P_{i,t}=j \mid z) = \int_{f(v_{l-1})}^{f(v_l)} \int_{f(v_{j-1})}^{f(v_j)} f_{\varepsilon,pi|z}(t_1 - m(z_{i,t}), t_2 - mp(z_{i,t}))\partial t_2 \partial t_1.$

Now I face the problem of choosing a suitable parametric family for $F_{\varepsilon,pi|z}(.)$ that can be identified from expressions 2), 4) and 5). I make the assumption that $(y^*_{i,t}, p^*_{i,t})$ are normal distributed, with means $(m(z_{i,t}), mp(z_{i,t}))$, standard-deviations $(c(z_{i,t}), cp(z_{i,t}))$ and correlation-coefficient $\rho(z_{i,t})$. All the components in the terms $\varepsilon_{i,t}$ and $pi_{i,t}$ are specified to be uncorrelated with z. The difference $mp(z_{i,t}) - m(z_{i,t})$ can be interpreted as respondents' average bias, while $cp(z_{i,t})$ denotes the heterogeneity of agents' private information. Note that $cp(z_{i,t})$ is a parameter for intra group heterogeneity of information and not a parameter of each agent's subjective uncertainty. $cp(z_{i,t})$ represents how much agents in the same group z differ between themselves. The correlation coefficient can be interpreted as a measure of the quality of respondents' private information.

Under the normality assumption, the prediction error of the respondents, $p^*_{i,t} - y^*_{i,t}$, is completely described by its bias and variance components, being distributed as: $N(mp(z) - m(z), c(z)^2 + cp(z)^2 - 2\rho(z)c(z)cp(z))$. Agents can only be rational if $mp(z) - m(z) = 0$ and $cp(z)^2 - 2\rho(z)c(z)cp(z) \leq 0$. The term $mp(z) - m(z)$ examines how respondents' bias varies across demographic groups. Rational agents will have a 0 prediction bias, while overconfident students will have positive values of $mp(z) - m(z)$. For the respondents with private information, the variance of their prediction errors (that is, the variance of $\varepsilon_{i,t} - pi_{i,t}$) is equal to $c(z)^2 + cp(z)^2 - 2\rho(z)c(z)cp(z)$. Note, however, that a rational agent with no private information (that is, with $pi_{i,t} = 0$) would have a prediction error variance of $c(z)^2$. Therefore the term $cp(z)^2 - 2\rho(z)c(z)cp(z)$ measures the excess variance component in families' predictions errors $(\varepsilon_{i,t} - pi_{i,t})$ that a rational agent with $pi_{i,t} = 0$ would not have. Respondents that use their private information efficiently to improve their forecasts will have a lower prediction variance than a rational agent without private information, that is: $cp(z)^2 - 2\rho(z)c(z)cp(z) \leq 0$. If the term $cp(z)^2 - 2\rho(z)c(z)cp(z)$ is positive, then the families are showing an "excess variance" in their forecast errors relative to a rational agent.

The model can also show if families are using observable information in an optimal way. For instance, families could put too much or too little weight on their older scores. The estimated coefficients for the previous school marks $(A_{i,t-1})$ will say whether people are updating observable information too slowly or too quickly by comparing the weights given to recent school performance by the parents and students to the optimal weights a rational agent would have.

3.2 Identification proof

The normal model is identified by imposing location and scale normalizations: A.1) $f(v_0) = -\infty$, $f(v_1) = 0$, $f(v_J) = +\infty$; A.2) $c(x_0) = 1$ for some x_0 with positive density. In this case the probability of a student getting a mark lower than j is given by:

$$6)\ \Pr(Y \leq j \mid z) = \Phi\left(\frac{f(v_j) - m(z)}{c(z)}\right) \Longrightarrow m(z) = f(v_j) - c(z)\Phi^{-1}(\Pr(Y \leq j \mid z)).$$

Using expression 6) for categories j and 1 evaluated at x_0 identifies the cut-off points $f(v_j)$:

$$7)\ f(v_j) = \Phi^{-1}(\Pr(Y \leq j \mid z)) - \Phi^{-1}(\Pr(Y \leq 1 \mid z)).$$

Using expression 6) evaluated at any x for category j and category $k < j$ identifies $c(z)$:

$$8)\ c(z) = \frac{f(v_j) - f(v_k)}{\Phi^{-1}(\Pr(Y \leq j \mid z)) - \Phi^{-1}(\Pr(Y \leq k \mid z))}.$$

$m(z)$ is then identified from expression 6). The identification of $mp(z)$ and $cp(z)$ follows the same argument, except that the scale normalization of the variance is unnecessary since the cutoffs are already identified. The reason why this model allows for heteroscedasticity across groups is that the variance is standardized for only one group of agents and cutoffs are assumed to be the same for all groups. Also, the correlation of private information with the unobserved factors affecting student achievement, $\rho(z)$, is identified from equation 5) (Zellner and Lee 1965).

3.3 Dynamic structure of achievement and expectation updating

Parents and students report their beliefs in several periods and it is reasonable to assume that expectations are correlated over time. Therefore I estimate a parsimonious structure that accounts for the initial private information each agent has and how this information is updated every period.

Student achievement has both permanent, α_i, and transitory components, $v_{i,t}$. The AR(1) structure for $v_{i,t}$ implies that all the previous information shocks count, but their effect decays in a geometric way. In the same way there is a permanent component of the expectations of each agent, α_i^p, and a component, $v_{i,t}^p$, that is updated in each period as the agent gets more information:

9) $\varepsilon_{i,t} = \alpha_i + v_{i,t} = \alpha_i + (u_{i,t} + \lambda(z_{i,t})v_{i,t-1}) = \alpha_i + u_{i,t} + \sum_{h=1}^{t-1} u_{i,t-h} \prod_{l=1}^{h} \lambda(z_{i,t-l+1})$, with $\varepsilon_{i,1} = \alpha_i + v_{i,1} = \alpha_i + u_{i,1}$ and the standardization $\lambda(z_{i,0}) = 1$,

10) $pi_{i,t} = \alpha_i^p + v_{i,t}^p = \alpha_i^p + (u_{i,t}^p + \lambda^p(z_{i,t})v_{i,t-1}^p) = \alpha_i^p + u_{i,t}^p + \sum_{h=1}^{t-1} u_{i,t-h}^p \prod_{l=1}^{h} \lambda^p(z_{i,t-l+1})$, with $pi_{i,1} = \alpha_i^p + v_{i,1}^p = \alpha_i^p + u_{i,1}^p$ and the standardization $\lambda^p(z_{i,0}) = 1$,

where the random effects $(\alpha_i, \alpha_i^p) \sim N((0,0), [\sigma_\alpha(z_i), \sigma_\alpha^p(z_i), \rho_\alpha^p(z_i)])$ and the time-varying unobservables are normally distributed, $(u_{i,t}, u_{i,t}^p) \sim N((0,0), [\sigma_u(z_{i,t}), \sigma_u^p(z_{i,t}), \rho_u^p(z_{i,t})])$. It is also assumed that observations are independent across children (that is, independence across i) and that $\alpha_i \perp u_{i,t} \forall t$, $u_{i,t} \perp u_{i,j}\ \forall j \neq t$, $\alpha_i^p \perp u_{i,t}^p \forall t$, $u_{i,t}^p \perp u_{i,j}^p\ \forall j \neq t$. I apply the standardizations $c_0 = -\infty$, $c_1 = 0$, $c_J = +\infty$ and $Var(\varepsilon_{i,1} \mid z_i^0) = 1$ for some x_i^0 with positive density. The AR coefficients $-1 \leq \rho(z_{i,t}) \leq 1$ and $-1 \leq \rho^p(z_{i,t}) \leq 1$ are restricted to be within the unit-root circle. Note that other distributions are possible. The same model was estimated using a Multivariate Beta distribution. The Multivariate Normal distribution was chosen because the log likelihood values of the estimated model were higher and therefore provided a better fit to the empirical data[14]. The results for the Multivariate Beta distribution are available in an online appendix.

The random-effects and AR(1) structure can be interpreted in a simple way. The Random-effect α_i represents the permanent factors that impact the student achievement measured by the teachers, while α_i^p denotes the private information or permanent knowledge that the respondents (parents[15]

[14] The worse fit of the Multivariate Beta distribution is not surprising, because the panel dimension of the problem is high (2 endogenous variables, marks and predictions, available for several years). The Fairlie-Gumbel-Morgenstern and the Sarmanov family of Multivariate Beta distributions are not flexible enough to allow the correlation to vary between -1 and +1 and severely restricts the correlation among time periods (Ting Lee 1996). The Multivariate Normal is a special distribution that allows the correlation between all dimensions to vary between -1 and +1, which is a flexibility that is not possible in several other multivariate probability distribution families (Ting Lee 1996).

[15] For simplicity, the model does not distinguish whether the parents' predictions come from the same respondents or if parents and other relatives switch among themselves, therefore it is implicitly assumed that in families with multiple respondents for the parental survey there is some sharing of information. As shown in Table 1.1, mothers are the main parental respondent and they tend to persist as the same respondent for all surveys. In practice, due to the small sample size of the BSS and the small fraction of families that switch the parental respondent, it is not feasible to estimate a model where parents decide to switch among themselves as the survey's respondent. In Table A.1.4 of the web appendix, the estimates from an ordered probit for the Math marks and predictions suggest that including the dummy variable "Always the same parent respondent" is a significant predictor for the Math marks in the Fall, but it is not a significant predictor for the Parents and Students predictions. For the English/Reading marks and predictions, a similar ordered probit suggests that "Always the same parent respondent" lowers the expectations

or students) have of such factors. The AR(1) process represents the transitory factors in student achievement ($u_{i,t}$) and the knowledge agents have of these components ($u^p_{i,t}$). This is important because the model allows to estimate whether transitory unobservable factors are more difficult for families to forecast in relation to temporary ones. The private information (or knowledge) of the agents (α^p_i, $u^p_{i,t}$) is imperfect, therefore it is not perfectly correlated with the teachers' measures (α_i, $u_{i,t}$). All the components of $\varepsilon_{i,t}$ and $pi_{i,t}$, whether the permanent effects (α_i,α^p_i) or the transitory components ($u_{i,t}$,$u^p_{i,t}$), are unobservable to the econometrician.

3.4 Estimation

Let $\bar{Y}_{i,t} \equiv (\hat{Y}_{i,1},..,\hat{Y}_{i,t}) = (j_1,...,j_t) = g_t$, $\tilde{Y}_{i,t} = (Y_{i,1},..,Y_{i,t}) = (l_1,...,l_t) = q_t$, $\tilde{z}_{i,t} \equiv (z_{i,1},..,z_{i,t})$, be the vectors of agent i's guesses, academic scores and covariates until time t. The agents' outcome probability and the overall likelihood function are given by:

11) $\Pr(\bar{Y}_{i,T} = g_T, \tilde{Y}_{i,T} = q_T \mid \tilde{z}_{i,T}) = \prod_{t=1}^{T} \Pr(\hat{Y}_{i,t} = j_t, Y_{i,t} = l_t \mid \tilde{z}_{i,t}, \bar{Y}_{i,t-1})$

12) $L = \sum_{i=1}^{N} \log(\Pr(\bar{Y}_{i,T} = g_T, \tilde{Y}_{i,T} = q_T \mid \tilde{z}_i))$

This model is estimated by Simulated Maximum Likelihood, using the GHK procedure (Geweke et al. 1994) with 30 draws. The draws are obtained using the Modified Latin Hypercube Sampling (MLHS) method, which has been shown to strongly outperform other numerical simulators in many applications (Hess et al. 2004). The MLHS method to obtain R multivariate draws basically starts with an equal spaced sequence of values, $\varphi(j) = \dfrac{j-1}{R}$ for $j = 1,...,R$, in each dimension. Then a pseudo-uniform number h is added to the draws of each dimension to get $\tilde{\varphi}(j) = \varphi(j) + \dfrac{h}{R}$ for $j = 1,...,R$. This allows the econometrician to get an equal-spaced coverage of each dimension. To get the pseudo-uniform number h I use scrambled Halton draws, which have been shown to outperform standard uniform numbers (Hess et al. 2004).

---

and parental bias for the Fall marks predictions, although it does not affect the marks and students' predictions. The regressions also show that when the father is the respondent for the first survey, the students show higher marks for both Math and English/Reading, although the children's predictions are lower (for Math in the Fall and English/Reading in the Spring). This result can be due to several factors, since the presence of the father in the first survey can be correlated with an intact and more supportive home environment or also with higher family income.

Finally, for the purposes of estimation I assume the following functional forms:

13) $mj(z_{i,t}) = c_{mj} + z_{i,t}\beta^{mj}$, for the conditional means of the latent variables, with $mj = m, mp$;

14) $\sigma_{\alpha j}(z_{i,1}) = (\frac{\exp(c_{\alpha j} + z_{i,1}\gamma_{\alpha j})}{1+\exp(c_{\alpha j} + z_{i,1}\gamma_{\alpha j})})\exp(c_{\alpha j} + z_{i,1}\gamma_{\alpha j})$ for the standard-deviation of the random-effects, with $\alpha j = \alpha, \alpha^p$ and the standardization $c_\alpha = 0$;

15) $\sigma_{uj}(z_{i,1}) = (1 - (\frac{\exp(c_{\alpha j} + z_{i,1}\gamma_{\alpha j})}{1+\exp(c_{\alpha j} + z_{i,1}\gamma_{\alpha j})})^2)^{1/2}\exp(c_{\alpha j} + z_{i,1}\gamma_{\alpha j})$ and $\sigma_{uj}(z_{i,t}) = \exp(c_{uj} + z_{i,t}\gamma_{uj})$ for $t > 1$, the standard-deviation of the time specific factors, with $uj = u, u^p$ and $\alpha j = \alpha, \alpha^p$;

16) $g(z_{i,t}) = 2(\frac{\exp(c_g + z_{i,t}\theta_g)}{1+\exp(c_g + z_{i,t}\theta_g)}) - 1$ with $g = \lambda, \lambda^p, \rho_\alpha^p, \rho_u^p$ standing for the geometric weight functions of the previous idiosyncractic factors, the correlation between the random-effects and the correlation between the time-specific factors.

I estimate the model with different coefficients for each of the main school levels: First grade, Elementary, Middle, and High school. The coefficient vectors $m(z)$, $c(z)$, $mp(z)$, $cp(z)$, $\rho(z)$ also control for time-dummies for both races. The parameters of the autocorrelation, dispersion and correlation of unobservables use race, gender and maternal education as explanatory variables.

This dynamic model involves a large set of parameters, therefore I focus only on the most representative results of the regressions for Math. The results for English are qualitatively similar and available in a web appendix. Furthermore, the web appendix also includes all the Matlab and Stata codes necessary to replicate all the analysis in this article.

4 Estimation results

4.1 Bias of respondents' expectations across demographic groups

Tables 2.1 to 2.4 show the results of the empirical model for Math[16]. Table 2.1 shows the coefficients affecting the mean achievement and mean expectations of parents and students during elementary school. The mean expectations across demographic groups in first grade, middle and high school

[16] Standard-errors and t-statistics were also estimated with 100 bootstrap sample replications or with the outer product of the gradient, but the results are similar to the inverse Hessian and are not reported.

are qualitatively similar to the elementary school, therefore for brevity those results were left to the web appendix.

Table 2.1: Math marks and expectations in Elementary school - Q1, 83/84-87/88

	Marks (w/ students)	**Parents**	**Students**
	$m(z_{i,t})$	$mp(z_{i,t})^{Parents}$	$mp(z_{i,t})^{Students}$
Mean of latent marks/ expectations:			
Constant	0.42 (0.14)*	1.84 (0.15)*	2.73 (0.3)*
black race	-0.24 (0.06)*	0.11 (0.05)*	-0.01 (0.09)
female	0.09 (0.05)**	-0.03 (0.05)	0.01 (0.09)
Student's month of birth	-0.01 (0.01)	0 (0.01)	0.01 (0.01)
age of mother at birth ($\times$ 10)	0.01 (0.05)	-0.14 (0.05)*	-0.06 (0.08)
mother's years of education	0.05 (0.01)*	0.02 (0.01)	0.01 (0.02)
nr of failed grades	0.25 (0.06)*	-0.04 (0.05)	0.05 (0.08)
CAT score	0.44 (0.08)*	0.24 (0.06)*	-0.12 (0.09)
Average of all past marks	3.78 (0.82)*	2.2 (0.67)*	0.45 (1.18)
Average of last 4 quarters	0.91 (0.6)	2.05 (0.53)*	2.3 (1)*
(controls for year-dummies)	yes	yes	yes
Nr of observations (Elementary school$^{a)}$)	2,359	1,493	1,307
Nr of observations (All school levels)	10,223	3,533	7,177

a) Nr of observations in Elementary School (Fall quarter) after the First Grade (1982).

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

Table 2.2 shows the standard-deviation of the heterogeneity in the temporary unobservable private information for each group and how the private information of parents and students is correlated with the actual unobserved factors affecting student achievement. Table 2.3 shows the standard-deviation of the private information for the persistent factors for parents and students and how it correlates with the actual unobserved persistent factors affecting students' school performance. Finally, Table 2.4 shows how parents and students are using information from recent school scores versus older performance in order to update their expectations.

The results of Table 2.1 show that black parents expect significantly higher marks but their children do worse than average, implying black parents have a higher optimism bias. Children that are retained in lower grades have substantially higher marks than others, but the expectations of parents and students are similar to their peers, which implies these respondents are less biased than average. Female students are better at Math while their parents have similar expectations to others, which implies a lower overconfidence bias of parents in relation to girls. This result is consistent with parental stereotypes for genders (Carlana 2019). However, results for English/Reading are reversed, with parents being more overconfident for girls than for boys. The results of higher parental overconfidence for boys in Math and for girls in English/Reading happen across all school levels, being observed in Elementary School, Middle School and High School[17]. Children of parents with more age and education have higher marks but their parents have similar expectations as others, which again indicates a lower bias[18].

Parents use CAT scores to form their Math expectations, but students do not appear to do so, therefore in elementary school children adjust less to new information than parents.

Similar results were found for Math at the other school levels (first grade, elementary, middle and high school levels) and for the Reading/English marks, which are available in the web appendix. The only significant difference is that while repeating students have higher scores in elementary school, retention does not seem to help their grades during middle and high school.

4.2 Knowledge of persistent versus temporary unobserved factors

Table 2.2 shows the standard-deviation of the heterogeneity of temporary factors affecting students' elementary school performance and how the private information of families is correlated with this term. The estimates for the dispersion of the unobserved factors affecting current performance

[17] This can be verified in the Tables A.2.1, A.3.1, A.3.3 and A.3.5 of the web appendix.

[18] It is possible that age could be correlated with other factors such as school experience from previous siblings. Estimating a simple ordered probit model for the marks and respondents' predictions in the web appendix, I show that the parental predictions for Math (Fall) are lower with the number of older siblings. Also, students with older siblings have lower expectations for their English/Reading marks (in the Spring). This suggests that indeed the number of older siblings could be an additional factor. However, even after accounting for the number of older siblings, the models in the web appendix show that both the Math and English/Reading marks are higher for families with older mothers, while the respondents' expectations (parents' predictions for Math and English/Reading, plus the children's predictions for Math) are lower. This shows that indeed the respondents' overconfidence bias falls with the mother's age, even after accounting for the possible knowledge from other siblings.

show that female students and children of higher educated mothers depart less from the mean, therefore it is easier to predict their marks. Parents with higher education have a smaller deviation of unobserved private knowledge, which indicates that these parents follow less noisy prediction rules. Black parents, however, appear to have a greater dispersion in the knowledge about their children. Students of highly educated parents also have private information that is more correlated with actual achievement (Table 2.2), which shows their better use of private information.

Table 2.2: Unobserved heterogeneity of Math marks and expectations in Elementary school - Q1, 83/84-87/88

Standard-deviation of unobservables:

$\sigma_u(z_{i,t}) = \exp(z_{i,t}\gamma_u)$ $\quad \sigma_u^{p,b}(z_{i,t}) = \exp(z_{i,t}\gamma_u^{p,b})$, $b = Parents, Students$

Coefficients:	γ_u	$\gamma_u^{p,Parents}$	$\gamma_u^{p,Students}$
Constant	-0.14 (0.12)	-0.38 (0.12)*	0.13 (0.13)
black race	0.07 (0.05)	0.12 (0.07)**	0.04 (0.08)
female	-0.08 (0.05)	-0.09 (0.07)	-0.03 (0.08)
mother's years of education	-0.03 (0.01)*	-0.06 (0.02)*	-0.02 (0.02)

Controls for year-dummies are included.

Correlation of unobservables with marks:

$$\rho_u^{p,b}(z_{i,t}) = 2\left(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}\right) - 1,\ b = Parents, Students$$

Coefficients:	$\theta_u^{p,Parents}$	$\theta_u^{p,Students}$
Constant	0.15 (0.15)	-0.03 (0.18)
black race	-0.2 (0.13)	-0.08 (0.17)
female	0.04 (0.12)	-0.05 (0.15)
mother's years of education	-0.01 (0.03)	0.07 (0.04)**

Controls for year-dummies are included.

Standard-deviations in (), * and ** denote 5% and 10% statistical significance.

I now compare families' private information about new factors affecting school performance (Table 2.2) with parents and students' private information about the persistent factors affecting their performance (Table 2.3). The constants for the correlation of parents and students' private information with persistent performance factors (Table 2.3) are larger than for the correlation of private information with temporary factors (Table 2.2). This indicates that parents and students know more about persistent factors affecting their performance than about new factors previously unknown. However, the standard-errors of the estimation are also large, therefore the model cannot reject the hypothesis that parents and students' know little or nothing about the unobserved elements affecting elementary school performance. The next section will show that the correlation of parents and students' private information with the actual factors affecting school performance increases substantially during middle school and high school.

Table 2.3: Heterogeneity of families' persistent expectations and Math marks

Coefficients of random effects

	Marks: $\sigma_\alpha(z_i)$	**Parents:** $\sigma_\alpha^{p,Parents}(z_i)$	**Students:** $\sigma_\alpha^{p,Students}(z_i)$
Coefficients:	γ	$\gamma^{p,Parents}$	$\gamma^{p,Students}$
Constant	-3.95 (2.08)**	-0.4 (0.29)	-1.41 (0.26)*
black race	-1.27 (1.04)	0.06 (0.27)	-0.3 (0.25)
female	-2.49 (5.67)	-0.15 (0.26)	0.08 (0.24)
mother's years of education	0.43 (0.26)*	0.11 (0.06)**	0.01 (0.07)

Correlation of random-effects between Expectations and Marks

$$\rho_\alpha^{p,b}(z_i) = 2\left(\frac{\exp(z_{i,1}\theta_\alpha^{p,b})}{1+\exp(z_{i,1}\theta_\alpha^{p,b})}\right) - 1,\ b = Parents, Students$$

	Parents:	**Students:**
Coefficients:	$\theta_\alpha^{p,Parents}$	$\theta_\alpha^{p,Students}$
Constant	1.25 (0.76)	2.01 (1.23)
black race	-1.17 (0.57)*	0.5 (0.94)
female	-0.56 (0.52)	0.66 (1.58)
mother's years of education	0.1 (0.11)	0.12 (1.65)

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

Parents of black race have a random-effect information that is less correlated with the actual random-effect in student achievement (table 2.3), which indicates that they have less access to good private information relative to white families.

4.3 How do respondents use past information to form their expectations?

Table 2.4 presents the value of parents and students' coefficients for the past marks and compares them with the coefficients that are given by the actual process of students' marks. In elementary school both parents and students use recent academic performance (the last 4 marks) to update their expectations. However, the coefficients for predicting student achievement actually show that parents and students in elementary school should rely much more on the entire past school performance of the student, rather than just on last year's performance. Also, in the spring quarter of each academic year students do not rely as much as they should on their four most recent grades.

Table 2.4: Families' expectations use of the average marks in the last year and all past performance

	Elementary school		Middle school		High school	
	Q1	Q4	Q1	Q4	Q1	Q4
Students' expectations: $mp(z_{i,t})^{Students} = [x_{i,t}, Y_{i,t-1}]\,\beta^{p,Students}$						
All past marks	0.45 (1.18)	-2.54 (1.17)*	0.82 (1.04)	-0.57 (0.83)		1.47 (1.35)
Last 4 quarters	2.3 (1)*	6.84 (1.19)*	3.01 (0.62)*	7.27 (0.88)*		6.13 (0.92)*
Parents' expectations: $mp(z_{i,t})^{Parents} = [x_{i,t}, Y_{i,t-1}]\,\beta^{p,Parents}$						
All past marks	2.2 (0.67)*		4.96 (0.92)*		3.93 (1.09)*	
Last 4 quarters	2.05 (0.53)*		2.7 (0.44)*		4.64 (0.62)*	
Mean Achievement: $m(z_{i,t}) = [x_{i,t}, Y_{i,t-1}]\,\beta$						
All past marks	3.78 (0.82)*	-1.02 (0.63)	5.28 (1.18)*	-0.05 (0.84)	3.81 (1.16)*	2.82 (1.17)*
Last 4 quarters	0.91 (0.6)	12.72 (1.38)*	3.78 (0.66)*	10.55 (1.22)*	7.41 (0.86)*	10 (1.19)*

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

In middle school and high school parents and students rely more on both the last year's school performance and their entire past school performance. This updating procedure is actually very close to the actual effect of past scores estimated for the mean process of student achievement. Parents and students, therefore, improved substantially their use of the observable information given by past marks from elementary school to middle school and high school.

The main difference between parents and students' predictions and the real process of student achievement in middle school and high school is that parents and students did not rely as much on the information in the last four academic quarters as it would be efficient. This means that parents and students are slower than optimal to update their expectations during each academic year. Similar conclusions can be drawn from the families' expectations for English marks.

4.4 Bias and variance of prediction errors: Overview of results

This section evaluates the mean values of bias and the variance of respondents' prediction errors adjusted for the rational variance term ($\sigma_{i,t}^{p2} - 2\rho_{i,t}\sigma_{i,t}\sigma_{i,t}^{p}$), across 4 demographic groups (the whole population, black students, repeating students and families with mothers that have completed high school). I then report how these mean values evolved over elementary (Table 3.1), middle (Table 3.2), and high school (Tables 3.3 and 3.4). Since these average values are complex functions of several parameters I report bootstrap standard-errors of the values of each group.

Table 3.1: Bias and variance of expectations (Math, Fall - Elementary school, 82-87)

Variables / Demographic group	All	Black	Repeaters	Mothers w/ high school
correlation of parents and marks	0.13 (0.05)	0.10 (0.05)	0.08 (0.06)	0.13 (0.06)
mean bias of parental expectations	0.75 (0.15)	0.86 (0.17)*	0.71 (0.18)	0.72 (0.14)
variance of parental expectations' errors	0.83 (0.23)	0.89 (0.26)	1.04 (0.32)**	0.76 (0.23)*
variance of parents - "rational" agent	0.26 (0.09)	0.29 (0.10)	0.36 (0.13)**	0.24 (0.12)
correlation of students and marks	0.14 (0.03)	0.14 (0.04)	0.13 (0.04)	0.16 (0.04)
mean bias of students' expectations	1.15 (0.14)	1.27 (0.15)*	1.30 (0.19)*	1.05 (0.13)
variance of students expectations' errors	1.15 (0.28)	1.18 (0.31)	1.29 (0.38)*	1.05 (0.27)
variance of students - "rational" agent	0.59 (0.16)	0.59 (0.18)	0.67 (0.21)*	0.53 (0.15)

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level.

The results confirm that black parents and students are significantly more biased than white families at all school levels, but they have a similar prediction variance as white families. Parents and students of families where the mother has a high school degree are better predictors than average, since both their bias and prediction variance are lower than their peers. Parents and students decreased both their bias and prediction variance substantially from elementary to middle and high school. This implies that families improved their predictions as children aged, due to an improved use of both the observable information given by past marks (Table 2.4) and a better knowledge of unobservable factors affecting their school performance (Tables 3.1 and 3.2).

Table 3.2: Bias and variance of expectations (Math, Fall - Middle school)

Variables / Demographic group	All	Black	Repeaters	Mothers w/ high school
correlation of parents and marks	0.22 (0.04)	0.20 (0.05)	0.18 (0.05)**	0.24 (0.05)
mean bias of parental expectations	0.56 (0.13)	0.61 (0.14)*	0.66 (0.16)*	0.50 (0.12)*
variance of parental expectations' errors	0.62 (0.29)	0.60 (0.28)	0.72 (0.33)**	0.56 (0.28)*
variance of parents - "rational" agent	0.13 (0.06)	0.15 (0.07)	0.18 (0.09)	0.11 (0.06)
correlation of students and marks	0.35 (0.04)	0.35 (0.05)	0.32 (0.05)	0.36 (0.05)
mean bias of students' expectations	0.58 (0.11)	0.62 (0.13)*	0.67 (0.13)*	0.52 (0.11)*
variance of students expectations' errors	0.64 (0.22)	0.63 (0.22)	0.80 (0.24)*	0.59 (0.22)*
variance of students - "rational" agent	0.11 (0.05)	0.13 (0.06)	0.23 (0.10)*	0.08 (0.06)**

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level.

Table 3.3: Bias and variance of expectations (Parents - Math, Fall, High school)

Variables / Demographic group	All	Black	Repeaters	Mothers w/ high school
correlation of parents and marks	0.22 (0.12)	0.05 (0.17)**	0.07 (0.15)*	0.28 (0.12)
mean bias of parents' expectations	0.44 (0.10)	0.51 (0.11)*	0.49 (0.15)	0.40 (0.09)**
variance of parents expectations' errors	0.32 (0.15)	0.35 (0.19)	0.41 (0.19)**	0.25 (0.14)*
variance of parents - "rational" agent	0.08 (0.07)	0.14 (0.11)	0.15 (0.10)**	0.04 (0.05)

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level.

Table 3.4: Bias and variance of students' expectations (Math - Spring, High school)

Variables / Demographic group	All	Black	Repeaters	Mothers w/ high school
correlation of students and marks	0.31 (0.06)	0.38 (0.06)**	0.35 (0.09)	0.28 (0.07)
mean bias of students' expectations	0.56 (0.18)	0.61 (0.20)	0.55 (0.19)	0.53 (0.18)*
variance of students expectations' errors	0.29 (0.24)	0.30 (0.26)	0.29 (0.25)	0.26 (0.26)
variance of students - "rational" agent	0.00 (0.03)	-0.02 (0.05)	-0.01 (0.06)	0.02 (0.03)

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level.

Note that the correlation of parents and students' private information with actual school performance increased substantially from elementary school (Table 3.1) to middle school (Table 3.2) and high school (Tables 3.3 and 3.4), especially among students. In elementary school, parents and students' private information only had a 14% correlation with actual school performance. However, during middle and high school the correlation of private information with actual unobserved factors affecting academic grades was 22% for parents and 30 to 35% for students. Parents are as knowledgeable as children at an early age, but teenagers have a higher assessment of both observable and unobservable factors affecting achievement than their parents. This result shows that the role of private knowledge of unobserved factors increases substantially as children grow older. Therefore economic models of education and human capital that ignore private information held by teenagers are making unrealistic assumptions. One policy implication of this result is for school systems to give better and more frequent updating on academic performance to parents. For instance, Bergman (2020) in an experiment at a Los Angeles K-12 school, also shows that parents' beliefs about their children's academic achievement and effort are upwardly biased. However, he finds it is relatively cheap to send frequent text reports to parents about their children's school assignments, which has a positive impact on the children's academic achievement through a combination of more accurate beliefs and improved monitoring. Bergman et al. (2018) confirms that simply sending parents more frequent information can improve school retention, but that other complementary policies can consider both increasing the information of parents and also their academic parenting skills through home visits. Furthermore, it is also possible that information sent to parents can have positive spillovers to other parents and children (Berlinski et al. 2016).

Similar results are found for the parents and students' English/Reading forecasts and for the students' spring quarter expectations, which are available in the web appendix.

5 Conclusions

This paper studies how families learn to predict their academic performance. The prediction errors of the agents can be explained by several factors: 1) excessive optimism; 2) slowness to adjust to new information on the students' academic performance; and 3) families rely on noisy private information. To evaluate the role of these factors, I specify and estimate a model of student achievement and expectation formation. I then use the results of this model to compare the bias and variance of predictions across demographic groups and how respondents use the information of past marks in relation to their true predictive value.

I show that respondents are overconfident and use noisy private information, with overconfidence being smaller for families with higher education. Female students, children of parents with more age and education have higher marks but their parents have similar expectations as others, which indicates a lower overconfidence of the parents in these groups.

There are also racial differences in how families form expectations. Black families are more overconfident than average, but over the years they made similar gains as white families in their ability to forecast. Parents' predictions in middle school and high school adjusted too slowly to new information in recent school scores. After controlling for parental education, black students are more likely to stay additional years studying before dropping out of high school (Lang and Manove 2011), even if they have similar returns to education as white families (Carneiro et al. 2005), a higher distaste for schooling and similar discount rates as white families (Lang and Rudd 1986). Black parents and students are more overconfident about their academic scores even after several years of schooling, which may explain the puzzle of why they stay longer in school.

In elementary school, students had a larger bias and prediction variance than their parents. Families also updated their expectations more slowly than would be efficient. In middle school, families' bias and prediction variance decreased substantially, particularly among students who became as good predictors as their parents. In addition, both parents and students made better use of the information available in past marks.

The model also shows that parents and students possess significant private information about unobservable factors. The accuracy of this private information - as measured by its correlation with actual academic scores - increased for both parents and students during middle school and high school. Students in particular have a much higher degree of private information than their parents during high school. This result shows that the role of private knowledge of unobserved factors increases substantially as children grow older. Economic models of human capital decisions, therefore, should take into account higher degrees of private information as students grow older.

The results of this paper also suggest a new view about the causes of why families fail to make more early childhood investments, despite their high return (Cunha et al. 2005). I show that a large proportion of parents at the start of elementary school expected their children to receive high marks and reach high levels of education. Overconfident parents could therefore fail to foresee that these investments are essential to their children's academic success.

Finally, several academic associations (such as the Econometric Society (ES), American Economic Association (AEA), the National Economics Association (NEA) and the European Economic Association (EEA)) have pointed out that the topic of racial inequality is still under-researched in economics. The diverse background of the BSS families could therefore prove a useful source for future research into the socioeconomic plight of disadvantaged minorities.

References

[1] Alexander, K., D. Entwisle and M. Thompson (1988), "Household composition, parental expectations, and school achievement", *Social Forces*, 67 (2), 424-451.

[2] Alexander, K., D. Entwisle and S. Dauber (2003), "On the Success of Failure: A Reassessment of the Effects of Retention in the Primary Grades," 2nd ed., Cambridge University Press.

[3] Alexander, Karl L.; and Entwisle, Doris R., 2004 -10-06, "The Beginning School Study, 1982-2002 (Log# 01293)", hdl: 1902.1/01293, http://id.thedata.org/hdl%3A1902.1%2F01293 Henry A. Murray Research Archive [distributor (DDI)].

[4] Attanasio, O., I. Almas and P. Jervis (2020), "Economics and measurement: new measures to model decision making," Presidential address to the World Congress of the Econometric Society, Bocconi University.

[5] Attanasio, O. and K. Kaufmann (2014), "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender," *Journal of Development Economics*, 109(C), 203-216.

[6] Bergman, P., E. Edmond-Verley and N. Notario-Risk (2018), "Parent Skills and Information Asymmetries: Experimental Evidence from Home Visits and Text Messages in Middle and High Schools," *Economics of Education Review*, 66, 92-103.

[7] Bergman, P. (2020), "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment," *Journal of Political Economy*, forthcoming.

[8] Berlinski, S., M. Busso, T. Dinkelman and C. Martinez (2016), "Reducing parent-school information gaps and improving education outcomes: Evidence from high frequency text messaging in Chile," IADB, mimeo.

[9] Berry, J. (2015), "Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India," *Journal of Human Resources*, 50(4), 1051-1080.

[10] Bugni, F. (2010), "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735-753.

[11] Carlana, M. (2019), "Implicit Stereotypes: Evidence from Teachers' Gender Bias," *Quarterly Journal of Economics*, 134(3), 1163-1224.

[12] Carneiro, P., J. Heckman and D. Masterov (2005), "Labor market discrimination and racial differences in premarket factors", *Journal of Law and Economics*, 48, 1-39.

[13] Conlon, J., L. Pilossoph, M. Wiswall and B. Zafar (2018), "Labor Market Search With Imperfect Information and Learning," NBER WP 24988.

[14] Cornand, C. and P. Hubert (2020), "On the external validity of experimental inflation forecasts: A comparison with five categories of field expectations," *Journal of Economic Dynamics and Control*, 110, 103746.

[15] Cunha, F., J. Heckman, L. Lochner and D. Masterov (2005), "Interpreting the evidence on life cycle skill formation", in *Handbook of the Economics of Education*.

[16] Delalibera, B. and P. Ferreira (2019), "Early childhood education and economic growth," *Journal of Economic Dynamics and Control*, 98, 82-104.

[17] Delavande, A. and S. Rohwedder (2011), "Individuals' uncertainty about future social security benefits and portfolio choice", *Journal of Applied Econometrics*, 26, 498-519.

[18] Entwisle, D., K. Alexander and L. Olson (1997), "Children, schools and inequality", *Social Inequality Series*, Westview Press.

[19] Geweke, J., M. Keane and D. Runkle (1994), "Alternative computational approaches to inference in the multinomial probit model", *Review of Economics and Statistics*, 76, 609-632.

[20] Giustinelli, P. (2016), "Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of the High School Track," *International Economic Review*, 57(2), 573-602.

[21] Giustinelli, P. and C. Manski (2018), "Survey Measures of Family Decision Processes for Econometric Analysis of Schooling Decisions," *Economic Inquiry*, 120(3), 359-397.

[22] Hess, S., J. Polak and K. Train (2004), "On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice", mimeo.

[23] Hoelzl, E. and A. Rustichini (2005), "Overconfident: Do you put your money on it?", *Economic Journal*, 115, 305-318.

[24] Houser, D., M. Keane and K. McCabe (2004), "Behavior in a dynamic decision problem: An analysis of experimental evidence using a Bayesian type classification algorithm", *Econometrica*, 72, 781-822.

[25] Jervis, P. (2017), "A Dynamic Model of Early Parental Investments in Children's Human Capital," PhD dissertation chapter.

[26] Lang, K. and P. Ruud (1986), "Returns to schooling, implicit discount rates and black-white wage differentials", *Review of Economics and Statistics*, 68, 41-47.

[27] Lang, K. and M. Manove (2011), "Education and Labor Market Discrimination," *American Economic Review*, 101(4), 1467-1496.

[28] Madeira, C. and B. Zafar (2015), "Heterogeneous Inflation Expectations and Learning," *Journal of Money, Credit and Banking*, 47(5), 867-896.

[29] Madeira, C. (2018), "Testing the rationality of expectations of qualitative outcomes", *Journal of Applied Econometrics*, 33(6), 837-852.

[30] Madeira, C. (2019), "Measuring the covariance risk of consumer debt portfolios", *Journal of Economic Dynamics and Control*, 2019, 109, 21-38.

[31] Manski, C. (2018), 'Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise', *NBER Macroeconomics Annual*, 32, 411-471.

[32] Oyserman, D. (2015), "Pathways to success through identity-based motivation," Oxford, UK: Oxford University Press.

[33] Rivkin, S. (1995), "Black/White Differences in Schooling and Employment," *Journal of Human Resources*, 30, 826-52.

[34] Stinebrickner, T. and R. Stinebrickner (2012), "Learning about Academic Ability and the College Dropout Decision," *Journal of Labor Economics*, 30(4), 707-748.

[35] Stinebrickner, R. and T. Stinebrickner (2014), "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model," *Journal of Labor Economics*, 32(3), 601-644.

[36] Ting Lee, M.-L. (1996), "Properties and applications of the Sarmanov family of bivariate distributions," *Communications in Statistics - Theory and Methods*, 25 (6), 1207-1222.

[37] van der Klaauw, W. (2012), "On the Use of Expectations Data in Estimating Structural Dynamic Models," *Journal of Labor Economics*, 30(3), 521–554.

[38] Zellner, A. and T. Lee (1965), "Joint estimation of relationships involving discrete random variables", *Econometrica*, 33, 382-394.

6 Data description

Table A.1.1 - Distribution of Students marks:

Letter mark (Reading/English, 1982-89)	1982-89 Q1	1982-89 Q2	1982-89 Q3	1982-89 Q4
1 - Unsatisfactory	23.2%	20.4%	17.9%	17.3%
2 - Satisfactory	52.3%	49.7%	48.2%	46.7%
3 - Good	20.2%	23.2%	26.1%	26.2%
4 - Excellent	4.3%	6.7%	7.8%	9.7%
Missing	38.2%	37.9%	37.0%	38.0%

All statistics are a percentage of the observed sample, except for the missing sample values which are a percentage of the whole population.

Table A.1.2 - Respondents' expectations	Parents		Students	
Best guess for the Reading/English (Fall) mark	**1982**	**1982-90**	**1982**	**1982-90**
1 - Unsatisfactory	5.1%	4.7%	3.3%	8.4%
2 - Satisfactory	37.3%	32.1%	16.8%	17.4%
3 - Good	44.8%	43.3%	36.5%	40.6%
4 - Excellent	12.8%	20.0%	43.4%	33.6%
Nr of observations	788	3,389	821	3,322

All percentage values of the samples are given as % of the non-missing sample that took the class.

Table A.1.3 - Difference between respondents' predictions and the mark obtained (English/Reading, Fall)	**Parents**		**Students**	
	1982	**1982-90**	**1982**	**1982-90**
-3		0.1%	0.1%	0.0%
-2	0.1%	0.5%	0.1%	0.4%
-1	3.7%	5.9%	4.1%	5.2%
0	30.7%	32.3%	16.3%	21.0%
1	47.9%	44.7%	32.2%	38.1%
2	16.7%	15.2%	34.9%	27.2%
3	0.8%	1.3%	12.2%	8.0%
Number of observations	724	2,584	748	2,619

Note: positive values correspond to optimism/overconfidence error.

Table A.1.4: Ordered Probit models for the Math marks and predictions with the added controls of the parents' identity and the number of older siblings

	Marks		Predictions		
VARIABLES	Fall	Spring	Parents (Fall)	Students (Fall)	Students (Spring)
Father in first BSS	0.122**	-0.0249	-0.0709	-0.174**	-0.116
survey dummy	(0.0646)	(0.0651)	(0.0834)	(0.0900)	(0.0722)
Other Relative	-0.0911	0.0991	-0.218**	-0.0208	0.0400
in first BSS survey	(0.0846)	(0.0861)	(0.113)	(0.112)	(0.0929)
Always the same	-0.0653**	-0.0321	-0.0326	-0.0529	-0.0272
parent respondent	(0.0391)	(0.0397)	(0.0517)	(0.0534)	(0.0438)
Nr of older siblings	-0.0212	-0.00934	-0.0463*	0.0251	0.0165
	(0.0148)	(0.0150)	(0.0195)	(0.0197)	(0.0166)
black race	-0.229*	-0.0825*	0.169*	0.0266	0.106*
	(0.0344)	(0.0347)	(0.0437)	(0.0466)	(0.0389)
female	0.161*	0.0397	0.0143	0.0171	-0.0442
	(0.0321)	(0.0323)	(0.0412)	(0.0436)	(0.0358)
Student's month of birth	-0.000793	-0.00745	-0.0110**	0.0104**	-0.00456
	(0.00459)	(0.00463)	(0.00589)	(0.00623)	(0.00511)
age of mother at birth	0.00698*	0.00883*	-0.00878*	-0.00805**	-0.00792*
	(0.00349)	(0.00350)	(0.00448)	(0.00473)	(0.00388)
mom's years of education	0.0530*	0.0230*	-0.0134	0.0207**	-0.00351
	(0.00818)	(0.00822)	(0.0103)	(0.0112)	(0.00912)
nr of failed grades	-0.0487**	0.0643*	-0.0350	0.0949*	-0.00288
	(0.0270)	(0.0280)	(0.0364)	(0.0397)	(0.0320)
Average of all past marks	0.0293*	0.0319*	0.0357*	0.0143*	0.000923
	(0.00269)	(0.00269)	(0.00341)	(0.00366)	(0.00289)
Average of last 4 quarters	0.0256*	0.125*	0.0351*	0.0174*	0.0586*
	(0.00185)	(0.00224)	(0.00248)	(0.00268)	(0.00211)
All the regressions control for year dummies.					
Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level					
Observations	4,906	5,838	3,059	2,678	4,122

Table A.1.5: Ordered Probit models for the Reading/English marks and predictions with the added controls of the parents' identity and the number of older siblings

	Marks		Predictions		
VARIABLES	Fall	Spring	Parents (Fall)	Students (Fall)	Students (Spring)
Father in first BSS	0.149**	0.0480	-0.0595	-0.0532	-0.181**
survey dummy	(0.0869)	(0.0880)	(0.0986)	(0.0994)	(0.0937)
Other Relative	0.0580	0.0991	-0.176	0.0159	0.177
in first BSS survey	(0.109)	(0.113)	(0.131)	(0.123)	(0.116)
Always the same	0.0423	-0.0166	-0.223*	-0.0117	-0.0291
parent respondent	(0.0508)	(0.0515)	(0.0595)	(0.0581)	(0.0545)
Nr of older siblings	-0.0250	-0.0319	-0.00783	-0.0168	-0.0546*
	(0.0194)	(0.0197)	(0.0223)	(0.0215)	(0.0205)
black race	-0.294*	-0.0789**	0.189*	-0.0586	0.0588
	(0.0449)	(0.0452)	(0.0507)	(0.0515)	(0.0482)
female	0.179*	0.0837*	0.165*	0.0407	0.133*
	(0.0419)	(0.0422)	(0.0474)	(0.0475)	(0.0447)
Student's month of birth	-0.00918	-0.0145*	0.00326	0.0132*	-0.000878
	(0.00593)	(0.00600)	(0.00670)	(0.00674)	(0.00635)
age of mother at birth	0.0109*	0.0134*	-0.0179*	0.00484	0.00267
	(0.00467)	(0.00468)	(0.00523)	(0.00522)	(0.00490)
mom's years of education	0.0859*	0.0591*	-0.00619	0.0178	-0.00898
	(0.0112)	(0.0112)	(0.0122)	(0.0127)	(0.0117)
nr of failed grades	0.0789**	0.00536	-0.112*	0.155*	-0.0178
	(0.0404)	(0.0409)	(0.0454)	(0.0446)	(0.0456)
Average of all past marks	0.0310*	0.0403*	0.0422*	0.00887*	0.00423
	(0.00357)	(0.00352)	(0.00411)	(0.00406)	(0.00368)
Average of last 4 quarters	0.0426*	0.159*	0.0341*	0.00790*	0.0405*
	(0.00297)	(0.00337)	(0.00332)	(0.00322)	(0.00302)
All the regressions control for year dummies.					
Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level					
Observations	3,059	4,011	2,331	2,264	2,636

7 Model coefficients for the respondents' expectations

7.1 Bias of respondents' expectations across demographic groups

Table A.2.1: Reading/English marks and expectations in Elementary school - Q1, 83/84-87/88

	Marks (w/ students)	**Parents**	**Students**
	$m(z_{i,t})$	$mp(z_{i,t})^{Parents}$	$mp(z_{i,t})^{Students}$
	Mean of latent marks/ expectations:		
Constant	0.12 (0.01)*	1.95 (0.22)*	2.75 (0.01)*
black race	-0.18 (0.03)*	0.3 (0.08)*	-0.08 (0.05)**
female	-0.02 (0.02)	0.15 (0.08)**	0.07 (0.04)
Student's month of birth	0 (0)	0 (0.01)	0.02 (0.01)
age of mother at birth (× 10)	-0.12 (0.01)*	-0.2 (0.08)*	0.02 (0.01)**
mother's years of education	0.07 (0)*	-0.01 (0.02)	0.02 (0.01)*
nr of failed grades	0.28 (0.02)*	-0.32 (0.08)*	-0.2 (0.02)*
CAT score	0.37 (0.01)*	0.25 (0.08)*	-0.16 (0.01)*
Average of all past marks	2.8 (0.01)*	4.4 (1.26)*	0.96 (0.05)*
Average of last 4 quarters	3.44 (0)*	3.73 (0.99)*	1.47 (0.08)*
(controls for year-dummies)	yes	yes	yes
Nr of observations (Elementary school$^{a)}$)	2,358	1,497	1,311
Nr of observations (All school levels)	6,720	3,335	6,904

a) Nr of observations in Elementary School (Fall quarter) after the First Grade (1982).

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

7.2 Knowledge of persistent versus temporary unobserved factors

Table A.2.2: Unobserved heterogeneity of English/Reading marks and expectations in Elementary school - Q1, 83/84-87/88

Standard-deviation of unobservables ($b = Parents, Students$):

	$\sigma_u(z_{i,t}) = \exp(z_{i,t}\gamma_u)$	$\sigma_u^{p,b}(z_{i,t}) = \exp(z_{i,t}\gamma_u^{p,b})$	
Coefficients:	γ_u	$\gamma_u^{p,Parents}$	$\gamma_u^{p,Students}$
Constant	-0.41 (0.01)*	-0.03 (0.12)	0.39 (0.12)*
black race	0.01 (0.01)	0.19 (0.07)*	0.11 (0.08)
female	-0.04 (0.03)	-0.02 (0.07)	-0.03 (0.04)
mother's years of education	0 (0.02)	-0.05 (0.02)*	-0.02 (0)*

Controls for year-dummies are included.

Correlation of unobservables with marks:

$$\rho_u^{p,b}(z_{i,t}) = 2\left(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}\right) - 1$$

Coefficients:	$\theta_u^{p,Parents}$	$\theta_u^{p,Students}$
Constant	0.31 (0.19)	0.17 (0.02)*
black race	-0.16 (0.16)	-0.16 (0.1)
female	0.07 (0.14)	-0.2 (0.05)*
mother's years of education	0.03 (0.03)	0.06 (0.03)*

Controls for year-dummies are included.

Standard-deviations in (), * and ** denote 5% and 10% statistical significance.

Table A.2.3: **Heterogeneity of families' persistent expectations and English/Reading marks**

Coefficients of random effects

	Marks: $\sigma_\alpha(z_i)$	**Parents:** $\sigma_\alpha^{p,Parents}(z_i)$	**Students:** $\sigma_\alpha^{p,Students}(z_i)$
Coefficients:	γ	$\gamma^{p,Parents}$	$\gamma^{p,Students}$
Constant	-2.05 (0.02)*	-0.49 (0.33)	-1.32 (0.07)*
black race	-1.04 (0.1)*	0.44 (0.28)	-0.48 (0.04)*
female	0.24 (0.03)*	-0.55 (0.27)*	0.24 (0.01)*
mother's education	-0.05 (0.01)*	-0.07 (0.05)	-0.11 (0)*

Correlation of random-effects between Expectations and Marks

$$\rho_\alpha^{p,b}(z_i) = 2\left(\frac{\exp(z_{i,1}\theta_\alpha^{p,b})}{1+\exp(z_{i,1}\theta_\alpha^{p,b})}\right) - 1,\ b = Parents, Students$$

	Parents:	**Students:**
Coefficients:	$\theta_\alpha^{p,Parents}$	$\theta_\alpha^{p,Students}$
Constant	-2.31 (2.97)	0.73 (0.01)*
black race	0.6 (269.85)	-0.27 (0.02)*
female	-1.68 (1.45)	0.14 (0.03)*
mother's education	1.21 (1.04)	0.44 (0.03)*

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

7.3 How do respondents use past information to form their expectations?

Table A.2.4: Families' expectations use of the average marks in the last year and all past performance (English/Reading)

	Elementary school		Middle school		High school	
	Q1	Q4	Q1	Q4	Q1	Q4
Students' expectations: $mp(z_{i,t})^{Students} = [x_{i,t}, Y_{i,t-1}]\,\beta^{p,Students}$						
All past marks	0.96 (0.05)*	1.43 (0.03)*	0.65 (0.67)	0.18 (0.58)		0.3 (1.06)
Last 4 quarters	1.47 (0.08)*	3.16 (1.50)*	1.99 (0.38)*	5.59 (0.34)*		2.93 (0.58)*
Parents' expectations: $mp(z_{i,t})^{Parents} = [x_{i,t}, Y_{i,t-1}]\,\beta^{p,Parents}$						
All past marks	4.4 (1.26)*		7.44 (1.43)*		5.69 (1.83)*	
Last 4 quarters	3.73 (0.99)*		2.75 (0.67)*		4.73 (0.93)*	
Mean Achievement: $m(z_{i,t}) = [x_{i,t}, Y_{i,t-1}]\,\beta$						
All past marks	2.8 (0.01)*	0.07 (0.01)*	2.26 (0.73)*	1.41 (0.57)*	2.34 (2.37)	0.62 (1.26)
Last 4 quarters	3.44 (0)*	-0.11 (0.01)*	3.43 (0.47)*	8.64 (0.42)*	11.01 (1.54)*	4.24 (0.86)*

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

7.4 Expectations for Math and Reading/English in other school levels

7.4.1 First quarter (Fall) of first-grade (first period of the model)

Table A.3.1: Math and Reading/English marks and expectations in 1st grade - Q1, 1982

	Marks (w/ students)	Parents	Students
	$m(z_{i,t})$	$mp(z_{i,t})^{Parents}$	$mp(z_{i,t})^{Students}$
Math - Mean of latent marks/ expectations:			
Constant	1.84 (0.26)*	2.38 (0.18)*	3.52 (0.36)*
black race	-0.51 (0.13)*	-0.04 (0.06)	-0.07 (0.14)
female	0.04 (0.11)	0.02 (0.06)	-0.19 (0.14)
Student's month of birth	-0.05 (0.02)*	-0.04 (0.01)*	-0.02 (0.02)
age of mother at birth (× 10)	0.07 (0.11)	-0.01 (0.06)	-0.01 (0.14)
mother's education	0.14 (0.03)*	0.11 (0.02)*	0.01 (0.03)
nr of failed grades	0.12 (0.38)	-0.15 (0.17)	-0.01 (0.38)
Observations (First Grade, Fall)	757	705	735
Observations (All school levels)	6,720	3,335	6,904
English/Reading - Mean of latent marks/ expectations:			
Constant	1.17 (0.04)*	2.68 (0.25)*	2.92 (0.08)*
black race	-0.42 (0.18)*	-0.03 (0.11)	-0.09 (0.11)
female	0.03 (0.02)	0.19 (0.11)**	0.13 (0+0.01i)*
Student's month of birth	-0.03 (0.02)	-0.05 (0.02)*	-0.01 (0.01)
age of mother at birth (× 10)	0.07 (0.01)*	0.1 (0.1)	0.02 (0)*
mother's education	0.12 (0.02)*	0.13 (0.03)*	0.01 (0.03)
nr of failed grades	-0.25 (0.22)	-0.55 (0.25)*	0.24 (0.25)
Observations (First Grade, Fall)	758	707	733
Observations (All school levels)	6,720	3,335	6,904

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

Table A.3.2: Unobserved heterogeneity of marks and expectations in First-grade: Q1, 1982

Standard-deviation of unobservables ($b = Parents, Students$):

	Math			**Reading/English**		
	$\sigma_u = e^{z_{i,t}\gamma_u}$	$\sigma_u^{p,b}(z_{i,t}) = \exp(z_{i,t}\gamma_u^{p,b})$		$\sigma_u = e^{z_{i,t}\gamma_u}$	$\sigma_u^{p,b}(z_{i,t}) = \exp(z_{i,t}\gamma_u^{p,b})$	
Coefficients:	γ_u	$\gamma_u^{p,Parents}$	$\gamma_u^{p,Students}$	γ_u	$\gamma_u^{p,Parents}$	$\gamma_u^{p,Students}$
Constant		-0.28 (0.11)*	0.24 (0.16)		0.1 (0.12)	-0.02 (0.04)
black race	0.08 (0.09)	-0.07 (0.07)	0.11 (0.11)	-0.07 (0.09)	-0.09 (0.08)	0.01 (0.03)
female	0.04 (0.09)	0.11 (0.07)	-0.15 (0.1)	-0.18 (0.01)*	0.06 (0.08)	-0.05 (0.03)
mother's education	-0.03 (0.02)	-0.01 (0.01)	-0.02 (0.03)	-0.03 (0.01)*	0.02 (0.02)	0.04 (0.01)*

Correlation of unobservables with marks:

		$\rho_u^{p,b}(z_{i,t}) = 2(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}) - 1$			$\rho_u^{p,b}(z_{i,t}) = 2(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}) - 1$	
Coefficients:		$\theta_u^{p,Parents}$	$\theta_u^{p,Students}$		$\theta_u^{p,Parents}$	$\theta_u^{p,Students}$
Constant		0.79 (0.25)*	0.19 (0.29)		0.65 (0.27)*	-0.52 (0.03)*
black race		0.07 (0.23)	-0.07 (0.26)		0.17 (0.28)	0.16 (0.11)
female		0.1 (0.22)	0.27 (0.26)		0.55 (0.27)*	0.41 (0.01)*
mother's education		-0.04 (0.05)	-0.05 (0.06)		0.02 (0.06)	0.09 (0.04)*

Standard-deviations in (), * and ** denote 5% and 10% statistical significance.

7.4.2 Middle school (Fall quarter)

Table A.3.3: Marks and expectations in Middle school - Fall

	Marks (w/ students)	**Parents**	**Students**
	$m(z_{i,t})$	$mp(z_{i,t})^{Parents}$	$mp(z_{i,t})^{Students}$
Mean of latent marks/ expectations:			
Exogenous variables		**Math:**	
Constant	-0.37 (0.3)	1.48 (0.21)*	1.75 (0.26)*
black race	-0.05 (0.09)	0.22 (0.07)*	0.1 (0.09)
female	0.12 (0.08)	0.09 (0.07)	-0.08 (0.07)
Student's month of birth	0.02 (0.01)**	-0.02 (0.01)	0.02 (0.01)
age of mother at birth (× 10)	0.1 (0.07)	-0.06 (0.06)	-0.02 (0.06)
mother's years of education	0.03 (0.02)	-0.03 (0.02)	0 (0.02)
nr of failed grades	-0.09 (0.08)	0.03 (0.05)	0.04 (0.06)
CAT score	0.06 (0.07)	0.01 (0.05)	0.06 (0.07)
Average of all past marks	5.28 (1.18)*	4.96 (0.92)*	0.82 (1.04)
Average of last 4 quarters	3.78 (0.66)*	2.7 (0.44)*	3.01 (0.62)*
(controls for year-dummies)	yes	yes	yes
Nr of observations (Middle School, Fall)	1,892	1,421	1,394
Nr of observations (All school levels)	6,720	3,335	6,904
Exogenous variables		**English/Reading:**	
Constant	0.33 (0.17)**	1.18 (0.27)*	1.81 (0.14)*
black race	-0.03 (0.06)	0.25 (0.1)*	0.07 (0.05)
female	0.17 (0.05)*	0.24 (0.09)*	0 (0.05)
Student's month of birth	0.01 (0.01)	-0.02 (0.01)*	0 (0.01)
age of mother at birth (× 10)	0.04 (0.04)	-0.06 (0.07)	-0.06 (0.04)
mother's years of education	0.03 (0.01)*	-0.03 (0.02)	0.02 (0.01)*
nr of failed grades	-0.09 (0.05)**	0.04 (0.07)	0.04 (0.04)
CAT score	0.01 (0.04)	0.11 (0.06)**	0.01 (0.03)
Average of all past marks	2.26 (0.73)*	7.44 (1.43)*	0.65 (0.67)
Average of last 4 quarters	3.43 (0.47)*	2.75 (0.67)*	1.99 (0.38)*
(controls for year-dummies)	yes	yes	yes
Nr of observations (Middle School, Fall)	1,886	1,396	1,367
Nr of observations (All school levels)	6,720	3,335	6,904

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

Table A.3.4: Unobserved heterogeneity of marks and expectations in Middle school

Standard-deviation of unobservables ($b = Parents, Students$):

	Math			Reading/English		
	$\sigma_u = e^{z_{i,t}\gamma_u}$	$\sigma_u^{p,b}(z_{i,t}) = e^{z_{i,t}\gamma_u^{p,b}}$		$\sigma_u = e^{z_{i,t}\gamma_u}$	$\sigma_u^{p,b}(z_{i,t}) = e^{z_{i,t}\gamma_u^{p,b}}$	
Coefficients:	γ_u	$\gamma_u^{p,Parents}$	$\gamma_u^{p,Students}$	γ_u	$\gamma_u^{p,Parents}$	$\gamma_u^{p,Students}$
Constant	0.23 (0.13)**	-0.27 (0.11)*	-0.09 (0.14)	-0.09 (0.06)	-0.07 (0.12)	-0.19 (0.07)*
black race	-0.08 (0.07)	0.02 (0.06)	-0.04 (0.07)	-0.07 (0.05)	0.09 (0.07)	0.07 (0.06)
female	0.05 (0.06)	-0.04 (0.06)	0.01 (0.07)	0.11 (0.05)*	-0.01 (0.06)	-0.08 (0.05)
mom's educ.	-0.02 (0.02)	-0.01 (0.01)	-0.02 (0.02)	-0.02 (0.01)**	-0.02 (0.02)	-0.05 (0.01)*

Correlation of unobservables with marks:

$$\rho_u^{p,b}(z_{i,t}) = 2\left(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}\right) - 1$$

Coefficients:		$\theta_u^{p,Parents}$	$\theta_u^{p,Students}$		$\theta_u^{p,Parents}$	$\theta_u^{p,Students}$
Constant		0.04 (0.16)	0.53 (0.23)*		0.45 (0.21)*	0.02 (0.16)
black race		0.08 (0.13)	-0.02 (0.17)		-0.09 (0.16)	0.08 (0.14)
female		0.01 (0.12)	0.13 (0.18)		0 (0.16)	0.25 (0.13)**
mom's educ.		0.01 (0.03)	0 (0.04)		-0.01 (0.04)	0.06 (0.03)*

Standard-deviations in (), * and ** denote 5% and 10% statistical significance.

7.4.3 High school (Fall quarter)

Table A.3.5: Marks and expectations in High school

	Fall Marks	**Spring Marks**	**Parents (Fall)**	**Students (Spring)**
	$m(z_{i,t})$	$m(z_{i,t})$	$mp(z_{i,t})^{Parents}$	$mp(z_{i,t})^{Students}$
Mean of latent marks/ expectations:				
Exogenous variables		**Math:**		
Constant	0.06 (0.32)	-0.03 (0.36)	1.34 (0.3)*	1.33 (0.36)*
black race	-0.2 (0.09)*	-0.01 (0.11)	0.21 (0.09)*	0.04 (0.12)
female	0.17 (0.08)*	-0.02 (0.11)	0.11 (0.08)	-0.18 (0.1)**
Student's month of birth	0.01 (0.01)	0 (0.01)	-0.02 (0.01)	-0.01 (0.01)
age of mother at birth (× 10)	-0.07 (0.08)	0.01 (0.1)	-0.2 (0.08)*	0.01 (0.09)
mother's education	0.03 (0.02)	0.01 (0.03)	-0.08 (0.02)*	-0.01 (0.02)
nr of failed grades	-0.04 (0.07)	-0.06 (0.09)	0.05 (0.07)	-0.06 (0.1)
CAT score	-0.15 (0.07)*	-0.11 (0.08)	0.12 (0.07)**	0 (0.08)
Average of all past marks	3.81 (1.16)*	2.82 (1.17)*	3.93 (1.09)*	1.47 (1.35)
Average of last 4 quarters	7.41 (0.86)*	10 (1.19)*	4.64 (0.62)*	6.13 (0.92)*
(controls for year-dummies)	yes	yes	yes	yes
Observations (High School)	769	931	619	606
Observations (All school levels)	5,020	6,227	3,533	4,476
Exogenous variables		**English/Reading:**		
Constant	-0.47 (0.66)	0.61 (0.33)**	1.49 (0.5)*	1.46 (0.26)*
black race	-0.17 (0.22)	-0.1 (0.12)	0.04 (0.18)	0.1 (0.09)
female	0.07 (0.17)	0.34 (0.1)*	0.25 (0.15)**	0.09 (0.08)
Student's month of birth	0.03 (0.02)	0.01 (0.01)	-0.01 (0.02)	-0.03 (0.01)*
age of mother at birth (× 10)	0.03 (0.13)	0.13 (0.08)**	-0.03 (0.12)	-0.07 (0.07)
mother's education	0.01 (0.04)	0.01 (0.03)	-0.02 (0.04)	0.03 (0.02)**
nr of failed grades	-0.29 (0.14)*	-0.25 (0.07)*	0.04 (0.11)	0.03 (0.07)
CAT score	-0.02 (0.14)	-0.05 (0.07)	-0.02 (0.11)	0.08 (0.06)
Average of all past marks	2.34 (2.37)	0.62 (1.26)	5.69 (1.83)*	0.3 (1.06)
Average of last 4 quarters	11.01 (1.54)*	4.24 (0.86)*	4.73 (0.93)*	2.93 (0.58)*
(controls for year-dummies)	yes	yes	yes	yes
Observations (High School)	626	705	442	446
Observations (All school levels)	4,870	5,921	3,335	4,485

Standard-deviations in (). * significant at the 5% level, ** significant at the 10% level

Table A.3.6: Unobserved heterogeneity of marks and expectations in High school

Standard-deviation of unobservables ($b = Parents, Students$):

	Math			Reading/English		
	$\sigma_u = e^{z_{i,t}\gamma_u}$	$\sigma_u^{p,b}(z_{i,t}) = \exp(z_{i,t}\gamma_u^{p,b})$		$\sigma_u = e^{z_{i,t}\gamma_u}$	$\sigma_u^{p,b}(z_{i,t}) = \exp(z_{i,t}\gamma_u^{p,b})$	
Coefficients:	γ_u	$\gamma_u^{p,Parents}$	$\gamma_u^{p,Students}$	γ_u	$\gamma_u^{p,Parents}$	$\gamma_u^{p,Students}$
Constant	0.07 (0.15)	-0.15 (0.15)	-0.25 (0.19)	-0.09 (0.13)	-0.15 (0.21)	-0.8 (0.16)*
black race	-0.08 (0.07)	0.09 (0.09)	-0.18 (0.12)	-0.25 (0.1)*	-0.01 (0.14)	0.47 (0.13)*
female	-0.04 (0.08)	-0.01 (0.09)	0 (0.12)	0.03 (0.09)	-0.07 (0.13)	-0.02 (0.1)
mom's education	-0.02 (0.02)	-0.05 (0.02)*	0 (0.03)	0 (0.02)	-0.03 (0.03)	-0.02 (0.02)

Correlation of unobservables with marks:

	Math		Reading/English	
	$\rho_u^{p,b}(z_{i,t}) = 2(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}) - 1$		$\rho_u^{p,b}(z_{i,t}) = 2(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}) - 1$	
Coefficients:	$\theta_u^{p,Parents}$	$\theta_u^{p,Students}$	$\theta_u^{p,Parents}$	$\theta_u^{p,Students}$
Constant	0.57 (0.28)*	0.53 (0.33)	0.61 (0.52)	0.29 (0.33)
black race	-0.1 (0.22)	-0.31 (0.25)	-0.72 (0.45)	-0.22 (0.28)
female	-0.02 (0.21)	0.03 (0.25)	0.03 (0.34)	0.04 (0.24)
mom's education	-0.07 (0.05)	0.08 (0.06)	0.04 (0.1)	0.13 (0.06)*

Standard-deviations in (), * and ** denote 5% and 10% statistical significance.

7.4.4 Elementary, Middle and High School (Spring quarter): Students' expectations and marks

Table A.3.7: Marks in the Spring quarter

	Marks		
School level	**Elementary**	**Middle**	**High school**
		$m(z_{i,t})$	
Mean of latent marks/ expectations:			
Exogenous variables	**Math (year-dummies controls included:**		
Constant	-0.04 (0.13)	-0.25 (0.26)	-0.03 (0.36)
black race	-0.02 (0.04)	0.05 (0.08)	-0.01 (0.11)
female	-0.01 (0.04)	0.07 (0.07)	-0.02 (0.11)
Student's month of birth	0 (0.01)	0 (0.01)	0 (0.01)
age of mom at birth ($\times$ 10)	0.04 (0.04)	0.08 (0.06)	0.01 (0.1)
mother's education	0 (0.01)	0.03 (0.02)**	0.01 (0.03)
nr of failed grades	0.03 (0.04)	0.03 (0.05)	-0.06 (0.09)
CAT score	0.06 (0.04)	-0.01 (0.05)	-0.11 (0.08)
Average of all past marks	-1.02 (0.63)	-0.05 (0.84)	2.82 (1.17)*
Average of last 4 quarters	12.72 (1.38)*	10.55 (1.22)*	10 (1.19)*
Observations (High Sch.)	3,198	2,098	931
Observations (All levels)		6,227	
Exogenous variables	**English/Reading (year-dummies controls included):**		
Constant	0.57 (0.02)*	0.12 (0.13)	0.61 (0.33)**
black race	0.31 (0.02)*	0.13 (0.04)*	-0.1 (0.12)
female	1.04 (0)*	0.06 (0.04)	0.34 (0.1)*
Student's month of birth	-0.05 (0.01)*	0.01 (0.01)*	0.01 (0.01)
age of mom at birth ($\times$ 10)	1.75 (0.01)*	-0.01 (0.03)	0.13 (0.08)**
mother's education	-0.21 (0)*	0 (0.01)	0.01 (0.03)
nr of failed grades	-1.07 (0)*	0.05 (0.03)	-0.25 (0.07)*
CAT score	-0.88 (0)*	-0.06 (0.03)*	-0.05 (0.07)
Average of all past marks	0.07 (0.01)*	1.41 (0.57)*	0.62 (1.26)
Average of last 4 quarters	-0.11 (0.01)*	8.64 (0.42)*	4.24 (0.86)*
(controls for year-dummies)	yes	yes	yes
Observations (High Sch.)	3,201	2,015	705
Observations (All levels)		5,921	

Table A.3.8: **Expectations in the Spring quarter (students)**

	Expectations		
School level	**Elementary**	**Middle**	**High school**
		$mp(z_{i,t})$	
Mean of latent marks/ expectations:			
Exogenous variables	**Math (year-dummies controls included:**		
Constant	2.9 (0.27)*	1.35 (0.22)*	1.33 (0.36)*
black race	0.12 (0.08)	0.16 (0.07)*	0.04 (0.12)
female	-0.05 (0.07)	-0.05 (0.06)	-0.18 (0.1)**
Student's month of birth	-0.01 (0.01)	0 (0.01)	-0.01 (0.01)
age of mom at birth (× 10)	-0.09 (0.07)	0.01 (0.06)	0.01 (0.09)
mother's education	-0.02 (0.02)	0 (0.02)	-0.01 (0.02)
nr of failed grades	-0.14 (0.07)*	0.07 (0.05)	-0.06 (0.1)
CAT score	-0.12 (0.08)	0 (0.05)	0 (0.08)
Average of all past marks	-2.54 (1.17)*	-0.57 (0.83)	1.47 (1.35)
Average of last 4 quarters	6.84 (1.19)*	7.27 (0.88)*	6.13 (0.92)*
Observations (High Sch.)	1,968	1,902	606
Observations (All levels)		4,476	
Exogenous variables	**English/Reading (year-dummies controls included):**		
Constant	2.64 (0.01)*	1.63 (0.13)*	1.46 (0.26)*
black race	0.06 (0.02)*	0.13 (0.05)*	0.1 (0.09)
female	0.08 (0.02)*	-0.05 (0.04)	0.09 (0.08)
Student's month of birth	0 (0)	0 (0.01)	-0.03 (0.01)*
age of mom at birth (× 10)	-0.03 (0.01)*	-0.11 (0.04)*	-0.07 (0.07)
mother's education	-0.01 (0)*	0 (0.01)	0.03 (0.02)**
nr of failed grades	-0.1 (0.01)*	-0.05 (0.03)	0.03 (0.07)
CAT score	-0.21 (0.01)*	-0.02 (0.03)	0.08 (0.06)
Average of all past marks	1.43 (0.03)*	0.18 (0.58)	0.3 (1.06)
Average of last 4 quarters	3.16 (1.50)*	5.59 (0.34)*	2.93 (0.58)*
(controls for year-dummies)	yes	yes	yes
Observations (High Sch.)	1,971	1,809	446
Observations (All levels)		4,485	

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

Table A.3.9: Unobserved heterogeneity of marks and expectations in the Spring quarter (Math)

Standard-deviation of unobservables:

	Marks:			**Students' expectations:**		
	Elementary	Middle	High	Elementary	Middle	High
		$\sigma_u = e^{z_{i,t}\gamma_u}$			$\sigma_u^{p,b}(z_{i,t}) = e^{z_{i,t}\gamma_u^{p,b}}$	
Coefficients:		γ_u			$\gamma_u^{p,Students}$	
Constant	-0.48 (0.12)*	0.03 (0.12)	0.07 (0.15)	0.2 (0.13)	-0.04 (0.13)	-0.25 (0.19)
black race	-0.08 (0.05)**	-0.21 (0.05)*	-0.08 (0.07)	0.14 (0.07)**	-0.21 (0.06)*	-0.18 (0.12)
female	-0.05 (0.05)	0 (0.05)	-0.04 (0.08)	-0.09 (0.07)	0.04 (0.06)	0 (0.12)
mom's education	-0.04 (0.01)*	-0.01 (0.01)	-0.02 (0.02)	-0.04 (0.02)*	-0.02 (0.01)	0 (0.03)

Correlation of unobservables with marks:

$$\rho_u^{p,b}(z_{i,t}) = 2(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}) - 1$$

Coefficients:					$\theta_u^{p,Students}$	
Constant				0.19 (0.16)	0.67 (0.21)*	0.53 (0.33)
black race				-0.16 (0.16)	-0.29 (0.16)**	-0.31 (0.25)
female				-0.06 (0.16)	0.16 (0.15)	0.03 (0.25)
mom's education				0.04 (0.04)	0.03 (0.04)	0.08 (0.06)

Standard-deviations in (), * and ** denote 5% and 10% statistical significance.

Table A.3.10: Unobserved heterogeneity of marks and expectations in the Spring quarter (Reading/English)

Standard-deviation of unobservables:

	Marks:			**Students' expectations:**		
	Elementary	Middle	High	Elementary	Middle	High
		$\sigma_u = e^{z_{i,t}\gamma_u}$			$\sigma_u^{p,b}(z_{i,t}) = e^{z_{i,t}\gamma_u^{p,b}}$	
Coefficients:		γ_u			$\gamma_u^{p,Students}$	
Constant	0.43 (0.01)*	-0.34 (0.06)*	-0.09 (0.13)	-0.04 (0.02)	-0.16 (0.06)*	-0.8 (0.16)*
black race	-0.23 (0.01)*	-0.05 (0.05)	-0.25 (0.1)*	0.06 (0.02)*	-0.07 (0.05)	0.47 (0.13)*
female	-0.28 (0.01)*	0.08 (0.05)**	0.03 (0.09)	-0.05 (0.03)	-0.03 (0.05)	-0.02 (0.1)
mom's education	-0.08 (0)*	-0.01 (0.01)	0 (0.02)	-0.03 (0)*	-0.05 (0.01)*	-0.02 (0.02)

Correlation of unobservables with marks:

$$\rho_u^{p,b}(z_{i,t}) = 2\left(\frac{\exp(z_{i,t}\theta_u^{p,b})}{1+\exp(z_{i,t}\theta_u^{p,b})}\right) - 1$$

Coefficients:					$\theta_u^{p,Students}$	
Constant				0.48 (0.01)*	0.79 (0.17)*	0.29 (0.33)
black race				0.27 (0.03)*	-0.17 (0.13)	-0.22 (0.28)
female				0.01 (0)	0.02 (0.13)	0.04 (0.24)
mom's education				0.27 (0)*	-0.05 (0.03)**	0.13 (0.06)*

Standard-deviations in (), * and ** denote 5% and 10% statistical significance.

7.4.5 AR coefficients for the Math and Reading/English marks plus the respondents' expectations

Table A.3.11: Parents and marks' AR coefficients (Fall quarter)

AR(1) coefficients: $\lambda^p(z_{i,t}) = 2(\frac{\exp(c_{\lambda^p} + z_{i,t}\theta_{\lambda^p})}{1+\exp(c_{\lambda^p} + z_{i,t}\theta_{\lambda^p})}) - 1$

	Parents (Math)			Parents (Reading/English)		
Exog. variables	Elementary	Middle	High school	Elementary	Middle	High school
Constant	0.17 (0.22)	0.69 (0.21)*	0.57 (0.28)*	0.06 (0.21)	0.17 (0.26)	0.87 (0.79)
black race	0.03 (0.19)	0.29 (0.19)	-0.34 (0.23)	-0.15 (0.18)	0.09 (0.23)	-0.44 (0.65)
female	-0.09 (0.19)	-0.29 (0.18)	0.24 (0.22)	0.15 (0.17)	-0.04 (0.2)	-0.02 (0.5)
mom's education	-0.01 (0.04)	-0.07 (0.04)**	-0.02 (0.05)	0.06 (0.04)	-0.01 (0.05)	-0.01 (0.11)
	Marks (Math)			Marks (Reading/English)		
Exog. variables	Elementary	Middle	High school	Elementary	Middle	High school
Constant	-0.24 (0.13)**	-0.2 (0.15)	-0.21 (0.22)	-0.3 (0.2)	-0.29 (0.18)	-0.11 (0.39)
black race	-0.16 (0.11)	0 (0.12)	0.14 (0.18)	0.09 (0.17)	0.06 (0.17)	0.25 (0.3)
female	0.03 (0.11)	0 (0.12)	-0.11 (0.17)	0.11 (0.15)	0.06 (0.15)	-0.28 (0.25)
mom's education	0.05 (0.02)*	-0.01 (0.03)	-0.05 (0.04)	0.03 (0.04)	0.01 (0.04)	0.01 (0.07)

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

Table A.3.12: Students and marks' AR coefficients (Fall quarter)

AR(1) coefficients: $\lambda^p(z_{i,t}) = 2(\frac{\exp(c_{\lambda^p} + z_{i,t}\theta_{\lambda^p})}{1+\exp(c_{\lambda^p} + z_{i,t}\theta_{\lambda^p})}) - 1$

	Students (Math)		Students (Reading/English)	
Explanatory variables	Elementary	Middle	Elementary	Middle
Constant	-0.19 (0.18)	0.32 (0.21)	-0.2 (0)*	0.44 (0.17)*
black race	0.36 (0.18)*	-0.08 (0.21)	0.34 (0.1)*	0.07 (0.15)
female	-0.09 (0.16)	-0.03 (0.19)	0.02 (0.08)	-0.28 (0.14)**
mother's education	-0.01 (0.04)	0.02 (0.05)	-0.01 (0.03)	-0.02 (0.03)
	Marks (Math)		Marks (Reading/English)	
Explanatory variables	Elementary	Middle	Elementary	Middle
Constant	0.3 (0.26)	0.2 (0.25)	-0.22 (0)*	-0.08 (0.09)
black race	-0.29 (0.26)	-0.09 (0.21)	-0.08 (0)*	0.07 (0.07)
female	-0.08 (0.24)	-0.07 (0.21)	0.05 (0.01)*	0.05 (0.07)
mother's education	-0.03 (0.07)	-0.01 (0.05)	0.01 (0)**	0.03 (0.02)

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

Table A.3.13: Students and marks' AR coefficients (Spring quarter)

AR(1) coefficients: $\lambda^p(z_{i,t}) = 2(\frac{\exp(c_{\lambda^p} + z_{i,t}\theta_{\lambda^p})}{1 + \exp(c_{\lambda^p} + z_{i,t}\theta_{\lambda^p})}) - 1$

	Students (Math)			Students (Reading/English)		
Exog. variables	Elementary	Middle	High school	Elementary	Middle	High school
Constant	0.24 (0.16)	0.13 (0.17)	0 (0.33)	0.1 (0.02)*	0.42 (0.15)*	0.47 (0.26)**
black race	0.02 (0.15)	0.22 (0.15)	-0.03 (0.27)	0.2 (0.04)*	-0.13 (0.13)	0.05 (0.27)
female	0.07 (0.15)	-0.1 (0.14)	-0.05 (0.26)	0.15 (0.06)*	-0.06 (0.12)	-0.31 (0.24)
mom's education	-0.03 (0.04)	0.02 (0.03)	0.05 (0.07)	-0.02 (0.01)**	0 (0.03)	0 (0.06)
	Marks (Math)			Marks (Reading/English)		
Exog. variables	Elementary	Middle	High school	Elementary	Middle	High school
Constant	-0.23 (0.11)*	0.16 (0.14)	-0.39 (0.26)	1.06 (0.02)*	0.03 (0.12)	0.6 (0.32)**
black race	-0.02 (0.1)	-0.19 (0.13)	0.34 (0.2)**	0.34 (0.01)*	0.01 (0.1)	0.02 (0.26)
female	0.08 (0.09)	-0.01 (0.11)	0.03 (0.22)	0.45 (0.01)*	0.12 (0.09)	-0.27 (0.24)
mom's education	-0.01 (0.03)	-0.02 (0.03)	-0.05 (0.04)	-0.87 (0)*	0 (0.02)	-0.01 (0.05)

Standard-deviations in (), * significant at the 5% level, ** significant at the 10% level

8 Bias and variance of the prediction errors: English/Reading

Table A.4.1: Bias and variance of expectations (Reading/English, Fall: Elementary sch., 82-87)

Variables / Demographic group	All	Black	Repeaters	Mothers w/ high school
correlation of parents and marks	0.21 (0.04)	0.20 (0.06)	0.11 (0.08)*	0.21 (0.06)
mean bias of parental expectations	1.57 (0.26)	1.77 (0.29)*	1.12 (0.24)*	1.50 (0.25)*
variance of parental expectations' errors	2.04 (0.63)	2.12 (0.68)	2.48 (0.86)	2.01 (0.66)
variance of parents - "rational" agent	0.72 (0.26)	0.78 (0.30)	1.03 (0.40)	0.69 (0.27)
correlation of students and marks	0.22 (0.09)	0.25 (0.10)	0.24 (0.11)	0.28 (0.11)*
mean bias of students' expectations	1.41 (0.40)	1.57 (0.44)*	1.50 (0.44)	1.19 (0.38)*
variance of students expectations' errors	2.92 (0.98)	2.90 (1.03)	3.18 (1.18)	2.76 (1.01)
variance of students - "rational" agent	1.56 (0.60)	1.53 (0.64)	1.66 (0.68)	1.39 (0.62)

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level.

Table A.4.2: Bias and variance of expectations (Reading/English, Fall - Middle school)

Variables / Demographic group	All	Black	Repeaters	Mothers w/ high school
correlation of parents and marks	0.19 (0.04)	0.15 (0.05)	0.13 (0.04)*	0.21 (0.05)
mean bias of parental expectations	0.53 (0.12)	0.58 (0.13)*	0.58 (0.13)**	0.47 (0.11)*
variance of parental expectations' errors	0.61 (0.26)	0.61 (0.25)	0.70 (0.29)	0.54 (0.25)
variance of parents - "rational" agent	0.77 (1.40)	0.82 (1.27)	0.96 (1.30)	0.79 (1.95)
correlation of students and marks	0.41 (0.05)	0.41 (0.05)	0.44 (0.05)	0.46 (0.05)*
mean bias of students' expectations	0.47 (0.11)	0.50 (0.12)*	0.55 (0.13)*	0.43 (0.10)*
variance of students expectations' errors	0.31 (0.15)	0.30 (0.14)	0.32 (0.15)	0.25 (0.13)*
variance of students - "rational" agent	0.02 (0.03)	0.03 (0.03)	0.00 (0.04)*	-0.01 (0.03)*

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level.

Table A.4.3: Bias and variance of expectations (Parents - Math, Fall, High school)

Variables / Demographic group	All	Black	Repeaters	Mothers w/ high school
correlation of parents and marks	0.13 (0.12)	-0.12 (0.13)*	0.15 (0.20)	0.17 (0.12)
mean bias of parents' expectations	0.52 (0.14)	0.61 (0.15)*	0.29 (0.18)	0.50 (0.14)
variance of parents expectations' errors	0.32 (0.17)	0.40 (0.18)**	0.38 (0.17)	0.30 (0.16)
variance of parents - "rational" agent	0.04 (0.05)	0.12 (0.06)**	0.02 (0.04)	0.04 (0.05)

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level.

Table A.4.4: Bias and variance of students' expectations (Math - Spring, High school)

Variables / Demographic group	All	Black	Repeaters	Mothers w/ high school
correlation of students and marks	0.36 (0.05)	0.33 (0.06)	0.22 (0.07)*	0.40 (0.05)*
mean bias of students' expectations	0.40 (0.12)	0.46 (0.14)*	0.52 (0.17)*	0.39 (0.12)*
variance of students expectations' errors	0.18 (0.13)	0.20 (0.14)	0.23 (0.17)	0.15 (0.12)*
variance of students - "rational" agent	0.02 (0.02)	0.03 (0.02)*	0.05 (0.03)	0.01 (0.01)

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level.

9 Bias and variance of students Spring prediction errors for all school levels: Math and Reading/English

Table A.4.5: Bias and variance of students' expectations in the Spring quarter (Math)

Variables Demographic group	All	Black	Repeaters	Mothers with H-S
Elementary school (1982-1987)				
correlation of students and marks	0.15 (0.05)	0.15 (0.06)	0.14 (0.05)	0.18 (0.06)
mean bias of students' expectations	1.52 (0.25)	1.76 (0.31)*	1.61 (0.29)	1.30 (0.22)*
variance of students expectations' errors	2.48 (1.29)	2.67 (1.43)	1.78 (0.59)	2.19 (1.19)
variance of students - "rational" agent	2.01 (1.15)	2.25 (1.32)	1.33 (0.47)*	1.76 (1.05)*
Middle school				
correlation of students and marks	0.35 (0.03)	0.31 (0.03)	0.33 (0.03)	0.38 (0.04)*
mean bias of students' expectations	0.79 (0.12)	0.85 (0.13)*	0.91 (0.15)*	0.73 (0.12)*
variance of students expectations' errors	0.81 (0.22)	0.79 (0.22)	0.95 (0.26)*	0.72 (0.21)*
variance of students - "rational" agent	0.13 (0.04)	0.19 (0.06)**	0.19 (0.06)*	0.10 (0.04)
High school				
correlation of students and marks	0.31 (0.06)	0.38 (0.06)**	0.35 (0.09)	0.28 (0.07)
mean bias of students' expectations	0.56 (0.18)	0.61 (0.20)	0.55 (0.19)	0.53 (0.18)*
variance of students expectations' errors	0.29 (0.24)	0.30 (0.26)	0.29 (0.25)	0.26 (0.26)
variance of students - "rational" agent	0.00 (0.03)	-0.02 (0.05)	-0.01 (0.06)	0.02 (0.03)

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level

Table A.4.6: Bias and variance of students' English/Reading expectations: Spring

Variables Demographic group	All	Black	Repeaters	Moms with H-S
Elementary school 1982-87				
correlation of students and marks	0.13 (0.05)	0.10 (0.06)	0.15 (0.06)	0.12 (0.06)
mean bias of students' expectations	2.38 (0.33)	2.63 (0.36)*	2.93 (0.48)*	2.03 (0.28)
variance of students expectations' errors	4.41 (1.20)	4.53 (1.28)	5.23 (1.61)	4.00 (1.10)
variance of students - "rational" agent	3.48 (0.98)	3.61 (1.05)	4.18 (1.36)	3.08 (0.88)
Middle school				
correlation of students and marks	0.22 (0.03)	0.19 (0.04)	0.22 (0.04)	0.22 (0.04)
mean bias of students' expectations	0.88 (0.18)	0.94 (0.19)*	0.95 (0.19)**	0.86 (0.18)
variance of students expectations' errors	1.35 (0.53)	1.39 (0.56)	1.40 (0.51)	1.32 (0.56)
variance of students - "rational" agent	0.40 (0.17)	0.43 (0.19)	0.37 (0.15)	0.40 (0.19)
High school				
correlation of students and marks	0.36 (0.05)	0.33 (0.06)	0.22 (0.07)*	0.40 (0.05)*
mean bias of students' expectations	0.40 (0.12)	0.46 (0.14)*	0.52 (0.17)*	0.39 (0.12)*
variance of students expectations' errors	0.18 (0.13)	0.20 (0.14)	0.23 (0.17)	0.15 (0.12)*
variance of students - "rational" agent	0.02 (0.02)	0.03 (0.02)*	0.05 (0.03)	0.01 (0.01)

Bootstrap standard-errors in () - 100 sample replications.

* significantly different from the "All" group at the 5% level, **significant at the 10% level

References

[1] Alexander, K., D. Entwisle and S. Dauber (2003), "On the Success of Failure: A Reassessment of the Effects of Retention in the Primary Grades," 2nd ed., Cambridge University Press.

[2] Alexander, Karl L.; and Entwisle, Doris R., 2004 -10-06, "The Beginning School Study, 1982-2002 (Log# 01293)", hdl: 1902.1/01293, http://id.thedata.org/hdl%3A1902.1%2F01293 Henry A. Murray Research Archive [distributor (DDI)].

[3] Madeira, C. (2019), "Learning your own ability", draft in R&R for the *Journal of Economic Dynamics and Control*.

Documentos de Trabajo **Banco Central de Chile**	**Working Papers** **Central Bank of Chile**
NÚMEROS ANTERIORES	PAST ISSUES
La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de Ch$500 si es dentro de Chile y US$12 si es fuera de Chile. Las solicitudes se pueden hacer por fax: +56 2 26702231 o a través del correo electrónico: bcch@bcentral.cl.	Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for order inside Chile the charge is Ch$500.) Orders can be placed by fax: +56 2 26702231 or by email: bcch@bcentral.cl.

DTBC – 919
Learning Your Own Ability
Carlos Madeira

DTBC – 918
Commodities Fundamental Model
Francisco Arroyo Marioli, Fernando Letelier

DTBC – 917
Firm Sorting, College Major, and the Gender Earnings Gap
Federico Huneeus, Conrad Miller, Christopher Neilson, Seth Zimmerman

DTBC – 916
Optimal Monetary Policy and Incomplete Information: Does the Real Exchange Matter?
Rodrigo Caputo, Felipe Leal

DTBC – 915
Estimating Shadow Policy Rates in a Small Open Economy and the Role of Foreign Factors
Jorge Fornero, Markus Kirchner, Carlos Molina

DTBC – 914
Measuring Small and Medium-Size Enterprises Contribution to Trade in Value Added: The case of Chile 2013-2016
Mario Marcel, Diego Vivanco

DTBC – 913
Toward a general framework for constructing and evaluating core inflation measures
Guillermo Carlomagno, Jorge Fornero, Andrés Sansone

DTBC – 912
Monetary Policy Press Releases: An International Comparison
Mario Gonzalez, Raul Cruz Tadle

DTBC – 911
The Credit Channel Through the Lens of a Semi-Structural Model
Francisco Arroyo Marioli, Juan Sebastian Becerra, Matias Solorza
DTBC – 910
Contracts, Firm Dynamics, and Aggregate Productivity
Bernabe Lopez-Martin, David Perez-Reyna

DTBC – 909
Optimal Spending and Saving Strategies for Commodity-Rich Countries
Alvaro Aguirre

DTBC – 908
Uncertainty, Risk, and Price-Setting: Evidence from CPI Microdata
Mario Canales, Bernabe Lopez-Martin

DTBC – 907
Earnings Inequality in Production Networks
Federico Huneeus, Kory Kroft, Kevin Lim

DTBC – 906
Price setting in Chile: Micro evidence from consumer on-line prices during the social outbreak and Covid-19
Jennifer Peña, Elvira Prades

DTBC – 905
Economic Growth at Risk: An Application to Chile
Nicolás Álvarez, Antonio Fernandois, Andrés Sagner

DOCUMENTOS DE TRABAJO Agosto 2021