Grade Repetition and Household Responses in a Low Income Setting

Tahir Andrabi[*], Ethan Matlin[†], Gabrielle Vasey[‡]

September 22, 2025

Abstract

Millions of children in low- and middle-income countries (LMICs) repeat a grade each year, yet little is known about how this affects students and their families. In many LMICs, grade repetition is common and discretionary, with decisions made informally by teachers rather than based on standardized rules. We study the academic and behavioral consequences of grade repetition in Pakistan, using matched panel data linking students, parents, teachers, and schools. Our empirical strategy exploits variation in retention decisions arising from differences in teacher thresholds and noise in perceived ability, controlling flexibly for prior achievement and student characteristics. We find that being retained lowers test scores by 0.27 to 0.44 standard deviations and increases dropout by 7 percentage points. Parents respond by revising downward their beliefs and reducing both short- and long-term educational investments. Retained students become less confident in the value of academic effort. In contrast, we find little evidence that teachers treat repeaters differently after the retention decision. These results suggest that retention acts as a salient information signal to parents, with lasting effects on household behavior.

Keywords: Education Policy, Achievement, Grade Retention, Parental Investments.

[*]Pomona College
[†]Harvard University
[‡]Concordia University

1 Introduction

Grade repetition is a common but understudied feature of education systems in low- and middle-income countries (LMICs). Each year, approximately 24 million children repeat at least one grade during primary school, and repetition rates in LMICs are often five to ten times higher than in high-income countries (UNESCO 2024).[1] Existing evidence on the impacts of retention comes primarily from high-income countries, especially the United States, where retention is governed by automatic promotion policies and standardized testing (e.g., Jacob and Lefgren, 2004; Eren et al., 2018; Figlio and Özek, 2019). In contrast, grade repetition in LMICs often reflects informal, subjective judgments by teachers, and its implications for students, households, and schools remain poorly understood.

This paper examines the academic consequences of grade repetition and the behavioral responses it triggers among students, parents, and teachers. We study these dynamics in Pakistan, where promotion decisions are made at the discretion of individual teachers and are not governed by centralized rules or standardized thresholds (Chohan and Qadir, 2011).

Using matched panel data from the Learning and Educational Achievement in Punjab Schools (LEAPS) project, we estimate the effects of being retained on test scores, dropout, beliefs, and educational investments. Our rich data link students to their households, teachers, and schools, allowing us to trace how each of these agents responds to a child not being promoted. We find that parents revise downward their expectations and investments, retained students become more discouraged about the returns to effort, and schools do not significantly adjust their behavior. Overall, grade repetition lowers academic performance: repeaters score 0.27 to 0.44 standard deviations lower across three core subjects and are 7 percentage points more likely to drop out than their peers.

To guide our empirical strategy, we develop a simple conceptual framework in which teachers observe a noisy signal of student ability and retain students whose perceived ability falls below a subjective threshold. In Pakistan, where there are no formal guidelines

[1] Data from 2022 http://data.uis.unesco.org.

for grade promotion in Grade 4, these thresholds vary widely across teachers and schools, and retention decisions are made based on informal criteria and teacher discretion. Survey evidence confirms that teachers play a central role in these decisions and rely heavily on subjective evaluations of academic performance.

This setting generates two sources of identifying variation. First, promotion thresholds differ across teachers. Second, within a given classroom, measurement error in perceived ability introduces quasi-random variation in who is retained. We use detailed controls for lagged test scores, teacher ratings, and student demographics to isolate residual variation that reflects these sources. Our main specification includes teacher fixed effects, allowing us to compare students exposed to different thresholds and account for systematic differences in promotion standards.

We find that grade repetition leads to large and persistent declines in academic performance. In the year following retention, repeaters score 0.27 to 0.44 standard deviations lower in math, English, and Urdu, and are 7 percentage points more likely to drop out, relative to similar peers. These effects do not vary meaningfully by gender and persist for at least two years.

To understand the mechanisms behind these academic losses, we examine how households, students, and teachers respond to the negative signal of retention. Parents substantially revise downward their expectations, rate their child as less capable, and reduce both short- and long-run educational investments. Students become less likely to believe that effort in school leads to improved life outcomes, suggesting a decline in academic motivation. In contrast, we find little evidence that teachers treat repeaters differently after the retention decision, either in terms of evaluations or classroom behavior. These findings suggest that grade retention provides new information to parents, shifting their beliefs and investment decisions in ways that may exacerbate preexisting disadvantages.

This paper makes three contributions. First, we provide new evidence on the consequences of grade repetition in a low-income country where promotion decisions are informal and teacher-driven. Much of the existing literature on grade retention focuses on high-income settings where retention is governed by centralized policies or score-based thresholds (e.g., Eren et al., 2018; Figlio and Özek, 2019; Jacob and Lefgren, 2004, 2009; Eide and Showalter, 2001; Borghesan et al., 2022). In contrast, we study a setting where retention is common, unregulated, and highly discretionary. The Pakistani context, with widespread private schooling, substantial teacher autonomy, and large within-grade age variation, resembles many other LMICs, suggesting broader relevance.

Second, we contribute to a growing literature on parental belief formation and household investment in education (Becker and Tomes, 1976; Tomes, 1981; Behrman et al., 1994; Dizon-Ross, 2019; Gan, 2021; Bergman, 2021). Existing work has shown that providing parents with new information can shift expectations and behaviors. We show that organic signals, such as a child being held back, can have similarly powerful effects, prompting parents to revise beliefs and reduce investments. These findings highlight how institutional actions, even when informal, can shape household decision-making in persistent ways.

Third, we situate our findings within a growing body of work on grade repetition in LMICs. Studies from Brazil, Uruguay, Colombia, and Mexico have used policy changes and quasi-experimental designs to estimate the effects of retention on academic achievement, grade progression, and dropout (Gomes-Neto and Hanushek, 1994; Manacorda, 2012; Ferreira Sequeda et al., 2018; Cabrera-Hernandez, 2022). In Sub-Saharan Africa, work from Senegal finds that repetition increases dropout risk (Glick and Sahn, 2010), and studies from China and Pakistan highlight the discretionary nature of promotion decisions (Hu and Hannum, 2020; King et al., 2016). While these studies focus primarily on long-run academic outcomes, few are able to trace how different actors interpret and respond to the retention decision. Our matched panel data allow us to examine these behavioral dynamics across students, parents, and schools, and to highlight belief formation as a key mechanism through which early academic setbacks may persist.

The paper proceeds as follows. Section 2 develops a conceptual framework to motivate our empirical strategy and highlight the sources of identifying variation. Section 3 describes the institutional setting and the LEAPS data. Section 4 outlines our empirical strategy and presents descriptive evidence on retention decisions. Section 5 reports the main results, including impacts on academic outcomes, household responses, student beliefs, and teacher behavior. Section 6 discusses robustness checks and threats to identification. Section 7 concludes.

2 Conceptual Framework

We develop a simple model to motivate our empirical strategy and highlight the sources of identifying variation. The model is grounded in the institutional setting of Pakistan, where there are no centralized criteria for grade promotion, and teachers decide whether to retain students based on their own assessment of student readiness.[2] This is supported by our survey data: 72% of school principals reported that the classroom teacher played a very important role and 25% an important role in the decision to retain a student. Moreover, 98% of principals cited "academic performance" as the primary basis for the retention decision, though without reference to any consistent benchmark. These responses suggest that teachers apply their own implicit thresholds and rely on subjective evaluations of student ability when making retention decisions.

Let each student i in period t have a latent ability θ_{it}, which is not directly observed by the teacher. Instead, teacher j observes a noisy signal:

$$s_{itj} = \theta_{it} + \varepsilon_{itj},$$

where ε_{itj} is mean-zero measurement error, capturing idiosyncratic factors or noise in the

[2] In government schools in Pakistan, automatic promotion applies through Grade 3. In Grade 4, there is no centrally determined grade retention policy. Standardized public examinations, such as the Punjab Examination Commission (PEC) exam in Grade 8, begin to influence promotion decisions in government schools at that stage. In contrast, private schools are largely unregulated and continue to rely on teacher or school discretion even in later grades.

teacher's perception. The teacher compares this signal to a teacher-specific threshold τ_j.[3] The student is retained if their signal falls below the threshold:

$$R_{itj} = \mathbf{1}(s_{itj} < \tau_j).$$

This model implies that the probability of being retained depends on both the student's latent ability and the teacher's threshold. Importantly, τ_j may vary across teachers and schools,[4] and ε_{itj} introduces within-school variation. These two components, heterogeneity in thresholds and idiosyncratic noise, generate identifying variation that allows us to estimate the causal effects of retention.

We examine the effects of grade retention on a wide set of outcomes measured in the following academic year ($t+1$). These include academic achievement (test scores), household investments in the child's education, and beliefs and attitudes reported by students, parents, and teachers. Rather than modeling the production function for each outcome explicitly, we take a reduced-form approach, and estimate whether being retained shifts these outcomes on average. This approach allows us to remain agnostic about the specific mechanisms, while providing empirical evidence on how households, students, and schools respond to the negative signal of retention.[5]

To identify these effects, we control flexibly for observable proxies for θ_{it}, including multiple lags of test scores and teacher ratings. Conditional on these controls, variation in R_{itj} arises from differences in τ_j and from ε_{itj}, both of which we argue are plausibly orthogonal to potential outcomes. Our main specification includes teacher fixed effects to absorb τ_j.

The model also highlights key threats to identification. First, teachers may select students to retain based on unobserved characteristics correlated with outcomes, violating the

[3] This threshold may change over time, however we are only using one year of data so we will drop the t subscript for simplicity.

[4] Most schools in our sample only have one classroom of students in the relevant grade, so teacher and school groups are interchangeable.

[5] We are also abstract from any notion of peer effects or optimal retention in this paper. It is possible that the class composition created by the retention decisions impacts the students' outcomes, however this is beyond the scope of this paper.

assumption that ε_{itj} is random noise. Second, the signal s_{itj} may itself be systematically biased by observable traits (e.g., gender, socioeconomic status), in which case retention is endogenous. We address these threats through rich covariate controls, balance checks, placebo tests, and a sensitivity analysis using the method proposed by Oster (2019), all of which are discussed in Section 6.

3 Data and Setting

3.1 Description of Data

We use data from the Learning and Education Achievement in Punjab Schools (LEAPS) project, a five-year survey of schools, school-children, and their households in the Attock, Faisalabad, and Rahim Yar Khan districts of Punjab province[6] conducted between 2004-2011 Surveyors randomly sampled 112 villages with at least one private school from the three districts.[7] Within each village, they administered questionnaires to students, teachers, head teachers, and households.

Surveyors also proctored tests in three subjects (Urdu, English, and Math) to assess student knowledge and learning. These tests were designed to cover mostly basic topics typically learned in lower grades.[8] The Urdu and English tests covered letters, word-recognition, sentence construction, and reading comprehension. The Math tests covered counting, addition, subtraction, multiplication, division, fractions, and word problems. Tests were scored using item response theory and normalized around zero using the first year score distribution (future years are in units standard deviations of the first year score distribution). Teachers did not receive results of these tests, therefore could not incorporate test scores into their

[6]Punjab is the most populous of Pakistan's provinces, containing 56% of the population.

[7]Since, for the most part, only larger villages contain private schools, there exists a potential selection bias, however due to the increasing private school market, this sample is likely to be more representative of the future (Das et al. (2012)).

[8]This decision in test construction was made to reflect students' lagged learning (rather than trying to assess where grade-level knowledge).

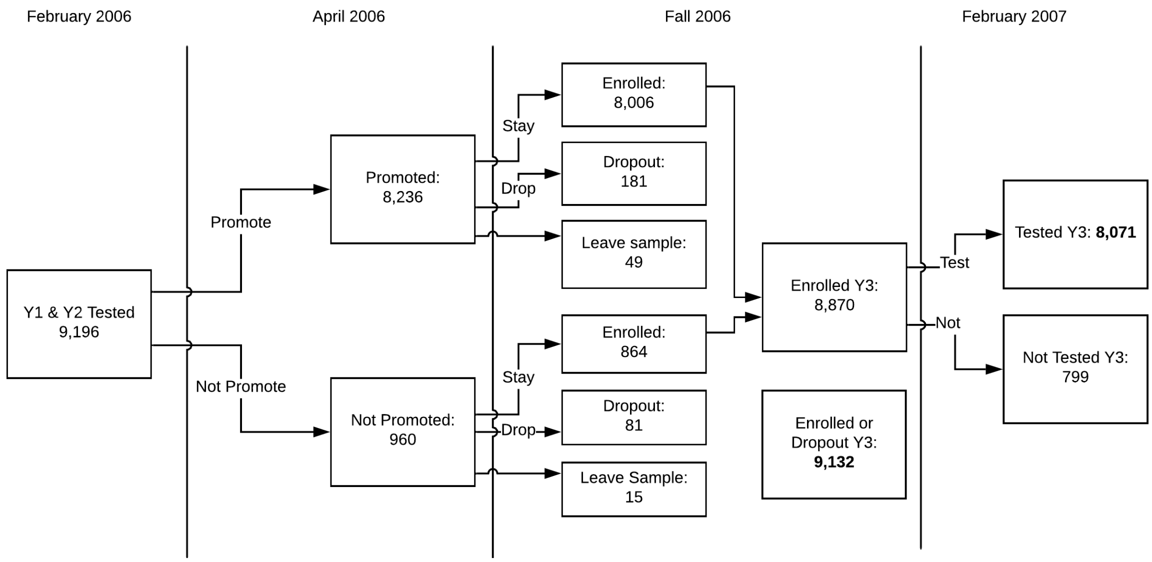
promotion/retention decisions.

In addition to the test score and enrollment data, we have rich demographic and survey data. For all students, we know their sex, age, and mother's education level. For a subsample of students, we also have data on their height and weight, and their feelings and attitudes towards school. From a matched household survey given to a subsample of students, we also have data on parental perceptions of their child, an asset-based wealth index, detailed educational expenditures, assessments of their child's teacher quality, and expectations of their child's ability. Finally, teachers were surveyed and answered the following question about a subsample of students in their class: "On a scale of 1 to 10, how good would you say that this student is in his/her studies?"

In the first year of the survey, 13,735 students in 804 schools which offered third grade instruction were surveyed. Of these schools, 485 (60%) were government schools and 319 private (40%).[9] At sampled schools, every 3rd grade student was sampled and we follow these initial students over time without adding new students to the sample (see Figure 1). Our analysis focuses on grade retention that occurs after the second year of the survey, which is 4th grade for the majority of the students. The outcomes in our analysis are variables from the third year of the survey.

[9]More specifically, there were 303 private schools and 16 NGO/Trust schools, however we group these schools together for our analysis.

Figure 1: Evolution of the sample



Note: There are 9,196 students between 7.5 and 15.5 years old in year 2 (4th graders are usually 9-10) with non-missing promotion status and test scores for both years 1 and 2. After year 2 (baseline), the teacher/school decided on each student's promotion status: should they be demoted to a lower grade, retained in their current grade, or promoted to the next grade? Promoted means the student advanced to the next grade while not promoted means the student was either retained in the current grade or demoted. 8,236 are promoted of which 8,006 remain enrolled and 181 dropout. 960 are not promoted of which 864 remain enrolled and 81 dropout. Thus, in year 3, there are 8,870 total enrolled students. Of these, 8,071 students are tested in year 3.

We primarily use four subsamples of the data. Sample 1, which is used for our results on academic outcomes, the subsamples are students who were tested in all three years (N=8,071)[10]. Sample 2 includes students who were tested all three years *and* whose teachers were surveyed about their performance (N=5,529), whereas Sample 3 includes students who were tested in all three years *and* matched to households given a more detailed household survey (N=698). Sample 4, used for the analysis of mechanism and household responses, uses data on students who were tested in the first two years and matched to household given the more detailed household survey (N=741). Table 1 presents baseline characteristics of

[10]This number is lower since some students were either new to the sample in the second year of the survey or were absent on the day of the test in year 1.

students for each of these samples. The patterns across each subsample are relatively similar: 7-9% of students are repeating a grade in the third year of the survey, just under half are female, and most students are 10 years old. Both teachers and parents slightly inflate student-ratings, however not overly so: teachers rate the average student 6/10 while parents rate their children 3.3/5. Finally, there appear to be no overwhelming differences between the subsamples (as we would expect since the subsamples are random).

Table 1: Baseline characteristics by sample

Variable	Child Sample	Teacher Sample	Parent Sample	Mech Sample
English score (Pre-)	0.306 (0.937)	0.363 (0.923)	0.252 (0.975)	0.238 (0.989)
Math score (Pre-)	0.254 (1.131)	0.304 (1.101)	0.304 (1.044)	0.287 (1.070)
Urdu score (Pre-)	0.380 (0.960)	0.420 (0.948)	0.373 (0.946)	0.352 (0.975)
Not Promoted	0.088 (0.284)	0.077 (0.266)	0.070 (0.256)	0.076 (0.264)
Female	0.467 (0.499)	0.470 (0.499)	0.446 (0.497)	0.441 (0.497)
Age	10.484 (1.394)	10.576 (1.404)	10.517 (1.409)	10.553 (1.488)
Mom Educated	0.277 (0.448)	0.355 (0.478)	0.287 (0.452)	0.265 (0.441)
Rating by teacher	- (-)	6.098 (2.323)	6.255 (2.354)	6.267 (2.377)
Parental Perception of Whether Child is Hard Working	- (-)	- (-)	3.304 (0.622)	3.294 (0.632)
Parental Perception of Child Intelligence	- (-)	- (-)	3.340 (0.621)	3.333 (0.625)
Inteligence Rating	-	-	-	3.333 (0.625)
Hardworking Rating	-	-	-	3.294 (0.632)
School Performance Rating	-	-	-	3.296 (0.656)
Educational Expenditures (short-run)	-	-	-	133.142 (100.163)
Lagged Educational Expenditures (long-run)	-	-	-	916.161 (413.651)
Family Help (any)	-	-	-	0.340 (0.474)
Hours of help with studies from family (median)	-	-	-	2.452 (3.887)
Hours reading or telling stories (median)	-	-	-	0.000 (0.532)
Any tutoring	-	-	-	0.182 (0.386)
Hrs. Tutoring	-	-	-	0.000 (4.749)
Know name of child's teacher	-	-	-	0.565 (0.496)
Met child's teacher	-	-	-	0.591 (0.492)
Look HW	-	-	-	0.560 (0.497)
Observations	8071	5529	698	741

Note: Entries are means with standard deviations in parentheses. Dashes indicate not available for that sample. "Observations" reports the number of units in each column. Column (1) includes all students with three years of test scores. Column (2) includes all students with three years of test scores whose teacher also rated their performance. Column (3) includes all students with three years of test scores matched to households that completed a more detailed household survey. Column (4) includes all students with two years of test scores matched to households that completed a more detailed household survey.

3.2 Institutional Setting: Grade Repetition in Pakistan

The setting for this paper is very different than in high-income countries like the U.S. Education levels in our sample are in general low: 72% of mothers in our sample did not complete primary school and 64% received no formal education. Mothers with no formal education spend ~ 0 minutes on a typical day for child educational needs at home. Learning levels for students are also low: by the end of 3rd grade, just over 50 percent of children have mastered the Mathematics curriculum for 1st grade (Andrabi et al. (2007)).

Schools in this setting also look very different than those in high income countries on average. Classes and schools are relatively small: the median class has 13 students and the median school has 119 students (across multiple grades). Classrooms have large age-ranges of students: the average classroom has a three to four year gap between the youngest and oldest student. There is a large low-cost private school sector, making up 40% of the schools in our sample and costing only a dime a day.

Grade repetition after fourth grade is common ($\sim 9\%$ in our sample), and teachers are mainly responsible for deciding who is retained. Although Pakistan does have automatic promotion at government schools until 3rd grade, our sample of students who would move from 4th to 5th grade aren't subject to this policy (Chohan and Qadir (2011). This differs from high-income countries, which in general have lower repetition rates (Eisemon (1997)).[11]

4 Identification and Empirical Strategy

4.1 Conceptual Sources of Variation

Our empirical strategy is motivated by the simple conceptual framework presented in Section 2, in which each teacher observes a noisy signal of student ability and retains students whose perceived ability falls below an implicit threshold. Because there is no formal grade retention policy for Grade 4 students in Pakistan, retention decisions vary substantially across schools and teachers. We leverage two sources of identifying variation:

1. Across-school variation in thresholds (τ_j): different schools apply different implicit

[11] Countries with clearer and properly enforced standards for repetition seem to have lower repetition rates than those with no or loosely-enforced state-mandated criteria for repetition. Repetition rates are highest in Sub-Saharan Africa (approximately 22% of primary students repeat a grade at any given time) and also high in the Middle East and North Africa (12%), however in developed countries only 1-5% ofstudents repeat a grade at any given time ((Eisemon (1997))).

standards for promotion.[12]

2. Within-school variation due to noise (ε_{ij}): among students with similar test scores and observed characteristics, retention decisions differ due to subjective perceptions or unobserved shocks.

4.2 Descriptive Evidence on Retention Decisions

The next step is to show that the conceptual framework, and therefore our identification strategy, is supported by the data. Table 2 confirms that, prior to the retention decision, students who would go on to be retained were scoring lower on average across all three tested subjects. This supports the assumption that retention decisions are based, at least in part, on student ability. Table 2 also shows that teachers rated the students they would eventually retain significantly lower than their peers. However, these averages mask substantial heterogeneity. Figure 2 plots the distribution of pre-retention average test scores for retained and promoted students. The figure shows clear overlap between the two groups. If teachers are making retention decisions based on ability, this overlap implies either that teacher-specific thresholds vary considerably or that the noise in the signal has high variance.

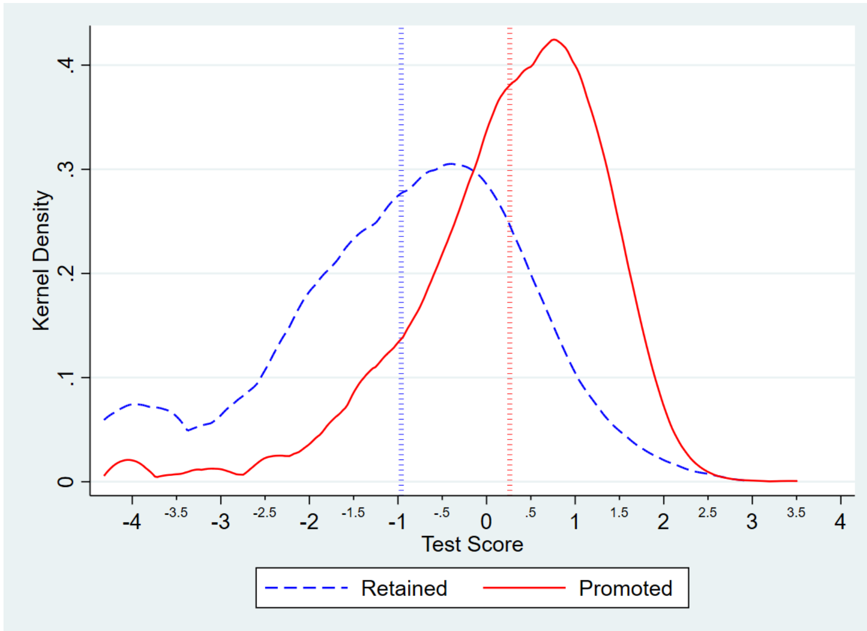
Table 2: Differences in test scores and teacher ratings before the retention decision

	Promoted	**Not Promoted**	**Difference**	**Total N**
English score	0.367 (0.893)	-0.532 (1.111)	-0.900***	9196
Math score	0.338 (1.051)	-0.910 (1.410)	-1.248***	9196
Urdu score	0.440 (0.914)	-0.581 (1.173)	-1.021***	9196
Rating by teacher	6.190 (2.264)	4.373 (2.422)	-1.817***	6269

Note: Entries in first two columns are means with standard deviations in parentheses. All variables are from year 2 (note that the retention decision happens between year 2 and year 3). The sample for the first three rows are all students with test scores in year 1 and year 2. The sample for the fourth row is all students with test scores in year 1 and year 2 and a teacher rating in year 2. In the third column, *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

[12] While the cross-teacher variation may sound reminiscent of a judge IV design, we cannot use that design in this setting for two primary reasons. First, the monotonicity assumption is violated: teachers are unlikely to have a common ranking for all students and differ only due to their cutoff. The lack of consistency within-teacher is evidence for this. Second, the exclusion restriction is violated: while judges affect outcomes only through sentencing, teachers also teach students and may affect learning trajectory.

Figure 2: Densities of test scores before promotion decision

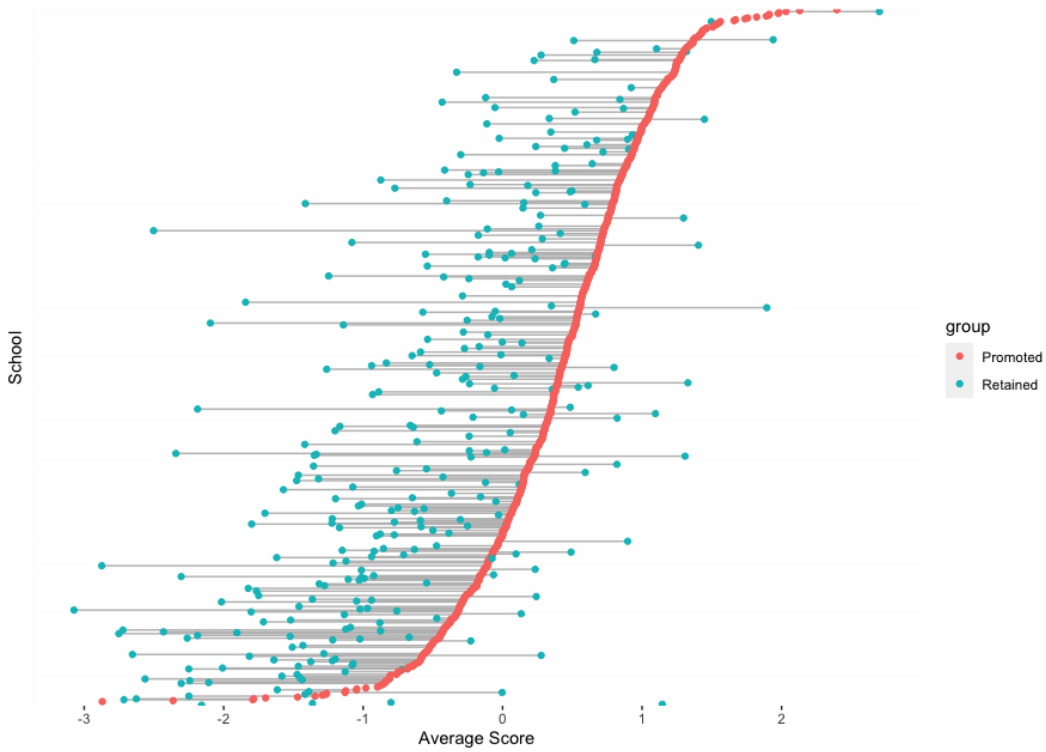


Note: Kernel density estimation of the distribution of math test scores for retained students (in blue) and promoted students (in red). The vertical lines represent the mean test score for each group. The test scores are from year 2 (note that the retention decision happens between year 2 and year 3).

To understand the impact of different teacher retention cutoffs, we look at differences across classrooms. Figure 3 plots two dots for each classroom: a red dot for the average test score of students who are promoted and a blue dot for the average test score of students who are retained. Test scores are all from the second year of the survey, before the retention decision. From this figure we can see that there is a large variation in teacher standards. The average test score of students who are retained at the top schools is higher than the average test scores of students who are promoted at lower schools. Further, in a few schools the average score of the retained students is higher (to the right) of the average score of the promoted students. We interpret this as evidence of measurement error.

To quantify the patterns suggested by the previous plots, we estimate a series of linear probability models (LPMs) where the dependent variable is an indicator for whether the student was retained. We report the R^2 and Adjusted R^2 for several of these LPMs in Table 3. Each of the four models contains flexible specifications of two lags of test scores. In a

Figure 3: Within school test scores before promotion decision



Note: For each school in the sample, two averages are computed: (1) the average test scores of students who are promoted (red) and (2) the average test scores of students who are retained (blue). These two averages are connected with a gray line. Schools who either promote or retain all students are included.

different setting with a centralized retention cutoff decision, whether a student repeats or not can be predicted with a high level of precision using the test score as an explanatory variable (or any other measures used by the school in the retention decision). In our setting, nearly all schools state that academic issues are the main reason for retention, yet the R^2 values in the first column of Table 3 are all less than 0.2.

The first model, Child Characteristics, also includes the gender and age of the student, as well as their mother's education.[13] The second and third model include the teacher's rating of the child and the parent's rating of the child respectively. Finally, the fourth specification combines all of the above information. Looking at the first column of R^2 values, we see that we are able to explain at most 20% of the variation in the retention decision. We then include teacher fixed effects, and present the R^2 and the Adjusted R^2 in the third and fourth column. The R^2 values get substantially higher, however it is important to also consider the Adjusted R^2, as we are including many fixed effects with a at times very small sample.

Table 3: Predicting retention with observables and teacher fixed effects

Specification	R^2	Adj. R^2	R^2 (FE)	Adj. R^2 (FE)	N
Child Characteristics	.14	.14	.36	.28	8071
Teacher Rating	.14	.14	.38	.27	5529
Parent Ratings	.2	.16	.67	.27	698
Child + Teacher + Parent	.18	.09	.77	.21	501

Note: Each row reports results from linear probability models with retention as the outcome. Columns 1–2 show results without teacher fixed effects; Columns 3–4 include teacher fixed effects. Adjusted R^2 accounts for model degrees of freedom. All specifications include two lags of test scores. Child Characteristics includes gender and age of the student, and their mother's education. Teacher Rating includes the year 2 teacher rating of the child. Parent Ratings includes the year 2 parent ratings of the child. The sample size decreases as more variables are included.

Together these results show us that observables, including lagged test scores, do not fully explain the retention decision, and that teacher differences in cutoffs are important. Overall, there is still a considerable amount of unexplained variation in the retention decision, and because of this, it is possible for there to be comparable students who face different

[13] In this setting, girls and boys are not retained at different rates. Table A.1 shows that this is true on average and also when controlling for age and test scores.

retention decisions simply because they were assigned different teachers or their teacher made a judgment error. The full regression tables are shown in appendix Table A.2.

4.3 Estimating Equation and Interpretation

The intuition behind our empirical specification is that we need to compare two students who would have similar outcomes in year $t+1$ if both progressed normally, however, in the data, one student was retained after year t. By comparing their outcomes in year $t+1$, we then see the impact of retention. The challenge in this estimation strategy is that we cannot see the counterfactual scores in $t+1$ when neither student was retained. Our solution is to estimate a very flexible specification with many controls, to compare similar students to the best of our ability.

Our main estimating equation is:

$$y_{ijt+1} = \beta R_{itj} + f(y_{ijt}, y_{ijt-1}) + X_i'\gamma + \mu_j + \varepsilon_{ijt}, \tag{1}$$

where y_{ijt+1} is the outcome in the year after retention, R_{itj} is an indicator for being retained at the end of year t, and $f(\cdot)$ is a fourth-order polynomial in two lags of test scores. X_i includes the age and gender of student i,[14] as well as their teacher's rating of them, and μ_j are school fixed effects. $\varepsilon_{i,j,t}$ is an idiosyncratic error and in the estimation all standard errors are clustered at the school level.

This specification compares students with similar observed characteristics, but different retention outcomes, which we assume are due to threshold heterogeneity and noise in perceived ability. School fixed effects control for systematic differences across schools in promotion criteria. We are identifying the effect of grade repetition from plausibly random

[14] We control for age because younger students may be less intellectually mature and more likely to be held back (Mahjoub (2017), Eide and Showalter (2001)) or older students may prioritize school less or are held back due to misbehavior. We control for gender because boys may be more disruptive in class, leading to retention for behavioral reasons, however their rowdiness may also distract them from their studies, biasing coefficients upwards. Note however, as shown in tables A.1 and 7 we find no difference based on gender, either in terms of the probability of retention or in differential outcomes.

variation in teachers' promote/retain decisions, i.e. mistakes and cross-teacher differences in standards. This is plausible because teachers are trained to teach, not assess student abilities, and may lack sufficient information to do so (Brophy (2006)). Without clearly defined achievement standards and corresponding mechanisms of assessment (which may not exist in many low-income countries), teachers' decisions should be relatively arbitrary (and we wouldn't expect that these arbitrary judgments systematically point in one direction or another). Therefore, their promotion decisions may be made with idiosyncratic error. This is especially true due to the binary nature of the retention decision; a small change in a teacher's noisy measurement of student ability may push a student into promotion or retention.

4.4 Differences Between Those Retained and Those Promoted

Our identification strategy rests on the idea that, conditional on controls, we can compare outcomes for retained and promoted students. This implicitly assumes that, after accounting for observed ability, the remaining differences between these two groups are random. In this section, we assess whether this assumption is supported by the data.

On average, students who are retained differ from those who are promoted. This is to be expected: students who are not promoted have significantly lower test scores and teacher ratings before the retention decision (Table 2). These patterns are consistent with a setting in which the retention decision is based, at least in part, on teacher perceptions of academic ability. We also see large differences between those promoted and those retained in the year after the retention decision, shown in Table 4. One goal of this paper is to determine how much, if any, of the difference in Table 4 is caused by being retained.

However, these averages mask the considerable overlap in ability between the two groups. Figure 2, for example, shows that many retained students outperform some promoted students on their pre-retention test scores. This is also reflected in the predictive power of observables for the retention decision. Table 3 shows that, even with detailed controls in-

Table 4: Differences in test scores and teacher ratings after the retention decision

	Promoted	Not Promoted	Difference	Total N
English score (Year 3)	0.837 (0.804)	-0.143 (0.940)	-0.981***	8071
Math score (Year 3)	0.953 (0.924)	-0.275 (1.221)	-1.228***	8071
Urdu score (Year 3)	0.974 (0.782)	-0.069 (0.970)	-1.044***	8071
Rating by teacher (Year 3)	6.326 (2.280)	4.754 (2.159)	-1.571***	6698

Note: Entries in first two columns are means with standard deviations in parentheses. All variables are from year 3 (note that the retention decision happens between year 2 and year 3). The sample for the first three rows are all students with test scores in year 1, year 2, and year 3. The sample for the fourth row is all students with test scores in year 1, year 2, and year 3, and a teacher rating in year 2. In the third column, *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

cluding lagged test scores, teacher and parent ratings, and basic demographics, we are only able to explain a modest share of the variation in retention outcomes (R^2 values below 0.20 without teacher fixed effects, and low adjusted R^2 even with them). This suggests that teacher discretion, often driven by unobserved or idiosyncratic factors, plays a meaningful role.

We now test more directly whether retained and promoted students are balanced once we condition on the full set of covariates used in our main specification. Specifically, we re-estimate a version of Equation 1 where the dependent variables are baseline characteristics, and the independent variable is an indicator for whether the student was retained. If our identifying assumption holds, then the coefficient on the retention dummy should be statistically indistinguishable from zero across these regressions.

Table 5 presents these conditional balance tests. Across a range of variables, including gender, age, and mother's education, we find that the two groups are similar once we control for prior achievement and include teacher fixed effects. Conditional balance tests for further variables are shown in the Appendix, in Table A.3. These results suggest that, conditional on ability and other controls, the remaining variation in retention status is plausibly as good as random. In other words, the residual differences between retained and promoted students appear to be driven by noise in the teacher's perception or variation in their implicit standards, rather than systematic differences in observed characteristics.

Table 5: Conditional balance tests – Primary covariates

Outcome Variable	**Coef. on Not Promoted**	**SE**	**p-value**
Female (1 = girl)	0.0021	0.0117	0.8579
Age (in years)	0.0249	0.0726	0.7315
Mother educated (1 = yes)	-0.0164	0.0237	0.4886

Note: In each row, Equation 1 is estimated with the Outcome Variable as the dependent variable. The coefficient on the retention variable, as well as its standard error and p-value are presented in the table.

Together, this evidence supports the internal validity of our empirical strategy: it is reasonable to interpret differences in later outcomes between retained and promoted students, conditional on observables, as reflecting the causal impact of grade repetition.

5 Results

We begin this section by confirming prior research that grade repetition reduces academic outcomes. Then, to understand the mechanisms by which these changes in academic outcomes occur, we focus on the reactions of parents, teachers, and students themselves to grade repetition. For every outcomes, we estimate Equation 1, which includes school fixed effects, teacher ratings of the student, and basic student-level controls. Lagged (and double-lagged) test scores enter in 4th order polynomial form to allow flexibility, and standard errors are clustered at the school level (Abadie et al. (2017) and Cameron and Miller (2015)). We then present robustness checks on these estimates, to validate our specification.

5.1 How are academic outcomes impacted by repeating?

The first outcomes we consider are test scores (Math, English, and Urdu) and dropping out, both one and two years after the retention decision.[15] Results are presented in Table 6, and

[15]By "drop out," we mean whether the student drops out before year $t+1$. The timeline is as follows. At the end of year t, teachers assign a promotion status to each student (either promoted to next grade or repeat current grade). Between years t and $t+1$, each student makes a decision to stay in school or dropout (they can also leave the mauza or leave the sample for unspecified reasons, however we ignore those options here). Our dropout variable is equal to one if the child dropped out of school and equal to zero if they are

show that repeating persistently lowers test scores and increases the probability of dropping out the year following the retention decision. For all three test scores, we find that repeaters score approximately 0.3 standard deviations lower on tests in all three subjects, and that this gap remains fairly constant even two years after the retention decision. The probability of dropping out the next year increases by 7 percentage points when students are retained, from a baseline average of 3%. However, the increase in dropout is not persistent, and the coefficient on retention two years later is a fairly precise zero.

Table 6: The effect of repetition on test scores and dropout

	Math		English		Urdu		Dropout	
	(1) 1 yr	(2) 2 yrs	(3) 1 yr	(4) 2 yrs	(5) 1 yr	(6) 2 yrs	(7) 1 yr	(8) 2 yrs
Not Promoted	**-0.440*** **	**-0.322*** **	**-0.278*** **	**-0.338*** **	**-0.371*** **	**-0.364*** **	**0.070*** **	**0.009**
	(0.055)	(0.076)	(0.036)	(0.050)	(0.043)	(0.056)	(0.014)	(0.019)
Lagged Score	**0.467*** **	**0.555*** **	**0.404*** **	**0.364*** **	**0.467*** **	**0.432*** **	**-0.005**	**-0.004**
	(0.027)	(0.037)	(0.023)	(0.033)	(0.024)	(0.028)	(0.007)	(0.011)
2yr Lagged Score	**0.247*** **	**0.240*** **	**0.195*** **	**0.142*** **	**0.226*** **	**0.174*** **	**-0.018** **	**-0.010**
	(0.027)	(0.035)	(0.020)	(0.032)	(0.027)	(0.030)	(0.008)	(0.010)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Rating	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag/Double Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	5529	3992	5529	3992	5529	3992	6222	4633
R^2	0.72	0.70	0.78	0.75	0.75	0.73	0.19	0.25

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. We run separate regressions for each school subject (math, English, and Urdu). Lagged score refers to a student's previous year score on the appropriate school subject (the lagged outcome variable).
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

The results in Table 6 control for gender, but it is possible that girls and boys react differently when retained. In this setting, girls and boys are not retained at different rates (see Table A.1). Table 7 then considers how the impacts of retention differ between girls and boys by allowing an interaction term between the retention dummy and a gender dummy. The results suggest that girls fare better than boys after being retained, in that their test scores do not decrease by such a large amount. However, the results are noisier than the main effects presented in Table 6 and not as large. For the remainder of the analysis, we will control for gender but we will not show the results separately.

still enrolled in school.

Table 7: The effect of repetition on test scores and dropout - By gender

	Math		English		Urdu		Dropout	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Not Promoted × Female		**0.1644**		**0.1695****		**0.1806****		**0.0036**
		(0.1062)		(0.0701)		(0.0845)		(0.0275)
Not Promoted	**-0.4403*****	**-0.5225*****	**-0.2783*****	**-0.3634*****	**-0.3706*****	**-0.4614*****	**0.0701*****	**0.0684*****
	(0.0553)	(0.0850)	(0.0364)	(0.0575)	(0.0434)	(0.0708)	(0.0145)	(0.0196)
Female	**-0.0341**	**-0.0391**	**0.0404***	**0.0353**	**0.0604****	**0.0550****	**-0.0036**	**-0.0037**
	(0.0279)	(0.0280)	(0.0218)	(0.0218)	(0.0256)	(0.0257)	(0.0080)	(0.0079)
Lagged Score	**0.4673*****	**0.4671*****	**0.4042*****	**0.4046*****	**0.4668*****	**0.4659*****	**-0.0046**	**-0.0046**
	(0.0270)	(0.0269)	(0.0231)	(0.0231)	(0.0238)	(0.0237)	(0.0074)	(0.0074)
2yr Lagged Score	**0.2473*****	**0.2457*****	**0.1951*****	**0.1946*****	**0.2263*****	**0.2256*****	**-0.0177****	**-0.0178****
	(0.0269)	(0.0270)	(0.0199)	(0.0198)	(0.0271)	(0.0271)	(0.0085)	(0.0084)
Constant	**0.7768*****	**0.7805*****	**0.6076*****	**0.6091*****	**0.6420*****	**0.6440*****	**0.0157****	**0.0158****
	(0.0314)	(0.0316)	(0.0261)	(0.0262)	(0.0255)	(0.0255)	(0.0080)	(0.0080)
AgeControls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Rating	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	5529	5529	5529	5529	5529	5529	6222	6222
R^2	0.72	0.72	0.78	0.78	0.75	0.75	0.19	0.19

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. We run separate regressions for each school subject (math, English, and Urdu). Lagged score refers to a student's previous year score on the appropriate school subject (the lagged outcome variable).
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

5.2 Mechanism - Teachers

As a first step in exploring mechanisms behind the negative effect of grade repetition on test scores, we consider the role of the teacher. To proxy for a teacher's opinion of a given student, we will use the teacher rating.[16] In Table 8 we present results where the outcome variable is the teacher's rating of the student, with specifications using the rating of the contemporaneous teacher (the teacher who holds the student back) as well as the next year's teacher (the teacher after the retention decision).

For the contemporaneous teacher, we find that they have a significantly lower rating for the child than we would predict given the student's lagged test scores and other information. This can be seen in column (1) of Table 8. If we consider the student's teacher for the next year, we see that they also have a negative view of the student (column (2)), however, when controlling for the students test scores in that year, the fact that they were retained has no significant impact on the teacher rating. In fact, comparing the R^2 of columns (3) and (4),

[16]In the main specification, we include teacher rating as a control. In these specifications we have removed the control, and instead use it as a dependent variable.

we learn that promotion status has no additional predictive value on the subsequent-year teacher's rating of the student above test scores.

Table 8: The effect of repetition on teacher ratings

	(1) Contemporaneous	(2) Next Year	(3) Next Year	(4) Next Year
Not Promoted	**-0.8575*** **	**-0.3729*** **	**-0.0398**	
	(0.160)	(0.137)	(0.140)	
Lagged Average Score	**1.1570*** **	**1.1220*** **	**0.5880*** **	**0.5891*** **
	(0.095)	(0.088)	(0.093)	(0.093)
2yr lagged Average Score	**0.4842*** **	**0.5342*** **	**0.2781*** **	**0.2786*** **
	(0.106)	(0.099)	(0.093)	(0.093)
Age	**-0.0791*** **			
	(0.026)			
Average Score: Y3			**1.0268*** **	**1.0318*** **
			(0.086)	(0.084)
Constant	**6.0961*** **	**5.4368*** **	**4.7356*** **	**4.7292*** **
	(0.289)	(0.085)	(0.106)	(0.104)
Age/Gender Controls	Yes	Yes	Yes	Yes
Fixed Effect	School	School	School	School
Lag Polynomial	4th Order	4th Order	4th Order	4th Order
N	5529	5446	5446	5446
R^2	0.45	0.49	0.51	0.51

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. The first column refers to the teacher who retained the student, and the remaining three columns refer to the teacher who taught the student the following year. Lagged score refers to a student's previous year average test score. Column (4) illustrates that promotion does not increase the R-squared of the regression.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

These results tell us two things. First, teachers do not seem to label students as "repeaters" and have a negative bias towards them. If anything, it seems as if each year the teacher views the students as i.i.d draws conditional on their test scores. Second, if we were worried that the contemporaneous teacher retained the student because of something that was observable to them but unobservable to us, it seems as if these traits were also unobservable to the student's teacher the following year. This supports our theory that teachers are trained to teach, not to recognize which students should be retained, and that some of the retention that we see in the data is caused by random noise.

5.3 Mechanism - Household

Next, we use our linked household-student-school panel data to provide new insights into how parents' attitudes and investments in relation to their children change following grade repetition.

We find that parents treat grade repetition as a signal of lower intellectual ability or potential. Table 9 has three parental perceptions as the outcome variables. Parents decrease their ratings of their child's performance in school, work ethic, and intelligence by 0.15 to 0.32 standard deviations after their child is retained. This effect holds even after controlling for lagged parent ratings, demonstrating that there is a clear change in parent evaluations of a child after she has been retained.

Table 9: The effect of repetition on parent perceptions

	School Performance		Work Ethic		Intelligence	
	(1)	(2)	(3)	(4)	(5)	(6)
Not Promoted	**-0.1516****	**-0.1321****	**-0.3281****	**-0.3120***	**-0.2552****	**-0.2405***
	(0.062)	(0.063)	(0.158)	(0.160)	(0.129)	(0.124)
Lagged Average Score	**0.1811*****	**0.1668*****	**0.2553*****	**0.2476*****	**0.3758*****	**0.3485*****
	(0.039)	(0.038)	(0.090)	(0.089)	(0.079)	(0.078)
2yr lagged Average Score	**-0.0366**	**-0.0586**	**0.0309**	**0.0225**	**-0.1467**	**-0.1673**
	(0.050)	(0.049)	(0.112)	(0.112)	(0.112)	(0.110)
Lagged School Performance		**0.1532*****				
		(0.027)				
Lagged Hardworking				**0.0830**		
				(0.061)		
Lagged Inteligence						**0.2366*****
						(0.061)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	741	741	741	741	741	741
N	0.13	0.17	0.08	0.08	0.10	0.12

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. Lagged score refers to a student's previous year average test score. Outcome questions are listed in full in the text.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 10 examines perceptions two years after the retention, and finds mixed results depending on the question. For school performance, there is no significant effect, perhaps signalling that parents have revised downward their expectations and their child is no longer "disappointing" them. The impacts on work ethic are persistent and similar in size two years after the retention. The impacts on perceived intelligence are no longer significant, but in terms of magnitude are still half the size as the initial estimates. Together, this shows that having their child repeat a grade causes parents to significantly downgrade their perceptions of their child, and some of these shifts are long-lasting.

Table 10: The effect of repetition on parent perceptions - Two years after repeating

	School Performance		Work Ethic		Inteligence	
	(1)	(2)	(3)	(4)	(5)	(6)
Not Promoted	**-0.0467**	**-0.0330**	**-0.3141***	**-0.2730***	**-0.1355**	**-0.1180**
	(0.136)	(0.134)	(0.165)	(0.161)	(0.165)	(0.161)
Lagged Average Score	**0.2392*****	**0.2323*****	**-0.0381**	**-0.0560**	**0.0262**	**0.0056**
	(0.067)	(0.067)	(0.092)	(0.092)	(0.093)	(0.091)
2yr lagged Average Score	**-0.0453**	**-0.0562**	**0.1236**	**0.1057**	**0.0713**	**0.0568**
	(0.080)	(0.080)	(0.108)	(0.106)	(0.123)	(0.122)
Lagged School Performance		**0.0850***				
		(0.044)				
Lagged Hardworking				**0.2042*****		
				(0.062)		
Lagged Inteligence						**0.2039*****
						(0.062)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	710	710	710	710	710	710
R^2	0.09	0.09	0.03	0.05	0.05	0.06

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. Lagged score refers to a student's previous year average test score. Outcome questions are listed in full in the text.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

While the parental perception questions are insightful, they are also somewhat vague. In the fifth round of surveys, approximately 5 years after the retention decision, a more concrete survey questions specifically asked parents what score they believe that their child could achieve on a test (they were asked for a minimum, maximum, and average). Table 11 presents suggestive evidence that parents substantially decreased all three of these measures if their child was retained.[17] This, along with the previous two tables, offers evidence that parents significantly lower their expectations of their child once they are retained, and that

[17]The sample size shrinks in this table because the question was asked in a follow-up survey for which not all households could be found and matched.

this effect persists for many years.

Table 11: The effect of repetition on parent beliefs

	(1) Average	(2) Maximum	(3) Minimum
Not Promoted	**-4.9701***	**-5.9614***	**-3.9929**
	(2.761)	(3.084)	(2.550)
Lagged Average Score	**1.1569**	**1.6451**	**0.3857**
	(2.246)	(2.510)	(2.120)
2yr lagged Average Score	**3.0027**	**2.4747**	**2.8306**
	(2.658)	(2.820)	(2.484)
Constant	**50.2885*****	**61.3707*****	**39.6159*****
	(1.944)	(2.097)	(1.803)
Age/Gender Controls	Yes	Yes	Yes
Lag Polynomial	4th	4th	4th
N	513	513	514
R^2	0.43	0.43	0.40

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. Lagged score refers to a student's previous year average test score. Outcome refers to how many questions parents expect a student to answer correctly on a test.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

In addition to revising their beliefs and expectations downwards, parents also reallocate resources away from repeating students. Table 12 shows annual expenditure on a repeater's education (which includes longer-term educational expenses including annual school fees, school uniforms, textbooks, and school supplies) decreases by approximately 176 PKR, or 20% of mean annual educational expenditure (column 3).[18] While there appears to be less of a decrease in monthly educational expenditures (which includes shorter-term expenses such as transportation costs, private tutoring, and pocket money for school), the point estimate is negative.[19] The magnitude of the decrease in expenditure gets even larger two years after the retention.

To confirm that these effects represent a resource reallocation away from the repeating student specifically (rather than from some confounding household-level shock), we estimate the same regression but change the outcome variable to be the average expenditure on all

[18]This drop in educational expenditure is not due to dropouts–all children in this regression's sample remained in school before and after grade repetition.

[19]We are lacking statistical power for this regression as it is identified off of only 56 repeaters in the smaller subsample of students whose household was included in the household survey.

Table 12: The effect of repetition on parent investments

	Monthly \$ $(t+1)$		Annual \$ $(t+1)$		Monthly \$ $(t+2)$		Annual \$ $((t+2))$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Not Promoted	**-17.7139**	**-15.0460**	**-176.1899*****	**-111.4625***	**-47.2153****	**-46.2378****	**-209.9473***	**-199.9287***
	(14.908)	(18.070)	(59.553)	(57.743)	(22.513)	(21.416)	(115.382)	(114.009)
Lagged Average Score	**12.3195**	**-0.0491**	**-15.5160**	**-20.4853**	**13.0877**	**8.5559**	**122.1426**	**121.3734**
	(15.489)	(13.301)	(47.249)	(45.470)	(17.020)	(16.707)	(77.506)	(77.480)
2yr lagged Average Score	**33.6261**	**10.5395**	**102.3851***	**65.1619**	**20.5382**	**12.0793**	**32.8975**	**27.1361**
	(22.248)	(15.056)	(59.911)	(57.953)	(21.311)	(21.183)	(102.077)	(101.723)
Lagged Monthly \$		**0.8893*****				**0.3258*****		
		(0.149)				(0.091)		
Lagged Anual \$				**0.4742*****				**0.0734**
				(0.055)				(0.085)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	741	741	741	741	741	741	741	741
R^2	0.05	0.33	0.05	0.16	0.04	0.06	0.05	0.05

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. Lagged score refers to a student's previous year average test score. Full descriptions of short-term and long-term investments are in the text.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

other siblings in the household. Table 13 shows that household expenditure on the siblings of repeaters may even rise, though the effect is not statistically distinguishable from zero (columns 1-4).[20] Correspondingly, parent perceptions of the repeater's siblings also rise or stay constant (Table 13 columns 5-10).

The previous evidence indicates that parents place the blame of retention on their child, and the following tables suggest that they do not put any on their child's teachers. In the parent surveys, parents were asked questions about how they view their child's teachers. Three of these questions were "How regular is your child's class-teacher overall?", "How good would you say that your child's class-teacher is in his/her teaching skills?", and "How good would you say that your child's class-teacher is overall?". Parents were allowed to respond that they "do not know", which is why the sample size decreases from the previous tables. Table 14 shows that parents of repeating students do not hold worse views of their child's teacher than parents of promoted students (conditional on expressing a view of their child's teacher). All coefficients are fairly precisely estimated zeros. Table 15 shows the same analysis, but for teachers the following year (the year during which the student is repeating the grade). The sample size is small, but it does not seem as if parents have any different

[20] The sample sizes in this table are smaller than in tables 11 and 12 due to students who have no siblings.

Table 13: The effect of repetition on investments in and perceptions of siblings

	Monthly \$ $(t+1)$		Annual \$ $(t+1)$		Performance (t+1)		Work Ethic (t+1)		Inteligence (t+1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Not Promoted	**0.5588**	**5.3413**	**29.2969**	**74.7752**	**0.1453**	**0.0234**	**0.2836***	**0.2352**	**0.0218**	**0.0073**
	(29.768)	(27.152)	(78.805)	(69.771)	(0.136)	(0.116)	(0.168)	(0.170)	(0.158)	(0.156)
Lagged Average Score	**26.8716**	**13.8832**	**92.1976**	**61.4781**	**0.1592***	**0.1126***	**0.1884****	**0.1850****	**0.0803**	**0.0649**
	(16.515)	(14.155)	(60.140)	(56.643)	(0.086)	(0.066)	(0.083)	(0.080)	(0.079)	(0.073)
2yr lagged Average Score	**19.2244**	**14.8686**	**18.2714**	**32.7411**	**-0.0202**	**-0.0223**	**-0.2928*****	**-0.2922*****	**-0.0812**	**-0.0163**
	(21.413)	(17.907)	(74.761)	(69.083)	(0.106)	(0.082)	(0.107)	(0.104)	(0.105)	(0.100)
Lagged Sibling Monthly \$		**1.9400*****								
		(0.267)								
Lagged Sibling Annual \$				**1.1220*****						
				(0.223)						
Lagged Sibling Overall						**0.4903*****				
						(0.036)				
Lagged Sibling Hardworking								**0.2441*****		
								(0.089)		
Lagged Sibling Inteligence										**0.3269*****
										(0.079)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	719	719	719	719	719	702	719	702	719	702
R^2	0.05	0.20	0.07	0.22	0.07	0.34	0.04	0.05	0.02	0.05

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. The table shows results from a student-level regression of average household expenditure on all siblings (leaving out the student) on whether the student repeated or not (columns 1-4) and of average parent evaluation of siblings (leaving out the student) on whether the student repeated or not. Lagged score refers to a student's previous year average test score. Full descriptions of short-term and long-term investments are in the text. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

views about the teachers if their child is held back or repeating the grade.

Table 14: The effect of repetition on parental ratings of teachers (current year)

	How Regular		Good Teacher		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)
Not Promoted	**-0.0538**	**0.0158**	**0.0130**	**0.0381**	**-0.0408**	**-0.0548**
	(0.054)	(0.062)	(0.091)	(0.108)	(0.092)	(0.109)
Lagged Average Score		**0.0062**		**-0.0268**		**-0.0785**
		(0.033)		(0.059)		(0.060)
2yr lagged Average Score		**0.0773***		**0.1743****		**0.0616**
		(0.045)		(0.075)		(0.074)
Age/Gender Controls	No	Yes	No	Yes	No	Yes
Lag Polynomial	-	4th	-	4th	-	4th
N	700	631	699	630	686	620
R^2	0.00	0.07	0.00	0.04	0.00	0.04

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. Lagged score refers to a student's previous year average test score.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 15: The effect of repetition on parental ratings of teachers (next year)

	How Regular		Good Teacher		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)
Not Promoted	**-0.0078**	**0.0184**	**-0.0102**	**0.0655**	**0.0711**	**0.0136**
	(0.074)	(0.086)	(0.104)	(0.116)	(0.107)	(0.121)
Lagged Average Score		**0.0876***		**0.0781**		**0.0765**
		(0.047)		(0.065)		(0.068)
2yr lagged Average Score		**0.0300**		**0.1659****		**0.0891**
		(0.061)		(0.084)		(0.088)
Age/Gender Controls	No	Yes	No	Yes	No	Yes
Lag Polynomial	-	4th	-	4th	-	4th
N	820	739	826	744	820	739
R^2	0.00	0.05	0.00	0.06	0.00	0.04

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. Lagged score refers to a student's previous year average test score.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

5.4 Mechanisms - Student

Being retained significantly lowers a child's test scores the next year, however, it is possible that there are other impacts as well, which could be harder to measure. Especially if retention has a random component (after controlling for test scores), it is possible that children would feel quite discouraged if they feel like they have been retained at random. To analyze this further, we consider several survey questions as outcomes in our main specification. These questions once again come from the fifth round follow-up survey, which takes place approximately five years after the retention decision. The sample size is smaller because of this. Table 16, column (1) shows that retention has a large negative impact on the student agreeing with the statement "If I study hard at school I will be rewarded with a better job in the future."[21] We take this as evidence that retention discourages students, and makes them doubt their own agency in their studies.

Columns (2) through (5) show that retention does not seem to discourage students in other areas of their life. There are no significant impacts of retention on any of the four

[21] To obtain the outcome variable for this regression, we recode a Likert scale response into a binary for agree/disagree.

other questions. The questions for each column are the following: (2) “If I try hard, I can improve my situation in life"; (3) “Other people in my family make all the decisions about how I spend my time"; (4) “I like to make plans for my future studies and work"; (5) “I have no choice about the work I do - I must work”. This specific decrease in belief that effort in school will benefit their life goals could be one reason why students who are retained see a decrease in their test scores.

Table 16: The effect of repetition on student beliefs

	(1) Study Hard	(2) Try Hard	(3) Others Decisions	(4) Make Plans	(5) Must Work
Not Promoted	**-0.2602****	**-0.0375**	**0.0110**	**-0.0584**	**0.0772**
	(0.102)	(0.032)	(0.084)	(0.087)	(0.091)
Lagged Average Score	**0.0210**	**-0.0226**	**-0.0378**	**0.0323**	**-0.0243**
	(0.046)	(0.015)	(0.061)	(0.043)	(0.057)
2yr lagged Average Score	**-0.0017**	**0.0277**	**-0.0312**	**0.0200**	**-0.0800**
	(0.062)	(0.021)	(0.078)	(0.056)	(0.068)
Constant	**0.8329*****	**0.9918*****	**0.5317*****	**0.8900*****	**0.1568*****
	(0.052)	(0.006)	(0.054)	(0.042)	(0.051)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order
N	578	578	578	578	578
R^2	0.35	0.27	0.30	0.32	0.20

Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. Lagged score refers to a student's previous year average test score. Outcome questions are listed in full in the text.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

6 Robustness Checks for the Main Results

We consider two primary threats to our identification strategy. First, measurement error in the teacher's signal may correlate with observable factors (e.g., if girls or poorer students are systematically underestimated). Second, teachers may select students to retain based on unobserved traits that also affect outcomes, such as behavior or motivation.

First, to address these concerns, we control for the age and gender of the student in all specifications, to ensure that we are picking up any biases the teacher may have on these two dimensions. We also control for the teacher's rating of the student, as this would likely pick up any variables that are unobserved to us but important to the teacher. Further, we present conditional balance tests in Table 5 and Table A.3.

Next, several of the tables already presented provide evidence against these threats. The results in Table 8 show that the year following the retention decision, teachers do not rate repeaters any differently than their peers (conditional on the controls and the most recent test score). This provides evidence that retention decisions are not being made based on some time invariant unobserved traits. If there was some persistent unobserved factor, for example poor behavior, that influenced the retention decision and also influenced our outcomes the next year, we would expect next year's teacher to also incorporate that into their ratings.

Table 13 doubles as a Placebo regression, as it shows that there is no impact of retention on the parental beliefs or investments of siblings. If retention and its negative effects were both caused by some negative shock at home that is unobserved to us, we would expect all the siblings in the household to have the same negative effects.

To assess the robustness to selection on unobservables, we run an Oster test (Oster, 2019). The results are shown in Table 17, and the implied values of Delta are generally above 1, suggesting that unobserved factors would need to be at least as important as observed controls to eliminate the estimated effects. The main specification controls for two lags of test scores, gender, age, and the teacher's rating, in addition to the retention indicator. It is challenging to come up with possible unobserved variables in this context that would have a comparable effect on the outcomes.

In Appendix A.1.2 we present further robustness checks of the main results. These tables include additional variables such as parent ratings, household wealth, child health, and a more flexible form of teacher rating. The results are robust to each of these specifications.

7 Conclusion

This paper provides new evidence on the consequences of grade repetition in a low-income country, where promotion decisions are made by teachers based on informal criteria. Using

Table 17: Oster test for robustness of main results

	Math	English	Urdu	Dropout
Coef. (Not promoted)	-0.440	-0.278	-0.371	0.070
Oster bound (β)	-0.020	0.021	-0.034	0.064
Delta	1.039	0.937	1.085	3.583
Observations	5529	5529	5529	6222

Note: This table reports results from the Oster (2019) test for robustness to omitted variable bias, corresponding to the estimates in Table 6. We restrict attention to the impacts measured one year after grade retention. Reported coefficients are the baseline estimates of not being promoted. The bound β^* indicates the estimated coefficient after adjusting for potential selection on unobservables, with δ measuring the relative degree of selection on unobservables versus observables. Following standard practice, we set R^* equal to 1.3 times the observed R^2.

detailed administrative and survey data from public and private schools in Pakistan, we include a rich set of controls for student ability and school fixed effects to estimate the impacts of retention on student outcomes and household responses.

We find that repeating a grade has a negative impact on test scores the next year, and these effects persist into the second year after the retention decision. In contrast, the effects on dropout are large but only immediately after being retained.

To understand the mechanisms behind these academic responses, we examine how teachers, households, and students react to retention using survey data. The results suggest that parents revise downward their beliefs about their child's academic ability and the value of investing in their education. Importantly, these responses are child-specific: parents do not reduce investments in siblings, nor do they shift blame toward the school or teacher. We also find that students themselves revise their beliefs about the relationship between effort and academic success, although beliefs about their own ability remain largely unchanged.

Taken together, these findings highlight that the effects of grade retention are not limited to academic outcomes, but also operate through behavioral channels in the household and for the student. This is a setting where the consequences of falling behind are steep and retention is viewed as a signal of failure. These results underscore the need to consider how teachers and schools are using grade retention as a policy, and how the optimal retention strategy may differ substantially depending on the setting.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Andrabi, T., J. Das, A. I. Khwaja, T. Vishwanath, and T. Zajonc (2007). Learning and educational achievements in punjab schools (leaps): Insights to inform the education policy debate. *World Bank, Washington, DC*.

Becker, G. S. and N. Tomes (1976). Child endowments and the quantity and quality of children. *Journal of political Economy 84*(4, Part 2), S143–S162.

Behrman, J. R., M. R. Rosenzweig, and P. Taubman (1994). Endowments and the allocation of schooling in the family and in the marriage market: the twins experiment. *Journal of political economy 102*(6), 1131–1174.

Bergman, P. (2021). Parent-child information frictions and human capital investment: Evidence from a field experiment. *Journal of political economy 129*(1), 286–322.

Borghesan, E., H. Reis, and P. E. Todd (2022). Learning through repetition? a dynamic evaluation of grade retention in portugal. *PIER Working Paper* (22-030), 1–80.

Brophy, J. (2006). *Grade repetition*. International Institute for Educational Planning.

Cabrera-Hernandez, F. (2022). Leave them kids alone! the effects of abolishing grade repetition: evidence from a nationwide reform. *Education Economics 30*(4), 339–355.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Chohan, B. I. and S. A. Qadir (2011). Automatic promotion policy at primary level and mdg-2. *Journal of Research & Reflections in Education (JRRE) 5*(1).

Das, J., P. Pandey, and T. Zajonc (2012). Learning levels and gaps in pakistan: A comparison with uttar pradesh and madhya pradesh. *Economic and Political Weekly*, 228–240.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review 109*(8), 2728–2765.

Eide, E. R. and M. H. Showalter (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education review 20*(6), 563–576.

Eisemon, T. O. (1997). Reducing repetition: Issues and strategies.

Eren, O., M. F. Lovenheim, and N. H. Mocan (2018). The effect of grade retention on adult crime: Evidence from a test-based promotion policy. Technical report, National Bureau of Economic Research.

Ferreira Sequeda, M., B. H. Golsteyn, and S. Parra-Cely (2018). The effect of grade retention on secondary school performance: Evidence from a natural experiment. Technical report, IZA Discussion Papers.

Figlio, D. N. and U. Özek (2019). An extra year to learn english? early grade retention and the human capital development of english learners. Technical report, National Bureau of Economic Research.

Gan, T. (2021). How parental educational investments respond to changes in ability belief: An application of big data techniques in education. *Working Paper*.

Glick, P. and D. E. Sahn (2010). Early academic performance, grade repetition, and school attainment in senegal: A panel data analysis. *The World Bank Economic Review 24*(1), 93–120.

Gomes-Neto, J. B. and E. A. Hanushek (1994). Causes and consequences of grade repetition: Evidence from brazil. *Economic Development and Cultural Change 43*(1), 117–148.

Hu, L.-C. and E. Hannum (2020). Red flag: Grade retention and student academic and behavioral outcomes in china. *Children and Youth Services Review 113*, 104896.

Jacob, B. A. and L. Lefgren (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of economics and statistics 86*(1), 226–244.

Jacob, B. A. and L. Lefgren (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics 1*(3), 33–58.

King, E. M., P. F. Orazem, and E. M. Paterno (2016). Promotion with and without learning: Effects on student enrollment and dropout behavior. *The World Bank Economic Review 30*(3), 580–602.

Mahjoub, M.-B. (2017). The treatment effect of grade repetitions. *Education Economics 25*(4), 418–432.

Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics 94*(2), 596–606.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Tomes, N. (1981). The family, inheritance, and the intergenerational transmission of inequality. *Journal of Political Economy 89*(5), 928–958.

A.1 Appendix

A.1.1 Additional descriptive tables and figures

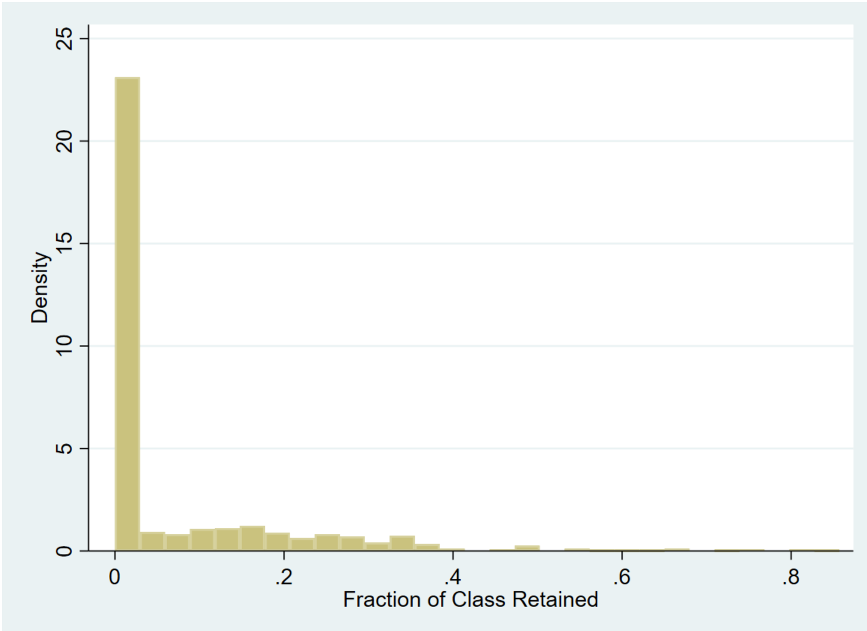
The first section of the Appendix provides additional tables and figures to support Section 4. First, Table A.1 shows that there is no difference in retention patterns across genders. Figure A.1 shows the distribution of the fraction of each class that was retained. Figure A.2 shows that teacher's ratings of students do not always align with their retention decisions. It shows one example classroom and the histogram plots the ratings of the students and whether they were retained or not. The full regression results associated with Table 3 can be found in Table A.2. Finally Table A.3 provides additional conditional balance tests to supplement those shown in Table 5.

Table A.1: The probability of repeating by gender

	(1)	(2)	(3)
Female	**0.0064**	**-0.0095**	**0.0003**
	(0.0057)	(0.0107)	(0.0117)
Lagged Score			**-0.0762*** **
			(0.0091)
2yr Lagged Score			**-0.0493*** **
			(0.0106)
Constant	**0.1035*** **	**0.0979*** **	**0.0652*** **
	(0.0038)	(0.0072)	(0.0117)
AgeControls	No	Yes	Yes
Teacher Rating	No	No	Yes
Fixed Effect	-	School	School
Lag Polynomial	-	-	4th
N	11869	11361	6113
R^2	0.00	0.18	0.34

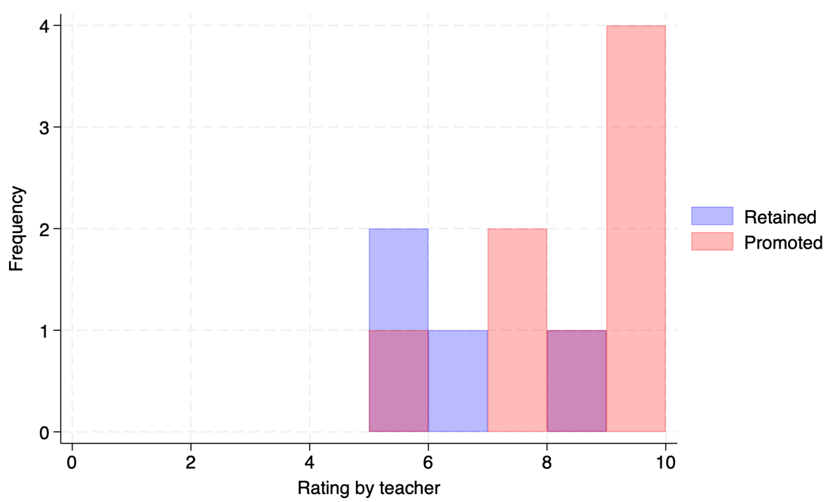
Note: Coefficient estimates bolded, standard errors (clustered at school-level) in parentheses below. The outcome variable is a dummy variable equal to one if the student repeats the grade. Lagged score refers to a student's previous year average score.
*** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level

Figure A.1: Across-teacher distribution of the fraction of class retained



Note: Histogram of the fraction of each teacher's classroom who is retained (including classrooms without any retained students).

Figure A.2: Example teacher: frequency histogram of number of promoted and retained students with each rating



Note: Plot shows the number of retained and promoted students (y-axis) with different teacher ratings (shown on x-axis)

Table A.2: Regression Predicting Repetition

	(1)	(2)	(3)	(4)	(5)	(6)	(7) int	(8) int
Lagged Average Score	**-0.1099***** (0.0079)	**-0.1037***** (0.0094)	**-2.3451***** (0.4743)	**-0.0872***** (0.0095)	**-0.1226***** (0.0324)	**-0.1304***** (0.0337)	**-0.0389** (0.1111)	**-0.0085** (0.1931)
2yr lagged Average Score	**-0.0463***** (0.0095)	**-0.0444***** (0.0111)	**0.4857** (0.5080)	**-0.0372***** (0.0110)	**-0.0328** (0.0303)	**-0.0199** (0.0450)	**-0.1719** (0.1433)	**-0.2791** (0.2704)
Mom Educated		**0.0010** (0.0078)	**0.0117** (0.1393)		**0.0011** (0.0076)			**-0.0771** (0.0594)
Child Wealth		**-0.0038** (0.0022)	**-0.0458** (0.0389)					**-0.0036** (0.0214)
Height		**-0.0020** (0.0046)	**-0.0786** (0.0778)					**-0.0001** (0.0396)
Weight		**-0.0002** (0.0052)	**-0.0136** (0.0875)					**0.0071** (0.0444)
Child Talent								
= 1				**0.1831***** (0.0206)	**0.1756***** (0.0208)			**0.1585** (0.1536)
= 2				**0.1180***** (0.0186)	**0.1114***** (0.0190)			**-0.1605** (0.1339)
= 3				**0.0713***** (0.0161)	**0.0687***** (0.0162)			**-0.1470** (0.1230)
= 4				**0.0231** (0.0131)	**0.0224** (0.0133)			**0.1044** (0.1299)
= 6				**-0.0208** (0.0117)	**-0.0253**** (0.0125)			**-0.1889** (0.1063)
= 7				**-0.0095** (0.0123)	**-0.0216** (0.0135)			**-0.1121** (0.1191)
= 8				**-0.0070** (0.0126)	**0.0052** (0.0155)			**-0.0778** (0.1150)
= 9				**0.0048** (0.0142)	**-0.0022** (0.0184)			**-0.0678** (0.1257)
= 10				**-0.0019** (0.0162)	**0.0314** (0.0227)			**0.1208** (0.2219)
Hardworking								
Below Average						**0.1650***** (0.0629)	**0.1945***** (0.0656)	**0.3568***** (0.1211)
Above Average						**-0.0330** (0.0470)	**-0.0600** (0.0521)	**-0.0636** (0.0949)
Intelligent								
Below Average						**-0.2381***** (0.0746)	**-0.2619***** (0.0790)	**-0.6469**** (0.2744)
Above Average						**0.0080** (0.0457)	**0.0331** (0.0508)	**-0.0598** (0.1267)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial 4th Order	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Test Score Components	No	No	Yes	No	Yes	No	Yes	Yes
Interact Scores, Child Chars	No	No	Yes	No	No	No	No	Yes
Interact Scores, TeachRatings	No	No	No	No	Yes	No	No	Yes
Interact Scores, ParRatings	No	No	No	No	No	No	Yes	Yes
Interact Par Ratings, TeachRatings	No	No	No	No	No	No	No	Yes
N	8071	5442	5434	5529	5529	698	698	487
Pseudo R^2	0.35	0.37		0.38	0.39	0.66	0.69	0.85

Note: This table shows coefficients from the regressions used to predict repetition used for computing the R^2's shown in table 3.

Table A.3: Conditional balance tests – Additional covariates

Outcome Variable	**Coef. on Not Promoted**	**SE**	**p-value**	**N**
Parent rating: School perf.	0.0607	0.1910	0.7511	545
Parent rating: Intelligence	0.1251	0.2063	0.5446	545
Asset-based wealth index	0.2555	0.4174	0.5410	546
Hours of family help	0.0435	1.6931	0.9795	543
Met child's teacher (1 = yes)	-0.0030	0.0223	0.8917	6263
Height-for-age z-score	-0.0119	0.0706	0.8657	6149
BMI z-score	-0.0263	0.0586	0.6537	6094

Note: Each row reports the coefficient from a separate regression of the listed baseline variable on a dummy for being retained, controlling for lagged test scores (4th order polynomial) and school fixed effects. Sample size varies across rows depending on variable availability.

A.1.2 Robustness of main results

Our rich data contains many variables, however not all students in the main sample have responses recorded for all variables. This means that we face a tradeoff: including more variables or having a larger sample. For the main analysis, we have included the most important variables. In the following subsections we replicate the main results while including other variables that readers may find relevant. Note that the sample size usually decreases, at times substantially. The following tables all replicate our main results for academic outcomes, shown in Table 6. Our results are robust to all of the inclusions.

A.1.2.1 Controlling for parent ratings

We add baseline parent ratings separately to the specification from teacher ratings since only very small subsample received ratings from *both* their teachers and their parents. We have two separate ratings from parents: one of a child's intelligence and one of a child's work ethic. Since the two are highly correlated, we introduce them to the model one at a time.

While inclusion of parent rating in the regression substantially reduces power,[22] the effect of repetition remains large and significant. Table A.4 (columns 2, 3, 5, 6, 8, and 9) shows that, accounting for lagged parental ratings, grade repetition reduces test scores by between .29 and .44 standard deviations for the three subjects.[23] Comparing the second and third columns for each subject to the first, we see that the repetition coefficient decreases by at most only 0.024 standard deviations when including the parent ratings, suggesting that our baseline model isn't missing any important dimension of latent student ability.[24] Additionally, none of the lagged parent ratings significantly explain test score variation, controlling for lagged scores.

[22] These variables come from a household survey which was administered only to ten randomly selected households per village. The sample is all tested students matched with a surveyed household. Due to the reduced sample and lack of intra-school variation, we switch from school- to mauza-level fixed effects.

[23] Note that due to the smaller sample size, we switch from school- to mauza-level fixed effects here– otherwise, identification comes from only 36 repeaters and 58 non-repeaters in classrooms for which we observe both repeaters and non-repeaters.

[24] Columns 10-12 show the specification with dropping out as the outcome variable. While we lack sufficient power to claim an effect, it is reassuring that the coefficient barely changes between column 10 (the baseline specification on the smaller subsample) and columns 11 and 12 (the specification including parent ratings). Hence, the lack of significance seems to be due to small sample size rather than because the effect disappears when controlling for parent ratings.

Table A.4: Effect of repetition on test scores - Considering characteristics observable to parents

	Math			English			Urdu			Dropout		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Not Promoted	**-0.433****	**-0.409****	**-0.436****	**-0.305****	**-0.291****	**-0.311****	**-0.393*****	**-0.369*****	**-0.389*****	**0.045**	**0.045**	**0.044**
	(0.179)	(0.181)	(0.179)	(0.143)	(0.143)	(0.142)	(0.143)	(0.137)	(0.141)	(0.036)	(0.036)	(0.036)
Lagged Score	**0.554*****	**0.549*****	**0.538*****	**0.481*****	**0.472*****	**0.484*****	**0.491*****	**0.478*****	**0.483*****	**0.001**	**0.001**	**0.002**
	(0.058)	(0.059)	(0.058)	(0.047)	(0.049)	(0.048)	(0.057)	(0.056)	(0.057)	(0.009)	(0.009)	(0.009)
2yr Lagged Score	**0.321*****	**0.317*****	**0.324*****	**0.317*****	**0.322*****	**0.317*****	**0.199*****	**0.196*****	**0.197*****	**-0.007**	**-0.007**	**-0.007**
	(0.067)	(0.067)	(0.065)	(0.053)	(0.053)	(0.052)	(0.064)	(0.062)	(0.063)	(0.007)	(0.006)	(0.007)
Hardworking (Lag)												
Below Average		**-0.196**			**-0.130**			**-0.146**			**-0.007**	
		(0.140)			(0.104)			(0.109)			(0.006)	
Above Average		**0.058**			**0.027**			**0.102****			**-0.008**	
		(0.061)			(0.055)			(0.050)			(0.006)	
Intelligent (Lag)												
Below Average			**-0.303**			**-0.112**			**-0.053**			**-0.001**
			(0.173)			(0.134)			(0.127)			(0.005)
Above Average			**0.054**			**-0.049**			**0.069**			**-0.009**
			(0.058)			(0.049)			(0.047)			(0.006)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mauza Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	698	698	698	698	698	698	698	698	698	760	760	760
R^2	0.68	0.69	0.69	0.73	0.73	0.73	0.74	0.74	0.74	0.25	0.26	0.26

Note: This table replicates the results shown in Table 6. The first column for each outcome variable is the exact same regression as in the main text, however the sample size is restricted to students who have are associated with a household survey, and therefore have parental responses. The results are robust.

A.1.2.2 Specification checks on controlling for two lags of test scores

The purpose of the value-added model is to "difference out" time-invariant factors. One way to check whether it succeeds is by testing if results are sensitive to inclusion of covariates that (1) should bias results in the absence of the value-added model and (2) we expect to remain constant year to year. If coefficients remain largely unchanged whether we include the variable or not and the time-invariant covariate doesn't significantly explain student performance, it should be taken as evidence that the value-added model is effectively accounting for time-invariant factors. We consider three different time-invariant confounding factors at the household-, child-, and parent-level. Two of these (household wealth and child health) may be subject to small shocks year to year. In each case, absent the value-added specification, lack of inclusion would bias results. However, using the value-added model, results remain largely unchanged, indicating that the model is successful.

The first test variable is household wealth. It affects both achievement and repetition since wealthier households may spend more resources to help their children, including pressuring schools to promote their children to the next grade. Household wealth should also be relatively time-invariant (affecting past test scores as much as present test scores) so it should have little predictive power using the value-added model. We compute a principal components child wealth index based on 20 survey questions asking whether the child's household owned a variety of products.[25] Table A.5, column 2 shows results from our baseline specification plus the wealth index. The repetition coefficients remains similar in magnitude (differing by at most .002 standard deviations). Furthermore, as expected, the wealth index is a poor predictor of test scores, conditional on lagged test scores.

[25] These include beds, radio, television, refrigerator, bicycle, plough, small agricultural tools, tables, fans, tractor, cattle, goats, chicken, watches, motor rickshaw, motorcycle/scooter, car/taxi/van/pickup, telephone, and tubewell

Table A.5: Effect of repetition on test scores - Considering household wealth

	Math		English		Urdu		Dropout	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Not Promoted	**-0.468*** **	**-0.468*** **	**-0.313*** **	**-0.313*** **	**-0.400*** **	**-0.402*** **	**0.074*** **	**0.074*** **
	(0.054)	(0.054)	(0.037)	(0.037)	(0.042)	(0.041)	(0.015)	(0.015)
Lagged Score	**0.501*** **	**0.501*** **	**0.440*** **	**0.440*** **	**0.493*** **	**0.492*** **	**-0.008**	**-0.008**
	(0.026)	(0.027)	(0.023)	(0.023)	(0.024)	(0.024)	(0.007)	(0.007)
2yr Lagged Score	**0.262*** **	**0.262*** **	**0.213*** **	**0.213*** **	**0.249*** **	**0.251*** **	**-0.017** **	**-0.017** **
	(0.027)	(0.027)	(0.020)	(0.020)	(0.027)	(0.027)	(0.008)	(0.008)
Child Wealth		**-0.002**		**-0.001**		**-0.009** **		**-0.000**
		(0.005)		(0.004)		(0.004)		(0.002)
Constant	**0.775*** **	**0.774*** **	**0.641*** **	**0.641*** **	**0.661*** **	**0.659*** **	**0.019*** **	**0.019*** **
	(0.026)	(0.026)	(0.021)	(0.021)	(0.021)	(0.021)	(0.007)	(0.007)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	5543	5543	5543	5543	5543	5543	6239	6239
R^2	0.72	0.72	0.78	0.78	0.75	0.75	0.19	0.19

Note: This table replicates the results shown in Table 6. The first column for each outcome variable is the exact same regression as in the main text, however the sample size is restricted to students who have a measure of household wealth. The results are robust.

Second, child health affects both achievement and promotion decisions (malnourished children may be held back due to perceived immaturity and also perform worse due to lower energy/cognitive capacity), however it also should be relatively time-invariant (aside from short-term shocks). Again, table A.6 shows that the value-added model is successful: after inclusion of the health variables, coefficients change by at most 0.015 standard deviations and become *more negative* for all three subjects. Additionally, height/weight explain little of the variation in test scores.

Table A.6: Effect of repetition on test scores - Considering child health

	Math		English		Urdu		Dropout	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Not Promoted	**-0.463*** **	**-0.478*** **	**-0.309*** **	**-0.313*** **	**-0.389*** **	**-0.396*** **	**0.075*** **	**0.078*** **
	(0.055)	(0.056)	(0.037)	(0.037)	(0.042)	(0.042)	(0.015)	(0.015)
Lagged Score	**0.503*** **	**0.500*** **	**0.440*** **	**0.441*** **	**0.496*** **	**0.497*** **	**-0.007**	**-0.005**
	(0.026)	(0.027)	(0.023)	(0.023)	(0.024)	(0.024)	(0.007)	(0.007)
2yr Lagged Score	**0.261*** **	**0.266*** **	**0.213*** **	**0.212*** **	**0.249*** **	**0.250*** **	**-0.017** **	**-0.018** **
	(0.027)	(0.027)	(0.020)	(0.020)	(0.027)	(0.027)	(0.008)	(0.008)
Height		**-0.015**		**-0.012**		**-0.003**		**0.005**
		(0.012)		(0.009)		(0.009)		(0.003)
Weight		**-0.001**		**0.005**		**-0.015**		**0.008**
		(0.013)		(0.010)		(0.010)		(0.004)
Constant	**0.773*** **	**0.770*** **	**0.641*** **	**0.642*** **	**0.658*** **	**0.651*** **	**0.019*** **	**0.025*** **
	(0.026)	(0.027)	(0.020)	(0.021)	(0.021)	(0.022)	(0.007)	(0.007)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	5493	5442	5493	5442	5493	5442	6178	6123
R^2	0.72	0.72	0.78	0.78	0.75	0.75	0.19	0.20

Note: This table replicates the results shown in Table 6. The first column for each outcome variable is the exact same regression as in the main text, however the sample size is restricted to students who have a measure of child health. The results are robust.

A.1.2.3 Controlling for nonlinearities in teacher rating

We also show results are identical when including teacher rating as a categorical rather than continuous variable (only showed it as continuous in main table for convenience of exposition)

Table A.7: Main specification with factorized teacher rating

	Math		English		Urdu		Dropout	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Not Promoted	**-0.472*** **	**-0.440*** **	**-0.313*** **	**-0.278*** **	**-0.402*** **	**-0.371*** **	**0.073*** **	**0.070*** **
	(0.055)	(0.055)	(0.037)	(0.036)	(0.043)	(0.043)	(0.014)	(0.014)
Lagged Score	**0.502*** **	**0.467*** **	**0.439*** **	**0.404*** **	**0.496*** **	**0.467*** **	**-0.007**	**-0.005**
	(0.027)	(0.027)	(0.023)	(0.023)	(0.024)	(0.024)	(0.007)	(0.007)
2yr Lagged Score	**0.258*** **	**0.247*** **	**0.213*** **	**0.195*** **	**0.245*** **	**0.226*** **	**-0.018** **	**-0.018** **
	(0.027)	(0.027)	(0.020)	(0.020)	(0.027)	(0.027)	(0.009)	(0.008)
Child Talent								
= 1		**-0.122**		**-0.060**		**-0.114** **		**0.027**
		(0.066)		(0.050)		(0.056)		(0.019)
= 2		**-0.096**		**-0.098** **		**-0.066**		**-0.001**
		(0.054)		(0.040)		(0.043)		(0.013)
= 3		**-0.120*** **		**-0.052**		**-0.094** **		**0.021**
		(0.045)		(0.036)		(0.037)		(0.014)
= 4		**-0.071**		**-0.019**		**-0.036**		**0.003**
		(0.039)		(0.028)		(0.031)		(0.009)
= 6		**0.026**		**0.052** **		**0.052** **		**0.004**
		(0.030)		(0.024)		(0.024)		(0.007)
= 7		**0.033**		**0.081*** **		**0.049** **		**0.001**
		(0.030)		(0.025)		(0.024)		(0.007)
= 8		**0.093*** **		**0.133*** **		**0.102*** **		**0.000**
		(0.033)		(0.025)		(0.026)		(0.008)
= 9		**0.123*** **		**0.166*** **		**0.150*** **		**-0.007**
		(0.038)		(0.030)		(0.029)		(0.009)
= 10		**0.176*** **		**0.234*** **		**0.166*** **		**-0.009**
		(0.040)		(0.032)		(0.031)		(0.010)
Age/Gender Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag Polynomial	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order	4th Order
N	5529	5529	5529	5529	5529	5529	6222	6222
R^2	0.72	0.72	0.78	0.78	0.75	0.75	0.19	0.19

Note: This table replicates the results shown in Table 6. The results are robust.