**WORKING PAPER** · NO. 2026-20

# Interpreting Performance: Evidence on Signal Weighting in Human Capital Investment

*Derek Rury and Ariel Kalil*

JANUARY 2026

THE UNIVERSITY OF CHICAGO
BECKER FRIEDMAN INSTITUTE FOR ECONOMICS

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

# Interpreting Performance: Evidence on Signal Weighting in Human Capital Investment

Derek Rury (Oregon State University)
Ariel Kalil (University of Chicago)*

This version: January 30, 2026

**Abstract**

Parents invest in children's human capital based on signals of academic performance, but we do not know how they weigh each when having perfect information or when they contain conflicting signals. Using 23,321 investment decisions from a survey experiment with 2,079 U.S. parents, we provide the first evidence on how parents trade off grades against standardized test scores. Both signals affect investment: parents adopt compensatory strategies, investing more when either signal indicates poor performance. Parents also put a higher weight on grades than tests, on average. However, we document asymmetric crowd-out: when grades are high but test scores are low, parents do not invest—high grades crowd out the response that low test scores would otherwise trigger. When grades are low but test scores are high, parents invest. This asymmetry implies that grade inflation imposes costs beyond direct signal distortion by preventing remedial investment in struggling students. Hispanic parents exhibit particularly pronounced grade-weighting. Our findings suggest that information interventions providing test scores will have attenuated effects when parents already possess inflated grade information.

*Key Words*: Beliefs, Preferences, Investments
*JEL Codes*: I21, J16, D83, D91

*Email: ruryd@oregonstate.edu; akalil@uchicago.edu.
The AEA registry for this project can be found using the following ID: AEARCTR-0014945. This work was supported by funding from the Sijbrandij Foundation to the Center for Incentive Design at the Anderson School of Management, UCLA. We also thank seminar participants at the School of Public Policy at Oregon State University.

# 1 Introduction

Parents who seek to invest in their children's human capital face a fundamental inference problem: they cannot observe ability directly but must infer it from available signals. In the United States, two primary signals reach parents—teacher-assigned grades and standardized test scores—and these signals often conflict. We provide the first evidence on how parents weigh grades against test scores when making investment decisions. Using an online survey experiment with 2,079 U.S. parents who evaluated hypothetical children with varying grade and test-score combinations. We find that parents place more weight on grades than test scores when making investment decisions. We also find a striking asymmetry: high grades crowd out the investment response that low test scores would otherwise trigger, but low grades do not crowd out the response to high test scores. This asymmetric crowd-out pattern implies that grade inflation has welfare costs beyond direct signal distortion: it causes parents to underinvest in children whose struggles are masked by inflated grades.

How parents weigh competing signals has grown increasingly consequential. Grade inflation has accelerated in U.S. schools: average GPAs have risen even as test scores stagnated (Gershenson, 2018; Goldhaber and Goodman Young, 2024). The COVID-19 pandemic amplified this divergence, with documented learning losses of 0.2 standard deviations masked by grade distributions that failed to reflect these declines. Simultaneously, standardized testing is increasingly scrutinized. Both signals are contested, yet parents must rely on them. A recent survey found that among parents whose child receives a B but scores below grade level on a standardized test, 40 percent would "not worry" about the discrepancy (Learning Heroes, 2022). Understanding which signal dominates when the two conflict is essential for predicting how information interventions affect human capital accumulation.

Our approach presents both signals simultaneously to each respondent, allowing us to estimate the independent effect of each signal type while holding the other constant. This design represents a key methodological contribution: prior studies examining parental responses to academic information have manipulated a single signal (Dizon-Ross, 2019; Bergman, 2021; Cobb-Clark et al., 2021), leaving open how parents integrate multiple signals that point in different directions. Our final dataset comprises 23,321 investment decisions from 2,079 parents, each evaluating scenarios involving fifth-grade children with experimentally varied grades (A through F) and test scores (10th through 90th percentile). We document three main findings. First, parents respond to both grades and test scores: lower performance on either signal increases recommended investments, consistent with compensatory strategies. On average, a one-unit decrease in grades (e.g., from B to C) increases recommended investment by 0.14 standard deviations; a one-unit decrease in test score percentile category increases investment by 0.12 standard deviations. Both effects are precisely estimated and economically meaningful.

Second, we find clear evidence that parents place more weight on grades than test scores when making investment decisions. In our heterogeneity results, we fail to find that this is driven by parental income or education, perhaps due to low power. However, we do find that Hispanic parents have a strong preference for grades over test scores.

Third, and central to our contribution, we find asymmetric responses to conflicting signals. When test scores are high but grades are low, parents invest: they treat the low grade as actionable, recommending additional time and money for academic support. When grades are high but test scores are low, parents do not invest: their recommendations are statistically indistinguishable from the reference group with concordant average performance. High grades crowd out the investment response that low test scores would otherwise trigger. This pattern implies that parents receiving inflated grades will fail to make remedial investments that their children's actual achievement levels warrant.

Our approach relies on controlled laboratory-style survey data to study parental investment decisions because such data allow us to isolate informational mechanisms that would be extremely difficult to observe in real-world settings. This might make some readers skeptical. However, in natural environments, grades and test scores are jointly determined by schools, teachers, and parental responses themselves, making it nearly impossible to separately vary these signals while holding other factors constant. The experimental structure of our design, however, enables precise manipulation of grade and test information, yielding clean comparisons across informational regimes.

To mitigate concerns about external validity, we incorporate several design features. First, we include individual fixed effects, thereby identifying responses from within-parent variation rather than cross-parent differences. This approach controls for stable individual characteristics that might affect both survey responses and real-world behavior. Second, we restrict the analysis to responses that fall within a plausible range of real-world investment behavior (time investments under 7 hours per day; monetary investments under $81 per week), ensuring that the estimates reflect decision patterns that could feasibly arise outside the laboratory. Third, our comprehension checks allow us to verify that respondents understood the scenarios; over 75 percent answered questions about percentile interpretation correctly, and our results remain robust after controlling for comprehension. Fourth, prior research using similar hypothetical-scenario methodologies has demonstrated that stated preferences in such designs correlate meaningfully with actual parental behavior (Cunha et al., 2013; Boneva and Rauh, 2018; Giannola, 2024). Together, these design choices allow us to combine the internal validity of controlled experimentation with safeguards that enhance the relevance of our findings to real-world parental decision-making. We acknowledge, however, that respondents are providing advice about hypothetical children rather than making decisions for their own children, which may introduce some divergence from actual investment behavior.

This paper contributes to three literatures and policy areas. First, we extend the literature on parental beliefs and information frictions by examining responses to multiple simultaneous signals—the situation parents face in practice—rather than to a single experimentally provided signal. Dizon-Ross (2019) shows that Malawian parents overestimate their children's performance and adjust investments when provided objective information. Bergman (2021) documents that U.S. parents overestimate effort and standing;

providing accurate information improves outcomes. Cobb-Clark et al. (2021) find that Australian parents increase tutoring by 15 percent when test scores reveal unexpectedly negative information. Our finding that grades crowd out test score responses implies that information interventions providing test scores alone will have attenuated effects when parents already possess inflated grade information.

Second, we contribute to theoretical models of belief formation under multiple information sources. In canonical frameworks (Cunha and Heckman, 2007; Cunha et al., 2010), parents observe skill directly. In practice, they infer skill from noisy signals. Kinsler and Pavan (2021) document "local distortions": parents anchor beliefs to local rather than national distributions. Our crowd-out finding represents a different distortion: when a familiar, frequently received signal conflicts with a less familiar signal, the familiar signal dominates regardless of relative accuracy.

This pattern is consistent with insights from behavioral economics regarding the availability heuristic (Tversky and Kahneman, 1973): signals that are more frequently encountered and more easily recalled may receive disproportionate weight in decision-making. Grades are reported regularly throughout the school year via report cards and teacher communications, whereas standardized test scores are typically reported annually. This frequency differential may contribute to grade primacy independent of the signals' relative informativeness about underlying academic skill. Our findings have implications for models incorporating multiple information sources with heterogeneous salience.

Third, our results have direct policy implications. Grade inflation's costs extend beyond signal distortion: if high grades crowd out responses to low test scores, inflation insulates struggling students from remedial investment. This crowd-out mechanism implies that combating grade inflation may be more consequential for parental behavior than expanding test score dissemination. At the same time, our finding that test scores trigger investment when grades are concordantly low suggests that maintaining standardized testing preserves a valuable information channel—one whose influence depends on the quality of grade information parents already possess.

The paper proceeds as follows. Section 2 develops our theoretical framework and reviews evidence on parental beliefs, information frictions, and grade inflation. Section 3 describes the survey experiment. Section 4 presents summary statistics. Section 5 reports our main results: parents weigh both signals, adopt compensatory investment strategies, and exhibit asymmetric crowd-out when signals conflict. Section 6 documents exploratory heterogeneity analyses by parent race and ethnicity. Section 7 examines mechanisms. Section 8 concludes with policy implications.

# 2 Background

A substantial literature documents that parents hold systematically biased beliefs about their children's performance. Dizon-Ross (2019) finds that Malawian parents believe their child ranks at the 73rd percentile on average—an overestimation of 23 percentile points. Providing accurate information shifts investments, with effects concentrated among parents who previously overestimated. Bergman (2021) documents similar frictions in U.S. schools: parents overestimate effort and standing, and real-time information about grades and absences increases engagement and improves outcomes. These effects operate through belief updating: parents who learn of poor performance respond by investing more, consistent with compensatory motives.

Kinsler and Pavan (2021) identify a mechanism for belief distortion: local projection. Parents anchor beliefs about national standing to local distributions within their child's school. At schools with below-average achievement, parents overestimate their child's position by approximately 15 percentile points. This distortion has investment consequences: parents who overestimate spend less time in educational activities than would be optimal. Because grades are inherently local signals—reflecting performance relative to classmates—while test scores provide national comparisons, the local distortions framework suggests grades may systematically mislead parents about true standing.

Our study builds on and extends this theoretical foundation by examining a distinct but related phenomenon: asymmetric signal weighting when multiple performance indicators conflict. While Kinsler and Pavan (2021) document how local reference points distort beliefs about a single dimension of performance, we ask how parents reconcile two signals that may point in different directions. The local distortions framework predicts that grades (a local signal) should be more susceptible to inflation and less informative about national standing than test scores. Our design allows us to test whether parents behave as if they recognize this distinction or instead exhibit systematic preferences for one signal type regardless of its objective informativeness.

Boneva and Rauh (2018) show that disadvantaged parents are particularly susceptible to belief inaccuracies, suggesting information frictions contribute to inequality. Ziege and Kalil (2025) confirm that parents update beliefs and investments in response to report cards, but the magnitude depends on information modality and teacher beliefs. Collectively, this literature establishes that parental beliefs are malleable, consequential for investment, and systematically distorted—but no prior study examines how parents reconcile conflicting signals.

The informational value of grades has been compromised by inflation. Gershenson (2018) documents that high school GPAs rose substantially between 2005 and 2016 even as test scores remained flat. Tyner and Gershenson (2020) distinguish static inflation (grades overstating performance at a point in time) from dynamic inflation (deterioration of the grade-performance relationship over time). Both forms reduce signal quality. Goldhaber and Goodman Young (2024) show that COVID-19 accelerated this pattern: the share of students receiving A's increased even as test scores declined precipitously. Parents observing high grades may have been systematically misled about pandemic-era learning losses visible in test data.

Inflation is not uniform: it is more pronounced in schools serving disadvantaged populations (Silva et al., 2025; Tyner and Gershenson, 2020). Parents already susceptible to local distortions receive the most inflated signals. Beyond inflation, grades incorporate effort, behavior, and compliance in addition to mastery (Goldhaber and Goodman Young, 2024), introducing noise for parents attempting to infer the human capital stock that predicts future success. Test scores, designed for cross-school comparison, are less susceptible to local inflation—but face skepticism regarding cultural bias and appropriateness. Parents must navigate both imperfect signals.

The experimental literature establishes that parents respond to information. Cobb-Clark et al. (2021) exploit Australia's NAPLAN rollout and find that test score releases increase tutoring by 15 percent among parents receiving negative news. Andrabi et al. (2017) show that Pakistani parents respond to school report cards by increasing teacher engagement and switching schools. Bergman and Chan (2021), Kraft and Rogers (2015), and Barrera-Osorio et al. (2020) document that reducing information frictions improves outcomes across contexts.

No prior study examines how parents weigh multiple signals when those signals conflict. Dizon-Ross (2019) varies test information holding grades constant; Cobb-Clark et al. (2021) study test scores in isolation; Bergman (2021) transmits grade information without examining reconciliation with test data. Each makes important contributions, but none addresses how parents trade off across signal types. This gap matters because parents receive both signals in practice. If grades dominate when the two conflict, information interventions providing test scores will have limited effects, and grade inflation will impose costs beyond direct distortion by preventing remedial investment in struggling children.

# 3 Survey Experiment Details

We conducted our experiment on the survey platform Prolific, which is often used in theoretical (McGranaghan et al., 2024; Nielsen and Rehbeck, 2022) and experimental (Coffman et al., 2024; Coffman and Klinowski, 2025; Bordalo et al., 2024) work in the social sciences. The complete survey is available in our online appendix. All participants signed their consent to participate in the project and were paid $3.15 to take our survey. After signing the consent form, we asked participants questions about their family life, such as how many children they have and their education levels. We used these results to screen out non-parents from the study. We then presented parents with instructions for the survey, which consisted of three phases.

In the first phase, parents were presented with scenarios of hypothetical children who were enrolled in fifth grade.[1] We asked parents to imagine that the child was enrolled in a school similar to the participant's child.[2] We then asked participants to offer advice on how much the family of this hypothetical child should invest in their

[1]While we acknowledge that a subset of children do not yet receive letter grades in fifth grade, letter grades are ubiquitous in most countries around the world and we make the weak assumption that most parents are aware of how they function.

[2]This may have been confusing for parents whose children had not yet reached fifth grade. There was no text for parents to imagine what school their child would attend (e.g. schools in their neighborhood).

child's academic progress. In order to provide a sense of the child's academic ability, in each scenario, we exogenously provided participants information about the child's grades and a standardized test score. Grades were presented as letter grades (e.g. A, B, C, D and F) and results from tests were presented in percentile rank points, meaning they went from 1 to 100, with a 1 representing the lowest score and a 99 representing the highest score.

We presented parents with five percentile ranks: 10, 30, 50, 70, and 90. We also informed parents that these test scores compared the child's results with those of other children nationwide. Since the early 2000s, standardized tests have become nearly universal in U.S. public education due to federal mandates like the No Child Left Behind Act (2001), which required annual testing in reading and math for students in grades 3 through 8. These tests were intended to provide objective, comparable measures of student achievement and to hold schools accountable for performance across all demographic groups. As a result, by the mid-2000s, all 50 states had implemented statewide assessments, creating a national landscape in which standardized testing became a defining feature of elementary and secondary education. One such example is the Smarter Balanced Assessment Consortium (SBAC), which is currently used 19 states and offers feedback in the form of individual comparisons other children of the same age from this cohort.

As part of the experiment, we also varied the child's name in the scenarios from Robert to Stacey to capture differential treatment for boys and girls. The universe of scenarios totaled 25 for each sex. Participants saw three randomly selected grades and test results and were asked to advise on how much the child's parents should invest in the child's academic progress, yielding 12 scenarios per participant. Parents answered an equal number of advice questions for both Robert and Stacey.

In each scenario, participants were asked to advise the hypothetical child's parent on how much time and money to invest in the child's academic progress per week. For monetary investments, parents were informed that the family had up to $100/week available for this decision. To prevent participants from providing unmotivated answers, we also mentioned that the family could use the money not spent in investing in their child for other purposes, such as taking a vacation or paying off bills. For time investments, participants were asked to report time investments per day; responses were capped at 24 hours. We also gave examples of time investments, such as spending time helping with homework or reading to the child.

In the third phase, we asked participants questions about their feelings about grades and standardized tests across several dimensions. These questions elicited whether parents view tests as biased against certain groups; whether parents felt like a bad parent if their child gets bad grades (test scores); whether parental investments will pay off more if the child is a good student; whether spending on academic support makes the most difference for children who are strong students; whether they were more likely to invest in educational resources if their child is a weak student; whether the child's behavior in

We therefore assume that parents were making this assumption.

the classroom affects the grades he or she gets; whether tests are better than grades at measuring a child's academic skills; whether grades are a better reflection of their child's social and emotional skills than tests; whether tests mostly reflect parents' education and income; whether children who are struggling with class material benefit more from parental involvement in their school activities; and whether the parent tends to give more guidance on homework to their child if they already have a strong understanding of the material. Responses to these questions could be "agree", "neither agree nor disagree" or "disagree" for each statement. We also asked parents how they valued grades and tests when making investment decisions for their own child where parents could indicate the numerical value between 0 and 1 that represented whether standardized tests are more informative about a child's skill.

One concern is that participants had no incentive to provide truthful answers to our hypothetical advice scenarios. If this was the case, it would be difficult to estimate how parents' responses to grade and test information and tell a credible story about their weights for either signal. To address this, we administered comprehension checks throughout the survey to assess the validity of participants' responses. These took the form of asking participants to answer the previous question. In Section 5, we present results that control for participants' propensity to answer these questions correctly. In addition, we also limit parental responses to those that appear reasonable in our analyses.

# 4 Data and Descriptive Results

To participate in our survey, participants must meet these criteria: be a parent, live in the US, be between 25 and 50 years old, and have had their first child born between 2000 and 2010. To ensure our sample was evenly distributed across the income spectrum, we created four income brackets and corresponding surveys in Prolific: $0-29,999$, $30,000-$59,999, $60,000-$89,999, and $90,000 and above. We then recruited 500 participants from each bracket. Ultimately, we collected over 23,321 hypothetical investment decisions in response to grade and exam information. The median time to complete our survey was 10 minutes and 40 seconds. Just over 97 percent of participants who started the survey completed it. Table 1 presents descriptive results from our sample of parents.

We observe that most of our sample was white (68.4%), followed by 17.8% Black, 7% Hispanic, and 2.3% Asian. The average age was 41.2 years, and just over half (53.5%) held a college degree. Nearly 70% were mothers, and the average age of the respondents' oldest child was 17.5 years. Additionally, 41.8% had household incomes below $60,000. Participants recommended that parents spend $43.4 per week and a little over four hours a day on their child's academic progress.

Two key points emerge from the descriptive analyses. First, we observe that parents recommend lower investments for higher-achieving children and higher investments for lower-achieving children, suggesting that they view investments as compensatory rather than complementary to achievement. We also notice outliers, especially in time investments. Given that there are only 24 hours in a day, it seems unreasonable for parents to devote more than 6 hours per day to their child's academic progress. Additionally, many participants select "50" or "100," for monetary investments, for example, which might be genuine responses but could also be due to participant error. To prevent these outlier investment amounts from skewing our estimates, we restrict the sample to respondents who indicate reasonable investment amounts as a robustness check.

# 5 Main Results

To study how parents value information contained in grades and exams and whether they prefer one over the other when making investment decisions, we estimate the following linear model using ordinary least squares (OLS):

$$investment\,advice_{ic} = \alpha + \beta_1 grade_{ic} + \beta_2 test_{ic} + \boldsymbol{X}\gamma + \epsilon_{ic}$$

where *investment advice* is the advice given by the participant, *grade* is the grade the hypothetical child receives, *test* is the exam performance percentile the child receives, and $\epsilon$ is a random error term. *investment advice* is comprised of responses to the money and time scenarios, which are measured on two different scales. For analysis, we standardize both response types to have a mean of 0 and a SD of 1 and combine them to maximize power. Later, we study how results differ depending on whether the scenario asked for money or time advice.

We first estimate $\beta_1$ and $\beta_2$ without controlling for demographic characteristics. We also estimate two additional models - the first where we include respondent fixed effects which control for any level differences in responses within subjects. And the second which includes demographic controls and clusters at the participant level which allows for correlation in the error term within subjects. Both $grade_c$ and $test_c$ are ordinal variables, representing letter grades, in the case of grade and performance quintiles (e.g. 30th percentile) in the case of tests. This choice requires some defense, as ordinal variables are not typically used in regression analysis. For test scores, the ordinal orientation of the variable represents a continuous percentile score, so we may view the ordinal measures as capturing a true continuous variable (i.e. percentile scores). For grades, the assumptions required to include these in the regression are a bit stronger. The most generous view of grades is that they represent a mapping of performance – which is multivariate (e.g. effort, behavior, etc.) and continuous – onto ordinal measures. Below, we also create indicators for receiving bad grades, for example, and perform a similar analysis. Our results from this exercise are largely qualitatively identical to those from our preferred specification. When interpreting our results, coefficients represent a unit increase in either grades A = 1, B = 2, C = 3, D = 4, F = 5) or tests (10th = 1, 30th = 2, 50th = 3, 70th = 4, 90th = 5) on the SD in investment advice. We interpret estimates of $\beta_1$ and $\beta_2$ as parental weights for information contained in grades and tests. For each model, we also test the equality of $\beta_1$ and $\beta_2$. This test tells us whether parents, on average, prefer one signal over the other.

Results are presented in Table 2 and provide evidence for three points. First, both in panel A and B, we see from columns 1-3 that $\beta_1$ and $\beta_2$ are statistically different from zero, indicating that parents change their investment decisions in response to grade and exam information. Second, all coefficients are positive, indicating that parents in our sample view investments as compensatory rather than complementary to students, consistent with our earlier descriptive results. Third, for each pair displayed in the table, the p-values from tests of equality between $\beta_1$ and $\beta_2$ are all small, indicating that parents prefer grades to tests when making investment decisions.

As mentioned in section 3, we suspect some responses provided by participants are not truthful. This could be the case because participants were not incentivized to provide truthful responses. We therefore limit our sample to those participants who responded with time investments less than seven and monetary investments less than 81. For time investments, this represents nearly 90% of our sample and for monetary investments just over 86% of our sample. Results can be found in Table 3. We see here that results remain largely unchanged by making this restriction.

We also add a control that indicates whether respondents answered the first comprehension check correctly. This question presented participants with a hypothetical child's test results, presented in percentile terms, and asked what percentage of students performed better than this student. Just over 75 percent of participants answered this correctly. We view this variable as a proxy for participant interest or dedication in completing the survey truthfully. Appendix table A.1 presents the results which include this control. Again, we see that there is no difference between these results and our primary specification.

To investigate the existence of nonlinearities in the impact of receiving grade and test signals, we create indicator variables which capture the actual grade and test outcome and include these indicators in a single regression. Table 4 shows presents these results where receiving an "A" and performing in the 90th percentile on the test are the omitted categories. Again, we see that investment increases as the child receives lower grades and scores. We also note that there are no noticeable jumps, indicating a smooth continuous response.

An effective way to evaluate how parents weigh information from grades versus standardized test scores is to analyze cases where the two signals conflict. While earlier models estimate average weights for grades and tests across all situations, they do not directly show how parents resolve disagreements between signals when one indicates high performance and the other suggests low performance. If parents prioritize grades more than test scores, this preference should be especially noticeable in such conflict cases.

To study this, we classify each vignette into four mutually exclusive categories based on whether grades and test scores indicate relatively high or low performance. We define grades of A or B as high and grades of D or F as low; similarly, test scores at the 10th or 30th percentile are classified as low, while scores at the 70th or 90th percentile are classified as high. This yields four groups: (i) grades high and tests low, (ii) grades low and tests high, (iii) both signals high, and (iv) both signals low. We exclude cases with a C grade or a 50th-percentile test score from the analysis to focus on situations with clear signals. This analysis was not prespecified in our pre-analysis plan and, therefore, should be considered exploratory.

Table 5 reports the results. Across specifications, we find strong evidence that parents respond asymmetrically to conflicting information. When grades are low, and tests are high, parents recommend significantly higher investment levels than when grades are high, and tests are low. In the baseline OLS specification, the coefficient on the "grades bad, tests good" category is positive and statistically significant, while the coefficient on the "grades good, tests bad" category is smaller in magnitude and not statistically distinguishable from zero. An F-test rejects equality between the two coefficients at conventional levels (p = 0.0026).

This asymmetry persists when clustering standard errors at the respondent level and becomes even more pronounced in the specification with respondent fixed effects. In the fixed-effects model, investment advice is significantly lower when grades are high and tests are low, while the opposite conflict produces a smaller and statistically insignificant response. The null hypothesis that the two conflict coefficients are equal is again strongly rejected (p = 0.0002). These results indicate that, within the same parent, investment advice tracks grade information more closely than test information when the signals disagree.

# 6 Heterogeneity

We next present our heterogeneity analysis for each of our pre-registered characteristics. To study heterogeneity, we estimate the following equation:

$$Y_i = \alpha + \beta_1 \text{Grade}_i + \beta_2 \text{Test}_i + \beta_3 C_i + \delta_1 (\text{Grade}_i \times C_i) + \delta_2 (\text{Test}_i \times C_i) + \boldsymbol{X}\gamma + \epsilon_i$$

where $C_i$ represents parental demographic characteristics (e.g. parent race). To test whether parents of a given characteristic prefer information contained in grades and exams differently, we compare $delta_1$ and $delta_2$ which capture the differential impact for grades and tests for each subgroup. We also present the p-value of whether these coefficients are equal. Given the modest sample sizes for some subgroups (particularly Hispanic parents, n=145, and Asian parents, n=48), these results should be interpreted with appropriate caution and viewed as hypothesis-generating rather than definitive.

Table 5 presents the results. Consistent with our prior results, we find that $\beta_1$ and $\beta_2$ are positive and statistically significant for each model. We fail to reject the null hypothesis of no difference between signals for low-income, white, black, and low-educated parents. In addition, we find that parents do not perceive the signals differently with respect to investments in girls versus boys.

For both our continuous and binary models, we find suggestive evidence that Hispanic parents (n=145) prioritize grades over tests when making investment advice decisions. The interaction coefficient for Hispanic parents is negative for tests but not for grades, indicating a differential response; however, given the small sample size, these findings should be interpreted cautiously. We also observe a large but statistically insignificant preference for tests over grades among Asian parents (n=48). The very small sample size among Asian parents limits the strength of conclusions, though the pattern is suggestive and may warrant further investigation with larger samples.

# 7 Mechanisms

There may be several reasons why parents prefer grades over test results when making their investment decisions. These drivers could enhance the utility of grades, for example, by capturing socio-emotional skills that tests do not. They can also highlight that tests are inferior to grades, for example, by asserting that tests are biased against certain groups. To test whether any of these mechanisms are driving our results, we explore how parents respond to survey questions studying these topics.

As outlined earlier in Section 2, we asked parents several questions about how they view the content of grades and test results and whether they consider them reliable. Nearly 40% of parents believe that tests are biased against certain groups. We also find that 50% of parents feel distressed when their child receives poor grades, while just over 40% feel the same when their child earns poor test results. 27% of parents see tests as reflecting a family's income, and nearly 18% report that tests better capture students' skills than grades. 46% agree that grades better reflect emotion than tests, whereas 76% believe grades better reflect behavior than tests. When asked how important grades and tests are in making decisions for their own children, 71% responded that grades are more important, compared to just 8.5% who said the same for tests.

One question worth considering how parents responses to this questions could be used to study potential mechanisms. It is possible that parents suffer from one of several biases that influence self-reported accounts of their *own* behavior (Grimm, 2010). This motivated our choice to present parents with hypothetical scenarios from children who not only were not their own but did not exist. As mentioned earlier, this approach is often used in research studying parents and their beliefs about their children (Cunha et al., 2013; Boneva and Rauh, 2018; Giannola, 2024).

To explore what might be driving our effects, we interact the mechanism with both grade and test signals to see if the effects differ and if the interaction coefficients are statistically distinct. This method helps us identify which factors could serve as mechanisms behind the higher weights for grades over exams. Although we have a comprehensive set of measures capturing parents' views on grades versus exams, we generally fail to reject the null hypothesis that the effects are the same for each of our potential mechanisms. The only interaction that is marginally significant (p = 0.087) is for parents who state they prefer grades over exam when making decisions. Our results show that their investment choices align with this stated preference for grades over test scores when investing in their own children.

Several explanations for these null results are possible. First, our measures may be subject to measurement error if parents' stated beliefs about grades and tests do not accurately reflect their decision-relevant beliefs. Second, the true mechanisms may not be among those we measured; for instance, the availability heuristic discussed in Section 2 would operate through signal familiarity rather than through explicit beliefs about signal quality. Third, multiple mechanisms may operate simultaneously with offsetting effects, obscuring their individual contributions. Future research should address these possibilities.

# 8 Conclusion

Parents often make educational investment decisions, such as helping with homework or paying for tutoring, based on limited and sometimes inaccurate information about their child's academic performance. Our study analyzes data from nearly 24,000 hypothetical investment decisions collected through an online survey experiment to understand how parents interpret two standard signals of academic ability: grades and standardized test scores.

We find that both grades and test scores significantly influence parents' beliefs and investment decisions. On average, parents respond in a compensatory manner: they tend to endorse spending more time and money when a child is struggling. However, parents consistently prefer grades over test scores when deciding whether to invest. This higher weight for grades over test scores remains strong across different demographic groups and is robust across multiple model specifications. Heterogeneity analyses suggest that Hispanic parents (n=145) may show a particularly strong tendency to rely on grades rather than test scores, though given the modest sample size, this finding should be interpreted cautiously. Asian parents (n=48) show a suggestive but imprecisely estimated opposite pattern.

We also find evidence of crowd out of test scores in parental investment decisions. When grades and test scores provide conflicting information, parents are more like to invest when the child receives a poor grade and a good test score. Conversely, they are less likely to invest if the child receives a *good* grade with a poor test score.

We examined several possible reasons for this focus on grades. Many parents in our survey expressed doubts about standardized tests, often mentioning concerns that these tests are unfair, culturally biased, or reflect family background more than actual ability. Emotional reactions also varied: some parents reported anxiety or disappointment in response to low test scores, whereas grades were typically experienced as more familiar and positive feedback. However, we found that none of these self-reported beliefs fully explain why parents prefer grades. Parents may value tests selectively; for instance, they might pay more attention to high-stakes exams like the SAT/ACT as college approaches. Overall, our results suggest that parents tend to pay less attention to information from standardized test scores, even though those scores often provide a more objective measure of skills than grades. This pattern reveals an informational failure in the education market.

If parents rely primarily on grade reports, which can be influenced by inflation or inconsistent standards, they may develop overly optimistic views of their child's achievement. A recent survey revealed that 90% of parents of fourth graders believe their child is performing at grade level in math, based on report cards; however, only about 40% of those children demonstrate proficiency on an independent national test (CRPE, 2025). This mismatch can lead to parents' underinvestment in children's human capital, thereby contributing to a "leaky pipeline" in education.

This might explain differences in participation at higher levels of education. For example, Asian American students are over-represented in advanced high school math classes while Hispanic students are under-represented (CRPE, 2025). Our own analysis (available on request) of the American Time Use Survey years 2003-2024 shows that Hispanic parents of young children spend significantly less time on active educational care than White or Asian parents. These differences in parental effort, along with misconceptions about achievement, can worsen existing inequalities in skill development. By 2022, Hispanic children scored much lower than their White and Asian peers on national math and reading tests (NAEP, 2024), pointing to potential long-term costs of information gaps in academic success.

From an economic perspective, the divergence between grades and exam scores can be viewed through the lens of information asymmetry and signaling. In education markets, grades and tests serve as signals of a student's ability or knowledge. Suppose one signal (report card grades) becomes inflated or noisy relative to the actual level of achievement. In that case, both families and downstream decision-makers (e.g., high schools, colleges, or employers) face a challenge in accurately assessing a student's ability. Our results suggest that parents, as key investors in their children's human capital, may be misled by an overly optimistic signal. This misallocation of effort and resources can exacerbate achievement gaps and reduce the effectiveness of educational spending. In this light, improving the reliability and transparency of performance information emerges as a policy priority. Reliable signals are crucial for efficient investment: when signals fail, even well-intentioned parents may under-invest in their children's education, with lasting consequences for skill formation and economic opportunity.

Lastly, our study contributes to the growing literature on how parents form beliefs and make investment decisions in response to signals of ability. Previous work has highlighted the local distortion of beliefs (Kinsler and Pavan, 2021) and the role of test results in shifting parental perceptions and behavior (Cobb-Clark et al., 2021). Our findings corroborate these insights but also suggest an overlooked mechanism, particularly given the rise in grade inflation: the crowding out of test results by potentially inflated grades. Parental investment decisions are crucial for children's human capital. Policymakers should help make clear to parents which signals are most predictive of children's success.

## References


Andrabi, T., J. Das, and A. I. Khwaja (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review 107*(6), 1535–1563.

Barrera-Osorio, F., K. Gonzalez, F. Lagos, and D. J. Deming (2020). Providing performance information in education: An experimental evaluation in colombia. *Journal of Public Economics 186*, 104185.

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

Bergman, P. and E. W. Chan (2021). Leveraging parents through low-cost technology: The impact of high-frequency information on student achievement. *Journal of Human Resources 56*(1), 125–158.

Boneva, T. and C. Rauh (2018). Parental beliefs about returns to educational investments—the later the better? *Journal of the European Economic Association 16*(6), 1669–1711.

Bordalo, P., G. Burro, K. Coffman, N. Gennaioli, and A. Shleifer (2024). Imagining the future: memory, simulation, and beliefs. *Review of Economic Studies*, rdae070.

Cobb-Clark, D. A., T. Ho, and N. Salamanca (2021). Parental responses to children's achievement test results. Technical report, IZA Discussion Papers.

Coffman, K. and D. Klinowski (2025). Gender and preferences for performance feedback. *Management Science 71*(4), 3497–3516.

Coffman, K. B., M. R. Collis, and L. Kulkarni (2024). Whether to apply. *Management Science 70*(7), 4649–4669.

Cunha, F., I. Elo, and J. Culhane (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Technical report, National Bureau of Economic Research.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American economic review 97*(2), 31–47.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

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

Gershenson, S. (2018). Grade inflation in high schools (2005-2016). *Thomas B. Fordham Institute*.

Giannola, M. (2024). Parental investments and intra-household inequality in child human capital: Evidence from a survey experiment. *The Economic Journal 134*(658), 671–727.

Goldhaber, D. and M. Goodman Young (2024). Course grades as a signal of student achievement: Evidence of grade inflation before and after covid-19. *Journal of Policy Analysis and Management 43*(4), 1270–1282.

Grimm, P. (2010). Social desirability bias. *Wiley international encyclopedia of marketing*.

Kinsler, J. and R. Pavan (2021). Local distortions in parental beliefs over child skill. *Journal of Political Economy 129*(1), 81–100.

Kraft, M. A. and T. Rogers (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review 47*, 49–63.

Learning Heroes (2022, June). Hidden in plain sight: A way forward for equity-centered family engagement. `https://bealearninghero.org/research`. Parents 2022 — K-12 Parent, Teacher & Principal Survey Findings.

McGranaghan, C., K. Nielsen, T. O'Donoghue, J. Somerville, and C. D. Sprenger (2024). Distinguishing common ratio preferences from common ratio effects using paired valuation tasks. *American Economic Review 114*(2), 307–347.

Nielsen, K. and J. Rehbeck (2022). When choices are mistakes. *American Economic Review 112*(7), 2237–2268.

Silva, P. L., S. L. DesJardins, R. Biscaia, C. Sá, and P. N. Teixeira (2025). Public and private school grade inflation patterns in secondary education. *The BE Journal of Economic Analysis & Policy 25*(2), 305–342.

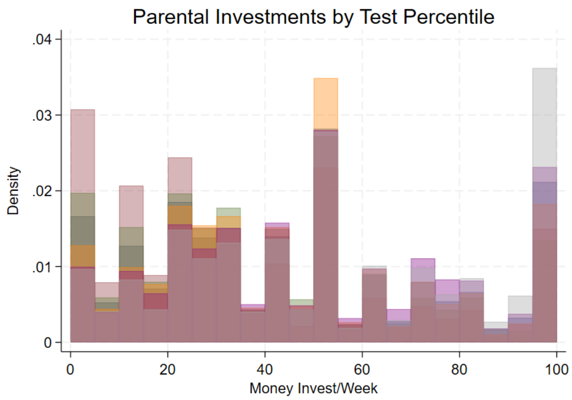
Tversky, A. and D. Kahneman (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology 5*(2), 207–232.

Tyner, A. and S. Gershenson (2020). Conceptualizing grade inflation. *Economics of Education Review 78*, 102037.

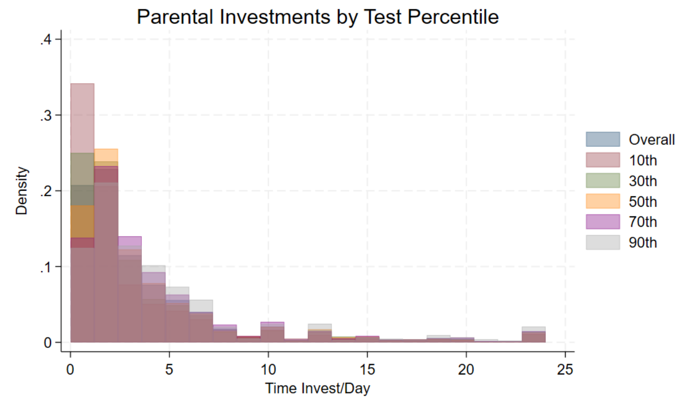
Ziege, E. and A. Kalil (2025). How information affects parents' beliefs and behavior: Evidence from first-time report cards for german school children. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2025-54).

# Figures and Tables

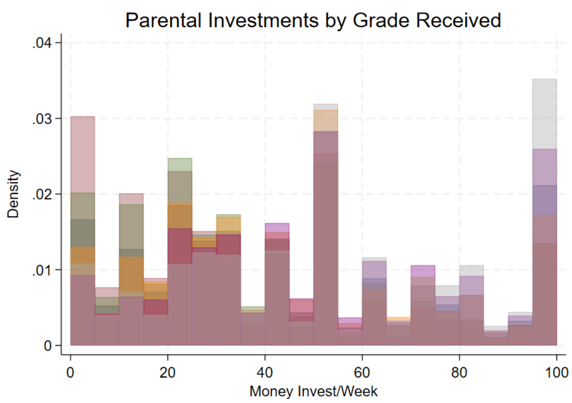
Figure 1: Investment Advice Amounts - Overall and By Grade



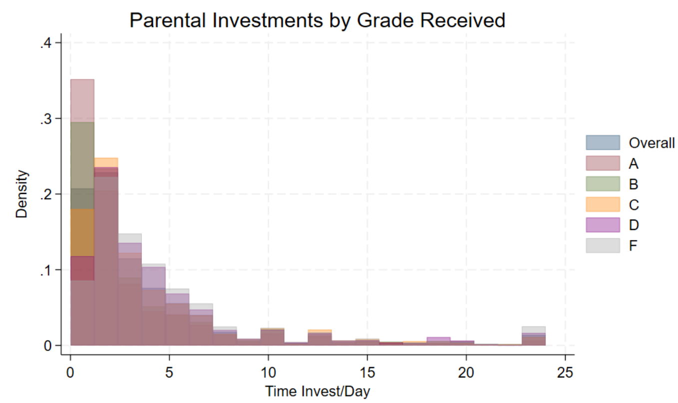

Panel A: Monetary investments per week




Panel B: Time investments per day




Panel C: Monetary investments per week




Panel D: Time investments per day

*Notes*: Each panel shows the density of responses to the hypothetical investment decisions in the vignettes, by signal type (grades vs. tests) for all scenarios as well as broken down by the value for grade and test scores in each scenario. Monetary values are weekly; time values are daily.

Table 1: Summary Statistics for Demographic and Outcome Variables

| **Variable** | **Count** | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|---|
| **Demographic Variables** | | | | | |
| White | 2079 | 0.684 | 0.465 | 0 | 1 |
| Black | 2079 | 0.178 | 0.383 | 0 | 1 |
| Hispanic | 2079 | 0.070 | 0.255 | 0 | 1 |
| Asian | 2079 | 0.023 | 0.149 | 0 | 1 |
| Age | 2048 | 41.209 | 7.612 | 0 | 83 |
| College Educ. | 2079 | 0.535 | 0.499 | 0 | 1 |
| Low Income (HH < $60k) | 2079 | 0.418 | 0.493 | 0 | 1 |
| Female | 2048 | 0.698 | 0.459 | 0 | 1 |
| Oldest Child Age | 2079 | 17.503 | 3.614 | 0 | 25 |
| Num Children | 2077 | 2.397 | 0.972 | 1 | 4 |
| **Outcome Variables** | | | | | |
| Money | 12402 | 43.424 | 29.890 | 0 | 100 |
| Time | 12360 | 4.023 | 4.541 | 0 | 24 |

*Notes*: All variables are based on self-reported survey responses. "Money" and "Time" represent hypothetical investments in response to vignettes. "College Educ." and "Low Income" are binary indicators for educational attainment and income level, respectively, both collected in the survey.

Table 2: Grades v.s. Test Results and Investment Advice from Parents

**Panel A: Continuous Specification**

| | No Controls | Participant FE | Controls+Clustering |
|---|---|---|---|
| Grades ($\beta_1$) | $0.140^{***}$ | $0.143^{***}$ | $0.142^{***}$ |
| | (0.0044) | (0.0034) | (0.0054) |
| Tests ($\beta_2$) | $0.124^{***}$ | $0.125^{***}$ | $0.126^{***}$ |
| | (0.0044) | (0.0034) | (0.0051) |
| $p$-value $\beta_1 = \beta_2$ | 0.0172 | 0.0003 | 0.0134 |

**Panel B: Binary Specification**

| | No Controls | Participant FE | Controls+Clustering |
|---|---|---|---|
| Grades = Bad ($\beta_1$) | $0.335^{***}$ | $0.344^{***}$ | $0.340^{***}$ |
| | (0.0110) | (0.0084) | (0.0130) |
| Tests = Bad ($\beta_2$) | $0.300^{***}$ | $0.305^{***}$ | $0.304^{***}$ |
| | (0.0110) | (0.0084) | (0.0126) |
| $p$-value $\beta_1 = \beta_2$ | 0.0047 | 0.0002 | 0.0009 |
| Observations | 23,221 | 23,221 | 22,865 |

*Notes:* Each panel shows estimated effects of grade and test information on parental investment decisions. Standard errors are shown in parentheses beneath the coefficients. Panel A reports linear models using continuous measures. Panel B restricts the sample to cases where either grades or tests indicate bad performance. The final row in each panel shows the $p$-value from a test of equality between the grade and test coefficients.

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 3: Effects of Performance Information on Investment with Restricted Investment Amounts

| | (1) OLS | (2) FE | (3) Clustered |
|---|---|---|---|
| **Panel A: Continuous** | | | |
| Grade ($\beta_1$) | 0.1406*** | 0.1440*** | 0.1227*** |
| | (0.0044) | (0.0033) | (0.0045) |
| Test ($\beta_2$) | 0.1232*** | 0.1252*** | 0.1067*** |
| | (0.0044) | (0.0034) | (0.0043) |
| $p$-value $\beta_1 = \beta_2$: | 0.0149 | 0.0003 | 0.0124 |
| **Panel B: Binary** | | | |
| Bad Grade ($\beta_1$) | 0.3667*** | 0.3674*** | 0.3240*** |
| | (0.0128) | (0.0099) | (0.0129) |
| Bad Test ($\beta_2$) | 0.3130*** | 0.3120*** | 0.2660*** |
| | (0.0128) | (0.0099) | (0.0120) |
| $p$-value $\beta_1 = \beta_2$: | 0.0046 | 0.0002 | 0.0008 |
| Respondent FE | No | Yes | No |
| Clustered SE | No | No | Yes |
| R-squared | 0.0728 | 0.5459 | 0.3765 |
| Observations | 23,221 | 23,221 | 22,865 |

*Notes:* This table presents results after restricting our analysis sample to time investments that were less than 7 hours/day and monetary investments that were less than 81 dollars per week. Each panel shows estimated effects of grade and test information on parental investment decisions. Standard errors are shown in parentheses beneath the coefficients. Panel A reports linear models using continuous measures. Panel B restricts the sample to cases where either grades or tests indicate bad performance. The final row in each panel shows the $p$-value from a test of equality between the grade and test coefficients..

Table 4: Regression Coefficients for Grades and Tests

| Variable | Participant FE | Controls + Clustering |
|---|---|---|
| **Grade B** | 0.1005*** | 0.0728*** |
| | (0.0150) | (0.0143) |
| **Grade C** | 0.2783*** | 0.2201*** |
| | (0.0151) | (0.0153) |
| **Grade D** | 0.4316*** | 0.3884*** |
| | (0.0150) | (0.0174) |
| **Grade F** | 0.5569*** | 0.4558*** |
| | (0.0150) | (0.0193) |
| **Test 70th** | 0.1275*** | 0.1017*** |
| | (0.0150) | (0.0153) |
| **Test 50th** | 0.2656*** | 0.2232*** |
| | (0.0150) | (0.0166) |
| **Test 30th** | 0.3875*** | 0.3197*** |
| | (0.0149) | (0.0177) |
| **Test 10th** | 0.4951*** | 0.4246*** |
| | (0.0150) | (0.0187) |
| **Constant** | -0.5276*** | -0.5450*** |
| | (0.0141) | (0.0787) |
| **N** | 23,256 | 22,900 |
| **R-squared** | 0.548 | 0.382 |

*Notes:* Each column reports coefficients from a separate OLS regression. The dependent variable is the amount of investment advice provided. Grade and test score categories are mutually exclusive dummies, with the omitted categories being Grade A and Test 90th percentile. Column 1 absorbs participant fixed effects (`pid`). Column 2 controls for parental demographics, clustering standard errors at the participant level. $^{***}p < 0.01$.

Table 5: Parental Investment Responses to Grade and Test Information

| | (1) OLS | (2) Clustered | (3) Individual FE |
|---|---|---|---|
| Grades good, tests bad ($\beta_1$) | $-0.028$ | $-0.028$ | $-0.041^{***}$ |
| | (0.019) | (0.019) | (0.015) |
| Grades bad, tests good ($\beta_2$) | $0.040^{**}$ | $0.040^{**}$ | 0.023 |
| | (0.019) | (0.019) | (0.015) |
| Grades good and tests good ($\beta_3$) | $-0.390^{***}$ | $-0.390^{***}$ | $-0.420^{***}$ |
| | (0.019) | (0.018) | (0.015) |
| Grades bad and tests bad ($\beta_4$) | $0.412^{***}$ | $0.412^{***}$ | $0.392^{***}$ |
| | (0.019) | (0.020) | (0.015) |
| Constant | $-0.005$ | $-0.005$ | 0.008 |
| | (0.011) | (0.017) | (0.008) |
| $\beta_1 = \beta_2$ (p-value) | 0.0026 | 0.0075 | 0.0002 |
| Observations | 23,199 | 23,199 | 23,199 |
| R-squared | 0.054 | 0.054 | 0.528 |

Notes: The dependent variable is parental investment. Column (1) reports OLS estimates. Column (2) reports OLS estimates with standard errors clustered at the individual level. Column (3) includes individual fixed effects using an absorbed-effects specification. The reported p-values correspond to F-tests of the null hypothesis $\beta_1 = \beta_2$. The omitted group is all combinations of grades and tests that include either a C, or 50 percent performanced combined with any other of each assessment (e.g. C, 10th percent or A and 50th percent). Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 6: Heterogeneous Effects of Grade and Test Signals by Parental Characteristics

**Panel A: Continuous Specifications**

| | **Time** | **Low Inc.** | **White** | **Hispanic** | **Black** | **Asian** | **Parent BA** | **Parent Female** | **Child Female** |
|---|---|---|---|---|---|---|---|---|---|
| Grade Effect ($\beta_1$) | 0.171*** | 0.129*** | 0.116*** | 0.122*** | 0.125*** | 0.123*** | 0.115*** | 0.110*** | 0.124*** |
| | (0.007) | (0.006) | (0.009) | (0.005) | (0.005) | (0.005) | (0.007) | (0.009) | (0.006) |
| Test Effect ($\beta_2$) | 0.151*** | 0.116*** | 0.090*** | 0.108*** | 0.109*** | 0.105*** | 0.095*** | 0.098*** | 0.108*** |
| | (0.007) | (0.006) | (0.008) | (0.004) | (0.005) | (0.004) | (0.006) | (0.008) | (0.006) |
| Demographic Effect ($\beta_3$) | 0.335*** | 0.236*** | -0.197*** | 0.247*** | 0.295*** | 0.243 | -0.092** | -0.148*** | 0.012 |
| | (0.034) | (0.059) | (0.061) | (0.089) | (0.059) | (0.153) | (0.043) | (0.047) | (0.029) |
| Grade × Demo ($\delta_1$) | -0.107*** | -0.016 | 0.009 | -0.007 | -0.018 | -0.033 | 0.014 | 0.017* | -0.005 |
| | (0.008) | (0.009) | (0.010) | (0.018) | (0.013) | (0.030) | (0.009) | (0.010) | (0.007) |
| Test × Demo ($\delta_2$) | -0.099*** | -0.024*** | 0.022** | -0.041** | -0.022* | 0.004 | 0.020** | 0.011 | -0.005 |
| | (0.008) | (0.009) | (0.010) | (0.017) | (0.013) | (0.027) | (0.009) | (0.010) | (0.007) |
| p-value: $\delta_1 = \delta_2$ | 0.471 | 0.501 | 0.352 | 0.163 | 0.821 | 0.244 | 0.609 | 0.638 | 0.988 |

**Panel B: Binary Specifications**

| | **Bad Time** | **Low Inc.** | **White** | **Hispanic** | **Black** | **Asian** | **Parent BA** | **Parent Female** | **Child Female** |
|---|---|---|---|---|---|---|---|---|---|
| Grade Effect ($\beta_1$) | 0.451*** | 0.341*** | 0.317*** | 0.322*** | 0.329*** | 0.325*** | 0.315*** | 0.286*** | 0.331*** |
| | (0.020) | (0.017) | (0.025) | (0.013) | (0.014) | (0.013) | (0.020) | (0.024) | (0.016) |
| Test Effect ($\beta_2$) | 0.378*** | 0.290*** | 0.226*** | 0.269*** | 0.269*** | 0.260*** | 0.225*** | 0.235*** | 0.267*** |
| | (0.019) | (0.016) | (0.024) | (0.012) | (0.013) | (0.012) | (0.018) | (0.022) | (0.016) |
| Demographic Effect ($\beta_3$) | -0.068*** | 0.159*** | -0.128*** | 0.151*** | 0.208*** | 0.176* | -0.023 | -0.099*** | -0.004 |
| | (0.019) | (0.048) | (0.047) | (0.052) | (0.035) | (0.091) | (0.026) | (0.028) | (0.012) |
| Grade × Demo ($\delta_1$) | -0.283*** | -0.044 | 0.008 | 0.005 | -0.040 | -0.120 | 0.013 | 0.051* | -0.017 |
| | (0.022) | (0.027) | (0.029) | (0.055) | (0.036) | (0.086) | (0.026) | (0.029) | (0.020) |
| Test × Demo ($\delta_2$) | -0.254*** | -0.069*** | 0.050* | -0.122*** | -0.049 | 0.034 | 0.067*** | 0.037 | -0.012 |
| | (0.021) | (0.025) | (0.027) | (0.046) | (0.036) | (0.090) | (0.024) | (0.026) | (0.021) |
| p-value: $\delta_1 = \delta_2$ | 0.341 | 0.492 | 0.275 | 0.073 | 0.866 | 0.132 | 0.127 | 0.710 | 0.892 |

Notes: Each column reports estimates from a separate regression of parental investment on grade and test signals, a demographic characteristic, and interactions. Panel A uses continuous measures; Panel B uses binary indicators. Robust standard errors clustered at the parent (pid) level are reported in parentheses. The final row reports the p-value from a Wald test of equality between the grade–demographic and test–demographic interaction terms. *, **, *** denote significance at the 10%, 5%, and 1% levels.

Table 7: Descriptive Statistics for Parent Belief Variables

| **Variable** | **N** | **Mean** | **SD** | **Min** | **Max** |
|---|---|---|---|---|---|
| Tests biased against certain groups | 2079 | 0.402 | 0.490 | 0 | 1 |
| Feel bad (Grades) | 2079 | 0.513 | 0.500 | 0 | 1 |
| Feel bad (Tests) | 2079 | 0.407 | 0.491 | 0 | 1 |
| Returns higher w/strong student | 2079 | 0.362 | 0.481 | 0 | 1 |
| Tests reflect parent income | 2079 | 0.280 | 0.449 | 0 | 1 |
| Returns higher w/weak student | 2079 | 0.730 | 0.444 | 0 | 1 |
| Grades capture behavior | 2079 | 0.766 | 0.424 | 0 | 1 |
| Tests capture skills | 2079 | 0.189 | 0.391 | 0 | 1 |
| Grades capture emotion | 2079 | 0.463 | 0.499 | 0 | 1 |
| Tests reflect parent skills | 2079 | 0.157 | 0.364 | 0 | 1 |
| Returns higher w/struggling student | 2079 | 0.793 | 0.405 | 0 | 1 |
| Returns higher w/more guidance | 2079 | 0.206 | 0.404 | 0 | 1 |
| Grades > Tests | 2079 | 0.708 | 0.455 | 0 | 1 |
| Grades < Tests | 2079 | 0.085 | 0.279 | 0 | 1 |

*Notes*: Table reports summary statistics for binary indicators of parent beliefs. SD is standard deviation. All variables have 2079 observations.

Table 8: Interaction Effects by Subgroup: Parent-Level Variables and Information Signals

## Panel A: Continuous Specifications

| | Grades More | Test Biased | Feel Bad (Grades) | Feel Bad (Tests) | Pay Off | Tests Parents | Spending | More Likely | Behavior (Grades) | Tests Skills | Grades Emotion | Struggling Benefit | More Guidance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grade Effect ($\beta_1$) | 0.100*** | 0.122*** | 0.107*** | 0.113*** | 0.124*** | 0.126*** | 0.126*** | 0.082*** | 0.115*** | 0.128*** | 0.112*** | 0.076*** | 0.123*** |
| | (0.009) | (0.006) | (0.006) | (0.006) | (0.006) | (0.005) | (0.005) | (0.009) | (0.010) | (0.005) | (0.006) | (0.012) | (0.005) |
| Test Effect ($\beta_2$) | 0.100*** | 0.106*** | 0.098*** | 0.100*** | 0.105*** | 0.108*** | 0.109*** | 0.062*** | 0.096*** | 0.109*** | 0.096*** | 0.078*** | 0.110*** |
| | (0.008) | (0.006) | (0.006) | (0.005) | (0.005) | (0.005) | (0.005) | (0.008) | (0.008) | (0.005) | (0.006) | (0.010) | (0.005) |
| Demographic Effect ($\beta_2$) | -0.186*** | -0.027 | -0.051 | 0.019 | 0.109** | 0.216*** | 0.262*** | -0.357*** | -0.095* | 0.227*** | -0.140*** | -0.284*** | 0.234*** |
| | (0.046) | (0.042) | (0.041) | (0.042) | (0.043) | (0.060) | (0.049) | (0.049) | (0.049) | (0.058) | (0.041) | (0.056) | (0.054) |
| Grade $\times$ Demo ($\delta_1$) | 0.031*** | 0.000 | 0.030*** | 0.022** | -0.005 | -0.025* | -0.011 | 0.055*** | 0.009 | -0.034*** | 0.020** | 0.058*** | -0.006 |
| | (0.010) | (0.009) | (0.009) | (0.010) | (0.010) | (0.013) | (0.011) | (0.010) | (0.011) | (0.012) | (0.009) | (0.013) | (0.012) |
| Test $\times$ Demo ($\delta_2$) | 0.008 | -0.002 | 0.015* | 0.013 | 0.002 | -0.019 | -0.012 | 0.060*** | 0.012 | -0.018 | 0.020** | 0.035*** | -0.023** |
| | (0.010) | (0.009) | (0.009) | (0.009) | (0.009) | (0.012) | (0.011) | (0.010) | (0.010) | (0.011) | (0.009) | (0.011) | (0.011) |
| p-value: $\delta_1 = \delta_2$ | 0.087 | 0.873 | 0.250 | 0.483 | 0.578 | 0.746 | 0.937 | 0.712 | 0.812 | 0.330 | 0.950 | 0.155 | 0.279 |

## Panel B: Binary Specifications

| | Grades More | Test Biased | Feel Bad (Grades) | Feel Bad (Tests) | Pay Off | Tests Parents | Spending | More Likely | Behavior (Grades) | Tests Skills | Grades Emotion | Struggling Benefit | More Guidance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grade Effect ($\beta_1$) | 0.269*** | 0.318*** | 0.280*** | 0.301*** | 0.327*** | 0.333*** | 0.334*** | 0.218*** | 0.314*** | 0.337*** | 0.294*** | 0.210*** | 0.328*** |
| | (0.024) | (0.017) | (0.018) | (0.016) | (0.016) | (0.013) | (0.015) | (0.024) | (0.028) | (0.014) | (0.018) | (0.032) | (0.014) |
| Test Effect ($\beta_2$) | 0.241*** | 0.268*** | 0.244*** | 0.256*** | 0.257*** | 0.267*** | 0.268*** | 0.160*** | 0.252*** | 0.267*** | 0.240*** | 0.196*** | 0.272*** |
| | (0.023) | (0.016) | (0.017) | (0.015) | (0.014) | (0.013) | (0.013) | (0.023) | (0.024) | (0.013) | (0.017) | (0.027) | (0.013) |
| Demographic Effect ($\beta_3$) | -0.108*** | -0.028 | 0.039 | 0.096*** | 0.098*** | 0.124*** | 0.215*** | -0.127*** | -0.042 | 0.117*** | -0.065*** | -0.096*** | 0.181*** |
| | (0.028) | (0.025) | (0.024) | (0.025) | (0.025) | (0.035) | (0.028) | (0.029) | (0.029) | (0.032) | (0.024) | (0.033) | (0.031) |
| Grade $\times$ Demo ($\delta_1$) | 0.076*** | 0.011 | 0.084*** | 0.056** | -0.011 | -0.065* | -0.037 | 0.143*** | 0.011 | -0.080** | 0.063** | 0.141*** | -0.025 |
| | (0.029) | (0.027) | (0.026) | (0.027) | (0.028) | (0.037) | (0.031) | (0.029) | (0.032) | (0.034) | (0.026) | (0.035) | (0.034) |
| Test $\times$ Demo ($\delta_2$) | 0.028 | -0.018 | 0.033 | 0.012 | 0.014 | -0.039 | -0.021 | 0.139*** | 0.012 | -0.036 | 0.046* | 0.080*** | -0.053* |
| | (0.027) | (0.024) | (0.024) | (0.025) | (0.026) | (0.034) | (0.029) | (0.027) | (0.028) | (0.033) | (0.024) | (0.030) | (0.030) |
| p-value: $\delta_1 = \delta_2$ | 0.223 | 0.399 | 0.147 | 0.232 | 0.518 | 0.577 | 0.701 | 0.923 | 0.987 | 0.339 | 0.635 | 0.188 | 0.534 |

*Notes*: Each column shows interaction coefficients for a specific parent or child subgroup. Coefficients indicate how information signals (Grade or Test) affect parental investment behaviors, conditional on subgroup membership. Standard errors, clustered at the student level, are in parentheses. "Characteristic" rows show baseline differences in subgroup investment. "Grade Interaction" and "Test Interaction" rows capture differential responses to each signal. $p$-values report tests for equality of the interaction coefficients. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

# A Tables

Table A1: Relationship Between Academic Measures and Investment

| | No Controls | Clustered + Controls |
|---|---|---|
| **Panel A: Continuous Measures** | | |
| Grade | 0.139*** | 0.122*** |
| | (0.004) | (0.005) |
| Test | 0.123*** | 0.106*** |
| | (0.004) | (0.004) |
| p-value: Grade = Test | 0.0112 | 0.0080 |
| **Panel B: Binary Measures** | | |
| Grade (Low) | 0.365*** | 0.324*** |
| | (0.013) | (0.013) |
| Test (Low) | 0.315*** | 0.264*** |
| | (0.013) | (0.012) |
| p-value: Grade = Test | 0.0054 | 0.0007 |

*Notes:* This table reports OLS estimates of the relationship between academic performance and weekly monetary investments. Panel A uses continuous measures; Panel B uses binary indicators for low performance. The "Clustered + Controls" specification includes demographic controls and clusters standard errors at the respondent level. All regressions include a control for whether the respondent correctly answered the percentage score attention check. Standard errors are shown in parentheses. *Significance levels:* *** $p$ <0.01, ** $p$ <0.05, * $p$ <0.1.