UNIVERSITÀ DEGLI STUDI FIRENZE
Dipartimento di Statistica, Informatica, Applicazioni "Giuseppe Parenti"
Eccellenza 2023-2027

Good Grades for Hard Work? A Lab-in-the-Field Study of Effort and Educational Inequality

Carlos J. Gil-Hernández, Alberto Palacios-Abad, Jonas Radl

DISIA WORKING PAPER
2025/09

Good Grades for Hard Work?

A Lab-in-the-Field Study of Effort and Educational Inequality

Carlos J. Gil-Hernández [1]

Alberto Palacios-Abad [2]

Jonas Radl [2,3] *

[1] University of Florence, Department of Statistics, Computer Science, Applications

[2] Universidad Carlos III de Madrid

[3] WZB Berlin Social Science Center

* Corresponding author: jonas.radl@uc3m.es

Abstract: Despite its importance for status attainment and meritocracy, measuring effort remains elusive, often relying on indirect proxies or unreliable self-reports. This study examines how objective (cognitive effort, CogEff) and subjective (teacher-perceived effort, TpEff) measures of student effort contribute to educational inequality. We examine the predictive capacity of effort for educational performance and test the mediating and moderating roles of effort in the relationship between parental socioeconomic status (SES) and school grades. Drawing on original, representative "lab-in-the-field" data from 1,270 fifth-graders in Spain and Germany, who performed three different incentivized real-effort tasks engaging various executive functions, four key findings emerge. First, both CogEff and TpEff predict grade point average (GPA), with TpEff having a powerful effect, more predictive even than IQ or parental SES. Second, effort—especially TpEff—is unequally distributed by parental SES and explains a substantial share of the SES-based GPA gap, on par with IQ. Third, roughly half of the GPA gap by social origin remains unexplained even after accounting for academic merit (IQ + effort). Fourth, while grading returns to CogEff are independent of SES, high-SES students are significantly less penalized for low TpEff than low-SES peers. Overall, effort predicts academic success and shapes educational (in)equality. High-SES students show higher average effort and can afford to be perceived as lazy, while hardworking low-SES students can overcome disadvantage through greater returns to teacher-perceived effort. We discuss the findings' implications for student agency, educational inequality, and fair evaluations.

Keywords: effort, socio-behavioral skills, inequality, grades, socioeconomic status, laboratory study

Funding: This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 758600).

1. Introduction

A prominent interdisciplinary literature addresses how socio-behavioral traits shape future life chances and (un)equal opportunity (Bowles and Gintis 2002; Farkas 2003; Heckman and Kautz 2012).[1] Effort, along with related traits such as conscientiousness, self-control, and grit (Duckworth et al. 2007), has a demonstrated influence on educational and labor market success (Smithers et al. 2018). Given unequal developmental opportunities and parenting practices (Kaiser 2016; Mcewen and Mcewen 2017), effort may also contribute to the intergenerational transmission of inequality (McGue et al. 2020).

Effort is a key concept in normatively charged debates about meritocracy and equal opportunity (Roemer and Trannoy 2016). Among the many factors influencing educational attainment, effort is often strongly emphasized by teachers and parents to improve learning (Schöner et al. 2024), as being hardworking is typically seen within students' agency. Accordingly, different theories of justice argue that inequalities are legitimate if they reflect differences in merit—effort, and IQ—rather than ascribed circumstances like race or social background (Asadullah et al. 2021.).

Despite its central conceptual role, effort remains an elusive factor in social stratification research due to measurement challenges (Radl and Miller 2021). Most existing studies rely on different self-reported or observational measures, which are subject to limitations in validity and reliability (Apascaritei, Demel, and Radl 2021). Reference group effects and social desirability often bias student self-reports (Chen et al. 2020), whereas teacher assessments may reflect biases related to student ascribed status (Owens 2022). This measurement deficit has left a blind spot in understanding how effort contributes to educational achievement and inequality. The present study contributes to the understanding of educational inequality by introducing a novel measure of cognitive effort, collected in a controlled laboratory setting and with strong claims to validity. It exploits incentivized real-effort tasks drawn from cognitive psychology and behavioral economics (Gill and Prowse 2012; Diamond 2013). Our original data stem from a laboratory experiment and surveys conducted with a representative school-based sample of fifth-grade students from the capital cities of Spain and Germany.

This article investigates four research questions on the predicting, mediating, and moderating roles of effort on educational achievement and inequality: (1) Does student effort predict academic performance? (2) Does student effort contribute to the intergenerational reproduction of educational inequality as a mediator of social origin effects? (3) How much inequality in academic performance remains after fully accounting for student academic merit (ability + effort)? (4) Does student effort moderate the association between SES background and academic performance?

First, we examine the association between students' objective and subjective effort measures and school grades. Using an innovative and objective measure of cognitive effort, we test its predictive power and ecological validity on grade point average, benchmarking against the effects of SES background and cognitive ability. The most direct way effort should affect grades is through engagement-based learning gains and good classroom behavior (Randall and Engelhard 2010). However, an indirect effect could arise if cognitive effort (CogEff) also influences teachers' perceptions of academic aptitude (Miller et al. 2017), as teachers are the ultimate evaluators of educational achievement. Therefore, we also study the impact of subjective, teacher-perceived effort on grade point average (GPA). To assess how effort is perceived and rewarded, we compare behavioral effort with teacher assessments. This triangulation allows us to examine how far subjective teacher perceptions align with objectively measured student effort. In doing so, we identify potential biases and additional factors teachers consider in grading (Bowers 2011).

Second, we investigate to what extent effort might mediate the effect of parental SES on educational achievement, contributing to the intergenerational transmission of educational inequality (McGue 2020; Attanasio, De Paula, and Toppeta 2025). The role of socio-behavioral traits in educational inequality is contested (Kröger, Palacios-Abad, and Radl 2024). While some studies suggest that they are less influenced by family background than cognitive skills (Holtmann, Menze, and Solga 2021), evidence also indicates that traits such as executive function, attention control, or delay of gratification are unequally distributed across socioeconomic status groups (Hackman, Farah, and Meaney 2010). Contrasting perspectives that stress the heritability of cognitive ability (IQ), Bowles and Gintis (1976, 2002) championed the view that personality factors play a strong role in the intergenerational persistence of economic status. The present study examines how CogEff, measured objectively through lab tasks, and teacher-perceived effort (TpEff), based on subjective teacher judgments, mediate the link between parental socioeconomic status (SES) and academic achievement. TpEff may reflect both actual behavior and teacher bias, making it exceptionally responsive to social background.

Third, our study investigates whether a residual association remains between parental SES and student grades after controlling for IQ as well as both objective and subjective effort measures (CogEff and TpEff). Beyond differences in measurable cognitive skills and effort, students from higher SES backgrounds often receive more favorable grades and evaluations than their lower-SES peers (Südkamp et al. 2012; Zanga and De Gioannis 2023). From an equality of opportunity standpoint, this raises the question of whether residual advantages in grading reflect unobserved student abilities, parental support for learning, or systematic teacher bias (van Huizen, Jacobs, and Oosterveen 2024). Prior research suggests that high-SES students are more likely to be perceived as competent (Tobisch and Dresel 2017), generate higher expectations (Wang et al. 2018), and benefit from subtle classroom advantages linked to cultural capital and teacher perceptions (Jæger 2022).

Fourth, we test two sociological mechanisms that explain how parental SES might moderate the effect of effort on grades: the compensatory advantage (Bernardi 2014) and resource substitution hypotheses (Shanahan et al. 2014). These theories suggest that high-SES families can offset adverse events or traits for status attainment or substitute a lack of ability with resources and investments (private tutoring; homework support) (Burger and Brack 2025). Consequently, the negative impact of low effort on GPA should be less severe for high-SES students, but previous research is inconclusive (Liu 2019). Teachers may also perceive cultural capital, derived behaviors and attitudes as relevant for grading, resulting in less penalization of low effort among high-SES students.

To answer these questions, we employ a novel research design based on "lab-in-the-field" experiments conducted in Madrid, Spain, and Berlin, Germany. Our sample includes 1,270 fifth-grade students from a representative sample of primary schools (public, private, and charter). In the lab, students completed three real-effort tasks, an IQ test, and a survey. The tasks, derived from cognitive psychology and behavioral economics, target different executive functions (Diamond 2013) to measure CogEff net of ability. TpEff is collected through interviews with students' teachers, while parental and student surveys provide socioeconomic information and school grades in math and language. This unique design exploiting a novel multilayered dataset opens a new window into the processes generating educational inequality in the classroom.

2. Theoretical Framework and Hypotheses

2.1. Effort and Educational Achievement

The early seminal work of Bowles and Gintis (1976) recognized the importance of personality characteristics for educational achievement. Since the early 2000s, the study of effort and related socio-behavioral traits (Smithers et al. 2018) has gained traction through models of skill and child development, focusing on the early determinants of later-life outcomes (Heckman and Katz 2012). Building upon psychological scholarship (Conti and Heckman 2014), social scientists have shown that a wide variety of personality scales, with varying degrees of construct validity (Duckworth and Kern 2011; McCrae and Mõttus 2019), are significant predictors of educational attainment and life outcomes (Smithers et al. 2018). Most previous studies have utilized self-reported personality scales related to the concept of effort, such as locus of control, self-control, grit, or self-discipline (Duckworth et al. 2007; Hsin and Xie 2017).

Personality traits, often assessed through the Big Five taxonomy, are among the most robust psychometric constructs (McCrae and Mõttus 2019). Conscientiousness is the Big Five trait most related to academic achievement and effort (Mammadov 2022), comprising sub-traits like self-regulation, attention, goal pursuit, planning and orderliness. Behaviors such as persistence, delaying gratification, or following rules are essential for school success, as they help children learn by staying focused, coping with frustration, and avoiding disruptive behavior. Self-regulation is a facet of conscientiousness and a key dimension of early child development, reflecting the capacity to mobilize costly cognitive resources, direct attention, and control impulses (Fleming, Heintzelman, and Bartholow 2016).

However, self-reports measuring personality are prone to measurement error because they do not readily generalize to actual behavior across real-life situations and do not account for performance-based incentives (Conti and Heckman 2014). Studies have shown that the link between self-reported traits and real effort is weak or even absent (Duckworth and Kern 2011; Apascaritei et al. 2021). An alternative is to use indirect, behavior-based measures (Palacios-Abad 2021). For instance, Borghans and Schils (2012) create a measure of test effort based on performance persistence over the 2-hour PISA test. This behavioral measure correlates with personality traits, such as conscientiousness, and predicts outcomes, including life satisfaction and alcohol use. Similarly, Zamarro et al. (2019) found that test persistence explains 32–38% of the variation in PISA scores across countries. Some such measures derived from test-taking behavior significantly predict later educational attainment, but still have limited internal validity (Borgonovi, Ferrara, and Piacentini 2023).

We expand on the previous literature by using an innovative direct measure of CogEff, which reflects actual behavior in the laboratory among a cross-national, representative sample of primary school students. We employ a measure based on three different real-effort tasks that integrates methodologies from both cognitive science and behavioral economics, aiming to capture the engagement of executive functions under incentivized conditions. Our first hypothesis is that there is a positive, sizable association between CogEff and educational performance, as measured by teacher-assigned school grades in math and language: **H1a.** *Cognitive effort positively predicts grade point average.*

While objective behavioral measures of individual effort are arguably more internally valid than survey-based self-evaluations by students, subjective measures are highly relevant when it comes to teachers' perceptions (Bowers 2011). As teachers are the evaluators of academic merit in the education system, their perceptions of student classroom behavior are essential. A meta-analysis by Südkamp et al. (2012) found that the correlation between teachers' judgments of student ability and standardized test scores is, on average, 0.63, leaving substantial room for student personality traits and teacher perceptions to influence the grading criteria (Borghans et al. 2016).

In many education systems, separate comportment grades formally assess traits such as orderliness, work habits and rule-following. Some teachers even openly admit factoring in behavioral aspects into subject grades (Schöner et al. 2024), indicating that effort effectively serves as an independent evaluation criterion. The appropriateness of such “pedagogical grades” is debatable, while its assessment is complicated by the difficulty of accurately observing grading schemes. In turn, teachers’ grading practices also affect students’ achievement and effort (Krohn and O’Connor 2005). Hence, students’ behavioral traits that influence achievement and grading, but which might not be fairly assessed by teachers, merit particular attention.

TpEff has been examined across various disciplines, including sociology (Carbonaro 2005), economics (Asadullah et al. 2021), and psychology (Upadyaya and Eccles 2015), with significant impacts on academic outcomes. Experimental and observational studies have found that while teachers primarily base grades on student achievement or curriculum knowledge, behavioral factors such as effort and behavior also influence final grades (Ferman and Fontes 2022; Randall and Engelhard 2010). Grading is multidimensional (Bowers 2011), with teachers often rewarding effort and conduct separately from achievement, partly due to their belief in these traits’ potential to boost future performance (Kelly 2008).

Thus, teacher reports may offer a direct classroom assessment of those student traits considered for grading; however, such data are seldom systematically collected. Hence, combining objective measures from behavioral (lab) settings with subjective classroom observations may improve both internal and ecological validity. In this study, we test how TpEff relates to students’ grades compared with students’ actual CogEff, exploring the accuracy of teacher judgments on the grading process. We posit the following hypothesis: **H1b.** *Teacher-perceived effort positively predicts grade point average, everything else being equal, and exerts a stronger influence than cognitive effort.*

2.2. Effort and Educational Inequality

Following Boudon's classical model (1974), two effects of social background mainly explain why educational inequality tends to persist across generations: primary effects, i.e., the mediating role of academic performance (GPA) between parental SES and educational transitions, and secondary effects, i.e., residual SES inequalities in educational transitions (e.g., decision-making), net of performance (Jackson 2013). Here, we focus on inequalities in academic performance, taking a step back to analyze: (1) the mediating and moderation roles of both objective and subjective measures of effort, net of ability (IQ), between parental SES and GPA; and (2) residual SES inequalities in grading, net of student ability.

2.2.1. The Mediating Role of Effort

Research on the socioeconomic stratification of personality traits has produced mixed findings (Kröger, Palacios-Abad, and Radl 2024). Numerous studies suggest that personality traits are less directly transmitted across generations than cognitive skills and are therefore less strongly shaped by family socioeconomic background (Holtmann, Menze, and Solga 2021; Gruijters, Raabe, and Hübner 2023). However, several studies also show that key socio-behavioral traits—such as delay of gratification (Watts et al. 2018), socioemotional skills (Attanasio, De Paula, Toppeta and 2025), attention control (Duncan and Magnuson 2011), and executive function (Hackman, Farah, and Meaney 2010)—are unequally developed across SES groups from early childhood due to disparities in resources, investments, and parenting styles.

To better understand these dynamics, we examine how both objective and subjective measures of effort may mediate the intergenerational transmission of educational inequality. We focus on two complementary mechanisms: (1) socioeconomic disparities in the development of CogEff; and (2) socioeconomic disparities in TpEff. Compared to lab-based effort measures, teacher perceptions may more accurately reflect students' actual classroom behavior and a broader set of socio-behavioral traits that teachers implicitly evaluate—traits which themselves may be shaped by students' socioeconomic background. At the same time, such "eye of the beholder" evaluations can also introduce bias, as they may reflect idiosyncratic judgments based on students' ascribed characteristics (Owens 2022). For instance, teachers may underestimate effort or overemphasize disruptive behavior among low-SES students, regardless of actual task engagement, ability or performance. Such biases can directly affect grades and influence long-term educational trajectories. Accordingly, we propose the following hypothesis: **H2a.** *Both cognitive effort and teacher-perceived effort are significantly influenced by parental socioeconomic status, with teacher-perceived effort showing a stronger association.*

Indeed, for personality to meaningfully mediate the intergenerational transmission of educational inequality, they must be associated with both parental SES and academic outcomes. However, the magnitude of this mediation depends on the personality trait, the validity of the measure, and the specific educational outcome under consideration—factors that have yielded mixed results in prior research. Some studies find that commonly used personality traits only weakly mediate educational inheritance or not at all (Holtmann, Menze, and Solga 2021). Similarly, a recent cross-national study using PISA data by Gruijters, Raabe, and Hübner (2023) shows that socio-emotional skills (e.g., growth mindset, self-efficacy, work mastery) are unequally distributed by SES, yet explain only a modest share (about 9%) of the socioeconomic achievement gap. In contrast, other studies report a stronger mediating role of personality in educational stratification (Shanahan et al. 2014; Damian et al. 2015; Hsin and Xie 2017).

Building on the rationale developed in hypotheses H1b and H2a—namely, that TpEff is both more strongly shaped by parental SES and more strongly predictive of grades than CogEff—we advance the following hypothesis: **H2b.** *Both cognitive effort and teacher-perceived effort significantly mediate the relationship between parental socioeconomic status and grade point average, with teacher-perceived effort serving as a stronger mediator than cognitive effort.*

2.2.2. The Residual Role of Parental SES: Teacher Bias or Unobserved Ability?

The intergenerational reproduction of education is mainly explained by unequal opportunity for skill development from an early age (Skopek and Passaretta 2021). Still, children from high-SES families tend to garner higher ability judgments (Südkamp et al. 2012), grades (Zanga and De Gioannis 2023), expectations (Geven et al. 2021), and track recommendations (Batruch et al. 2023) from their teachers. Ultimately, they thus advance to higher levels of education than their low-SES peers (Gil-Hernández 2021), even when academic performance is similar, be it measured by skills, test scores, or GPA. Students' objective ability is usually proxied by external standardized assessments (ideally covering the same school curriculum) or cognitive ability tests (Zanga and De Gioannis 2023).

Several factors might simultaneously account for residual SES gaps in grading. First, over and above students' observed ability and effort, high-SES students might benefit from parental resources and support to improve learning and knowledge of the official evaluable curriculum (Kalil, Ryan, and Corey 2012). Second, measurement error in cognitive tests and behavioral self-reported measures might attenuate the effect of skills and inflate residual SES effects (van Huizen, Jacobs, and Oosterveen 2024), as they tend to be positively correlated. Third, some teachers might display systematic grading biases against low-SES students due to various psychological mechanisms: implicit stereotypes or attitudes (Fazio et al. 2023), competence expectations by status group (Melamed et al. 2019), or statistical discrimination (Wenz and Hoenig 2020). Finally, cultural capital (Jæger 2022) might directly boost curriculum learning, teacher attention and privileges (Calarco 2014), or be misconceived as academic brilliance (Bourdieu and Passeron 1990). In line with all these factors, we stipulate the following hypothesis: **H3.** *Parental socioeconomic status has a residual positive association with students' grade point average, net of objective (CogEff and IQ) and subjective (TpEff) measures of student ability.*

2.2.3. The Moderating Role of Effort: Compensation or Reinforcement?

Socioeconomic resources and skills might be independent or interact, being substitutes or complements, in predicting status attainment (Damian et al. 2015). The *Matthew effect*, also termed *reinforcement* or *boosting*, is rooted in cumulative advantage theory to suggest that skills generate the highest returns for students from privileged backgrounds (Kwon and Erola 2022), triggering a multiplicative process so that "the rich get richer" (DiPrete and Eirich 2006).

Conversely, the *resource substitution hypothesis* posits that either skills or socioeconomic resources can drive status attainment, depending on which is more readily available (Burger and Brack 2025). Low-SES students with favorable cognitive or personality traits may offset their economic disadvantage, getting high returns for their skills. In contrast, high-SES students can rely more on family resources if they lack strong skills. Accordingly, several studies have found that socio-behavioral traits are more predictive of educational and status attainment among low-SES individuals (Schoon, Mele, and Burger 2025: academic self-concept and school engagement; Mele, Buchman, and Burger 2023: self-reported study effort; Liu

2019: index on approaches to learning, self-control and interpersonal skills; Damian et al. 2015: Big Five personality traits; Shanahan et al. 2014: Big Five), but findings are mixed.

The *compensatory advantage hypothesis* offers similar predictions to the resource substitution hypothesis, with more elaboration on the micro-level mechanisms. Drawing from rational action theories, it argues that educational inequalities are most pronounced among lower-performing students (Bernardi 2014). High-SES families are more averse to social demotion than low-SES families, displaying markedly inelastic expectations (Bernardi and Valdés 2021) and a strong motivation to use their resources for academic support of their children to reproduce their status. If their children perform poorly, high-SES parents might compensate by providing extra learning support (e.g., help with homework, private tutoring) (Park et al. 2011; Kaiser 2016), and school involvement (Forster and van de Werfhorst 2020). In contrast, low-SES families are more responsive to signals of academic aptitude and may lower their academic support and expectations in reaction to subpar grades (Holm et al. 2019). Teachers, in turn, may internalize or reinforce these expectations, sometimes misjudging student ability and (un)intentionally amplifying disparities, especially among low-skilled students.

While the compensatory hypothesis focuses on educational transitions and parental compensatory strategies under negative events (e.g., a poor report card) for status attainment (Bernardi and Valdés 2021), we focus on low effort and grading outcomes. Specifically, we examine the interplay between student effort and SES background, net of cognitive ability. We argue that teachers do not grade all children exerting low effort equally because they might further weigh in school engagement, disciplined behavior or appropriate interaction with the teacher, where high-SES students have an inherent advantage. Therefore, we expect the following hypothesis to hold: **H4.** *The GPA disadvantage associated with low cognitive effort or teacher-perceived effort is attenuated for high-SES students.*

3. Data, Variables and Methods

3.1. Data and Experimental Setup

Data stems from a lab experiment carried out with 1,270 5th-grade students from primary schools in the metropolitan area of Madrid (n = 789), Spain, and Berlin (n = 481), Germany, during the school year 2019/2020 and 2021/2022. To collect a representative sample of the fifth-grade population in each city, the schools were randomly selected from a sample stratified by neighborhood income quartile and type of school (public, private and mixed). The acceptance rate of the invited schools was about 25%, with most fifth-grade classes within participating schools taking part and limited within-class non-participation (due to illness or lack of parental consent). Children and parents compiled a sociodemographic questionnaire, and teachers reported information on participating classes and individual students. Additionally, students took a cognitive ability test.

The behavioral tasks to capture CogEff were conducted inside an on-campus university laboratory in Madrid and in a lab-in-the-field setting within the schools in Berlin. Students carried out three real-effort tasks (adopted from behavioral economics and cognitive psychology) on desktop/laptop computers under different incentive conditions (see supplementary materials C for details). We only used the tasks performed with the extrinsic piece-rate condition because the presence of stakes is crucial to extrapolate to real-life educational settings. We carried out a robustness check using a variable of CogEff constructed with all the incentive conditions in the supplementary materials (Table S.4.). The results are substantively similar to the results with our main extrinsic measure.

The computerized tasks were selected to minimizing the influence of ability and engage three core executive functions subdomains: (i) the “Slider Task”, a well-known task in experimental economics that focuses on goal maintenance, information processing and updating (Gill and Prowse 2019); (ii) the "Simon Task", an established cognitive psychology task that focuses on regulation, attention and inhibitory control (Cespón, Galdo-Álvarez, and Díaz 2016); (iii) the "AX-Continuous Performance Task", another standard psychological task that measures cognitive flexibility and switching between proactive and reactive control (Gonthier et al. 2016).[2] These real-effort tasks were designed to be demanding yet straightforward and new to the subjects, aiming to shield performance from the influence of ability, knowledge or previous experience.

Table 1. Descriptive statistics of the imputed analytical samples (n = 1,270)

Variable	Mean	SD	Min	Max
Z-GPA	0.00	1.00	-4.03	2.12
Z-Language Grade	0.00	1.00	-3.90	2.10
Z-Math Grade	0.00	1.00	-3.46	2.05
Language Grade	3.78	1.00	1	5
Math Grade	3.76	1.00	1	5
Z-IQ	0.00	1.00	-5.20	3.58
Z-Extrinsic Cognitive Effort	0.00	1.00	-3.65	2.43
Z-Full Cognitive Effort	0.00	1.00	-3.48	2.62
Z-Teacher-perceived Effort	0.00	1.00	-3.70	2.24
Teacher-perceived Effort	7.36	1.92	0	10
Parental Tertiary Education	0.52		0	1
Parental ISEI	47.79	17.55	10	89
Parental Salariat Class	0.48		0	1
Foreign-born Parents	0.28		0	1
Sex: Female	0.52		0	1
Age in Months	127.69	6.47	112	165
City: Madrid	0.62		0	1
School SES Quartile	2.49	1.12	1	4
Public School	0.67		0	1

3.2. Multiple Imputation

Some student data were missing due to incomplete or illegible survey responses. Missingness was especially common for parental SES variables (ranging from 14% to 20%), where nonresponse to the parent questionnaire could introduce selection bias in a complete-case analysis (e.g., by underrepresenting students from lower SES backgrounds). To prevent this, we employed multiple imputation by chained equations, incorporating all variables used in the analysis and relevant auxiliary variables to minimize bias under the missing-at-random assumption. We did not impute the dependent variable (GPA) (only 2.9% missing), as imputing the outcome can artificially inflate associations with predictors. Ten imputed datasets were generated, each representing 1,270 students and totaling 12,700 observations. Regression models pool the coefficients and standard errors across imputed datasets using Rubin's combination rules. Table 1 reports descriptive statistics for the variables in the imputed samples.

3.3. Variables

Grade point average (GPA). The main dependent variable is the students' GPA in language (Spanish in Madrid and German in Berlin) and math in the last official school report cards. This information was taken from the parental questionnaire, or from the children's questionnaire if the former was not available, on a 5-point scale, where 1 is Insufficient (1-4), 2 is Sufficient (5), 3 is Good (6), 4 is Noteworthy (7-8) and 5 is Excellent (9-10). To account for differences in teacher/school grading standards across classrooms (Calsamiglia and Loviglio 2019), grades in math and language were z-standardized within classrooms. These standardized scores were averaged to create a GPA, reflecting relative performance in both subjects within their classroom. Finally, this GPA measure was standardized again across all students to facilitate interpretation. As a robustness check (see supplementary materials Table S.1.), we replicated the main models by math and language grades separately.

Cognitive effort (CogEff). We measure CogEff across six rounds per student in the extrinsic piece-rate incentive condition—two rounds for each real-effort task. Each round's score represents the number of correct answers within a two-minute interval. If a participant chose to play the game instead of completing the task, that round's score was recorded as missing. Thus, our measure captures "effort intensity" rather than "effort direction" (Van Iddekinge et al. 2023). Student performance is recorded as the average number of correct responses per task round, standardized (z-score) within the distribution of scores for each task across all rounds, and then summed across the three tasks. To account for technical features of implementation that might affect performance but that are independent of effort, the scores are residualized for task order, computer videogaming frequency, and mouse use (which is relevant for the slider task), and finally re-standardized.

Teacher-perceived effort (TpEff). TpEff of the student was gathered in a survey administered to the teacher, who rated their perception of the amount of academic effort of each student in the class during the last school year on a scale from 0 to 10. We use two variants of this measure. One is z-standardized across the whole sample, and another is z-standardized by classroom to allow comparability across teachers/schools.[3] Figure S.4. (Supplementary materials B) shows the correlation of the objective (CogEff) and subjective (TpEff) effort measures, which is low at 0.19 (p-value < 0.000), suggesting high variability between behavioral and reported measures of effort, and a large room for teacher idiosyncratic judgments.

Fluid Intelligence (IQ). To ensure that effort is not capturing ability, and vice versa, we control for cognitive ability in all models. Cognitive ability is measured by fluid intelligence (abstract reasoning and problem-solving) using the standard version of the Raven's Progressive Matrices Test. Children had 5 minutes to complete as many matrices as possible. The total number of correct matrices was standardized across the full distribution of scores.

Parental SES. The socioeconomic status (SES) background of the students was measured with parental education. We constructed a dummy variable that is 1 for those students with at least one parent with tertiary education and 0 otherwise. We prioritized parental education as it aligns most closely with children's skill development and educational outcomes. Besides, parental education data exhibited the least amount of missingness (14%) and likely lower measurement error. As a robustness check (see supplementary materials Table S.3.), we replicated the findings with alternative, occupation-based measures such as the highest parental *International Socioeconomic Index of Occupational Status* (ISEI) or social class (1 = Salariat; 0 = else).

Sociodemographic covariates. In all models, we control for the following set of sociodemographic variables. At the individual level, we control for migration background (at least one parent [mother/father that filled the survey] was not born in the country = 1; else = 0); age in months; and sex (0 = Female; 1 = Male). At the school level, we control for the city/country where the experiment took place (0 = Madrid; 1 = Berlin); the type of school (0 = Private or semi-private; 1 = Public); and the school neighborhood income, indicating the quartile of the average household income of the neighborhood the school was in. As a robustness check, we separate analyses by the experiment city (see supplementary materials Table S.2.).

3.4. Methods

We use hierarchical two-level linear models with restricted maximum likelihood to account for the clustering of students (level 1) within classrooms (level 2). A random intercept is included to capture between-classroom variation in baseline GPA, reflecting the nested data structure. As a robustness check (see supplementary materials Table S.7.), we also estimate classroom fixed-effect models replicating the main multilevel random-effects models.

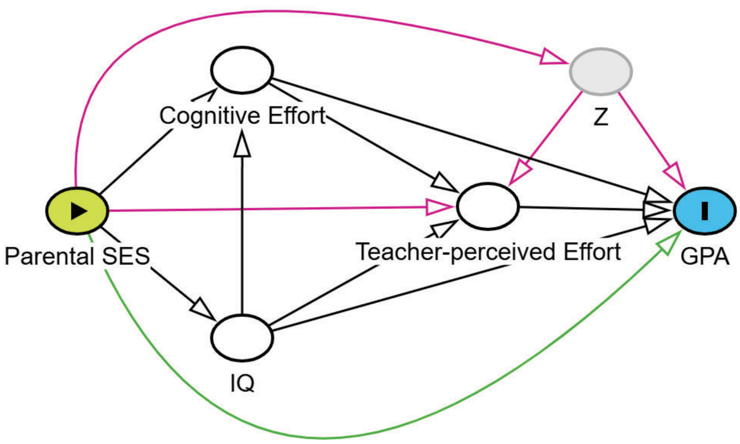
$$GPA_{ij} = (\beta_0 + \mu_j) + \beta_1 SES_{ij} + \beta_2 IQ_{ij} + \beta_3 CogEff_{ij} + \beta_4 TpEff_{ij} + \boldsymbol{Z_{ij}} + \boldsymbol{Z_j} + \varepsilon_{ij} \tag{1}$$

Equation (1) represents the main full model (M4), where GPA_{ij} is the academic grade of student i in the school classroom j. β_0 is the general intercept across all the clusters and μ_j is the random term that allows for variation around the intercept for each school classroom. At the student individual level, β_1 is the general slope of parental SES, β_2 stands for fluid intelligence, β_3 for CogEff, β_4 for TpEff, and $\boldsymbol{Z_{ij}}$ is a vector of covariates (age in months, sex, and migration background). At the school/classroom level, $\boldsymbol{Z_j}$ is a vector of covariates (city of the experiment, type of school, and school SES). Finally, ε_{ij} represents the error term. In the main analysis, we estimate four stepwise models: model 1 (M1) is a baseline model estimating the total effect of parental SES on GPA unconditional on student ability, only adjusting for sociodemographic covariates; model 2 (M2) includes fluid intelligence as a covariate; model 3 (M3) further adjusts for CE; and model 4 (M4) finally controls for TpEff. In M4, we can identify residual SES effects net of student ability: IQ and objective and subjective measures of effort. The order of inclusion of these variables implies a theoretical causal model that we develop in the following subsection on mediation.

3.4.1. Mediation

To implement the mediation analysis of parental SES on GPA, we estimate indirect effects via student ability and direct effects net of ability. Despite our data being cross-sectional, we assume a theoretical causal chain based on previous research on skill and child development. As theoretically illustrated in the directed acyclic graph in Figure 1, parental SES (educational attainment) is considered an ascribed variable (exogenous exposure) that precedes the sequential mediators and the outcome. Regarding the order of the mediators, we considered fluid intelligence and cognitive effort as quasi-parallel variables that are reciprocally influenced (Kyllonen and Kell 2018), but for simplicity we established a unidirectional arrow from IQ to CogEff. Both subsequently impact GPA directly and indirectly through teacher perceptions of student effort.[4] We assume that TpEff should not have a meaningful impact on relatively stable individual traits, such as IQ (proxied by fluid intelligence test) and CogEff (proxied by performance of executive functions tasks).

Figure 1. Directed acyclic graph of the theoretical causal model



Notes: Z is a potential unobserved confounder positively associated with the exposure, mediator TpEff and outcome I (or only with TpEff and I). The pink arrows represent potentially biased associations due to unobserved confounding. The green arrow illustrates the potential direct causal effect of SES.

We followed two empirical approaches. To assess whether effort mediates the intergenerational transmission of educational inequality, we first examine the association between parental SES and students' IQ, CogEff and TpEff. As meaningful mediation requires that the mediators be linked to both the independent (parental SES) and dependent (GPA) variables, we begin by estimating a series of linear multilevel random-effects models. These models assess the marginal effect of parental tertiary education on CogEff and TpEff, along with IQ for benchmarking, adjusting for relevant covariates at the student level (sex, age in months, migration background, and IQ; or CogEff in the IQ model), and the school level (school SES, type of school, and city). This preliminary step establishes the foundational associations required for mediation before proceeding to the full model, where we decompose the total effect on GPA with structural equation models.

Second, we estimated a near-saturated structural equation model (SEM) to simultaneously estimate all direct and indirect paths from parental SES to GPA through the sequential mediators IQ, CogEff, and TpEff, as depicted in Figure 1. The model also includes all covariates from the previous regression analyses. As shown in the correlation matrix in the supplementary materials B (Figure S.4.), these theoretical pathways are empirically supported, including through partial and semi-partial correlations with GPA. This approach captures both the sequential mediation chain (SES → IQ → CogEff → TpEff → GPA) and any direct effects, while accounting for correlations among mediators and measurement error inherent in psychometric constructs, thereby improving estimate accuracy. Following the single-indicator latent variable approach, we modeled IQ and CogEff as latent variables with (conservatively) fixed measurement error variances corresponding to a reliability of 0.7 (Cronbach's α), according to previous research for IQ (Langener et al. 2022) and our estimates for CogEff from the three tasks and two rounds. Unlike some causal mediation techniques (Imai, Keele, and Yamamoto 200), SEM models all mediators and their interrelations simultaneously within a unified framework. This allows for a comprehensive decomposition of the total effect of parental SES into relative shares mediated by each pathway, including complex sequential paths.

However, like causal mediation methods, SEM rely on key causal assumptions, particularly that there is no unmeasured confounding between SES, the mediators, and GPA (sequential ignorability). Violations of these assumptions, especially confounding between mediators and GPA, may bias indirect effect estimates and limit causal interpretation despite SEM's strengths in modeling complex mediation. A key limitation concerns potential confounding between mediators and the outcome, particularly for TpEff. If unobserved factors (represented as *Z* in Figure 1), such as parental involvement, student domain-specific knowledge, or classroom behavior, positively influence both TpEff and GPA, the estimated indirect effect of SES via TpEff may be upwardly biased, notably if they are also correlated with SES. As a result, the indirect effect via TpEff may reflect these unmeasured influences rather than a true causal mechanism, distorting the decomposition of total, direct, and indirect effects. In the supplementary materials A (Figure S.2.), sensitivity analysis to mediation estimates by unobserved confounding yields moderate robustness.

3.4.2. Moderation

To test our hypotheses on the heterogeneous returns of effort by parental SES, we estimate two additional hierarchical two-level linear models including the same vector of controls ($\boldsymbol{Z_{ij}} + \boldsymbol{Z_j}$) as in M4: in model 5 (M5) in equation (2), we introduce a two-way interaction term between parental SES and CogEff (β_5); and in model 6 (M6) in equation (3) we interact parental SES with TpEff (β_5), while controlling for the objective measure of effort (CogEff).

$$GPA_{ij} = (\beta_0 + \mu_j) + \beta_1 SES_{ij} + \beta_2 IQ_{ij} + \beta_3 CogEff_{ij} + \beta_5 SES_{ij} * CogEff_{ij} + \boldsymbol{Z_{ij}} + \boldsymbol{Z_j} + \varepsilon_{ij} \quad (2)$$

$$GPA_{ij} = (\beta_0 + \mu_j) + \beta_1 SES_{ij} + \beta_2 IQ_{ij} + \beta_3 CogEff_{ij} + \beta_4 TpEff_{ij} + \beta_5 SES_{ij} * TpEff_{ij} + \boldsymbol{Z_{ij}} + \boldsymbol{Z_j} + \varepsilon_{ij} \quad (3)$$

Finally, to account for potential non-linearities, we run a robustness check with nonparametric specifications of effort (see supplementary materials A; Figure S.1.), and the findings from the main parametric models replicate. Additionally, we run M5 controlling for TpEff and M6 without controlling for CogEff (see supplementary materials A; Table S.5.), and the interactions' sign and slope are highly consistent.

4. Results

Table 2 displays the main results from multilevel random-effects models to test the research hypotheses on student effort as a predictor of grades (M3: H1a; M4: H1b), the residual effect of parental SES, net of ability (M4: H3), and moderation (M5-M6: H4). Model 1 in Table 2 presents the total effect of parental SES on GPA, unconditional on students' ability or effort. Students with tertiary-educated parents achieve, on average, a GPA that is about half an SD unit higher ($\beta = 0.461$; p-value < 0.001) than students with non-tertiary-educated parents.

4.1. Effort as a Predictor of Academic Performance

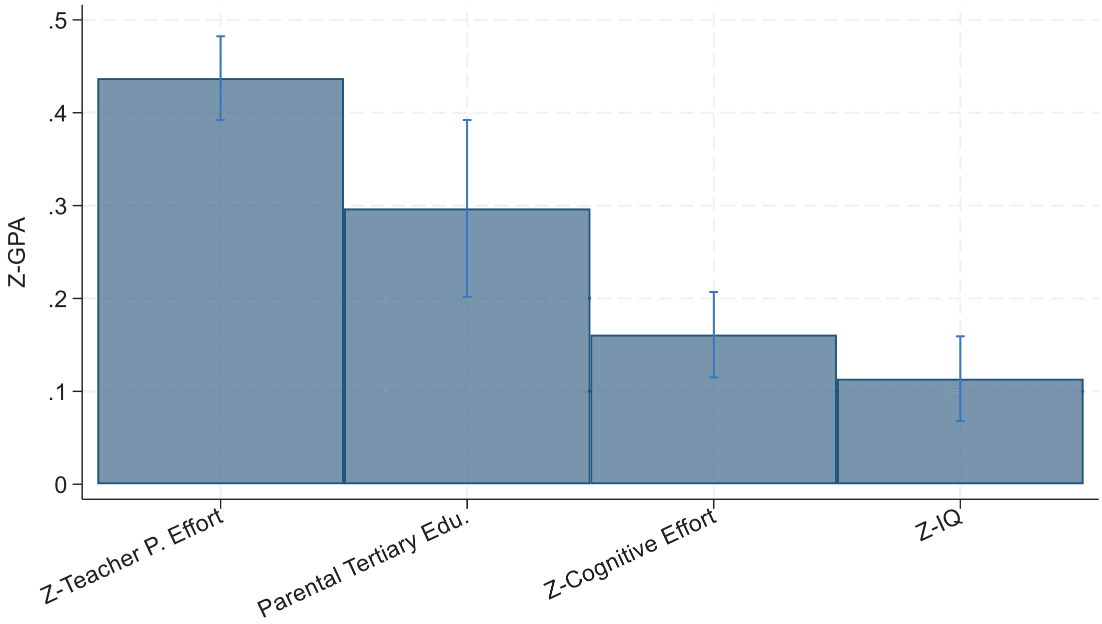
The results from M3 show that CogEff is indeed significantly and positively associated with students' school grades, adjusting for IQ, at 22% of an SD GPA unit ($\beta = 0.219$; p-value < 0.001). This finding provides support for H1a based on the predictive power of our behavioral measure of student effort, with a notable effect size between that of fluid intelligence and parental education. Moving on to M4 to test H1b on the predictive role of TpEff, this subjective measure of student effort has a strong association with GPA, equivalent to almost half an SD unit ($\beta = 0.469$; p-value < 0.001), and the total effect of parental SES, thereby validating H1b. Net of student objective cognitive ability and effort, students who are generally perceived as more hardworking by their teachers get a considerable grade premium. To better put effect sizes into perspective, Figure 2 displays the marginal effects from M4 in Table 2 for students' effort, ability, and family SES background. Further in line with H1b, TpEff has the largest net effect size, adjusting for the other predictors, more than doubling the impact of the direct effect of parental education ($\beta = 0.225$; p-value < 0.001), IQ ($\beta = 0.106$; p-value < 0.001) and CogEff ($\beta = 0.147$; p-value < 0.001).

Table 2. Linear multilevel random-effects models

	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Parental Tertiary Edu. (SES)	0.461***	0.383***	0.333***	0.226***	0.331***	0.225***
	(0.063)	(0.062)	(0.062)	(0.059)	(0.062)	(0.059)
Z-IQ		0.250***	0.187***	0.103***	0.186***	0.106***
		(0.028)	(0.028)	(0.026)	(0.028)	(0.026)
Z-Cognitive Effort (CogEff)			0.219***	0.148***	0.191***	0.147***
			(0.029)	(0.025)	(0.038)	(0.025)
Z-Teacher-perceived Effort (TpEff)				0.469***		0.547***
				(0.025)		(0.035)
SES*Z-CogEff					0.062	
					(0.055)	
SES*Z-TpEff						-0.156**
						(0.048)
R^2	0.078	0.132	0.171	0.360	0.171	0.368
Observations	1,270	1,270	1,270	1,270	1,270	1,270
Number of clusters	66	66	66	66	66	66

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05
Controls: sex, age in months and migration background (level 1); type of school; school SES; and city (level 2).

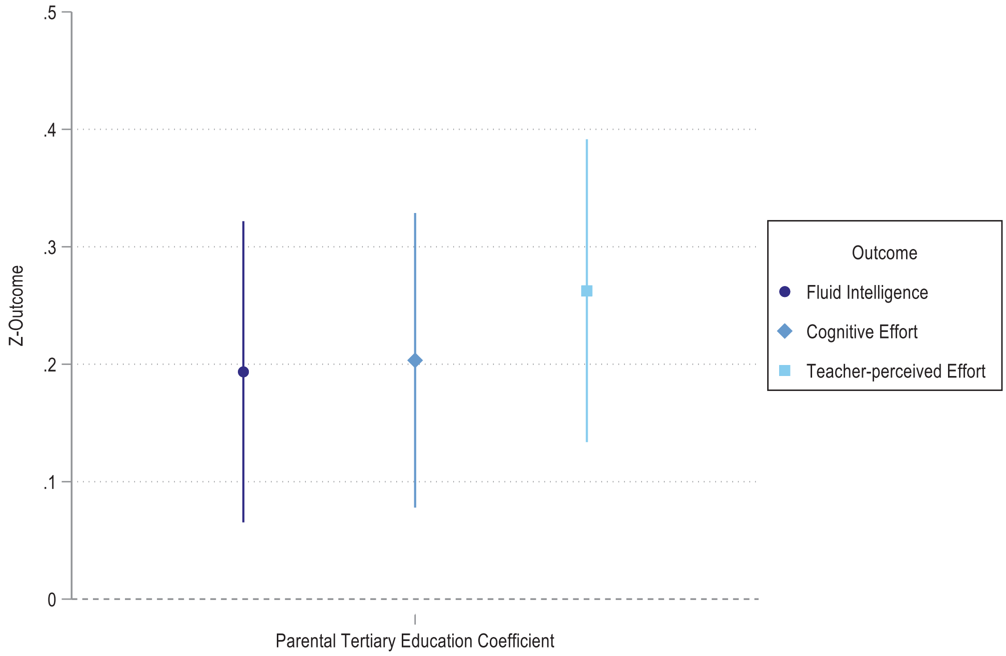
Figure 2. Marginal effects of predictors on Z-GPA (Model 4, Table 2) with 95% CI



4.2. Mediation

We present the results from the partial and full mediation analyses, which formally test H2a (Figure 3), H2b, and H3 (Table 3). Figure 3 displays the coefficients of parental tertiary education from linear multilevel random-effects models on different student outcomes, testing for SES inequalities as the first necessary step in the full mediation chain from parental SES to GPA. All student outcomes are positively and significantly related to parental education, validating H2a and the first condition for the subsequent mediation analysis. The association between parental SES and students' IQ and CE, net of each other, is equivalent to an SES ability gap of 19% (p-value < 0.05) and 20% (p-value < 0.001) of an SD, respectively. At the same time, the SES gap in TpEff, net of IQ, is slightly more substantial, with 26% of an SD, supporting H2b on larger SES gaps in subjective than objective effort.[5]

Figure 3. Marginal effects of parental tertiary education on effort and ability with 95% CI (n=1,270)



Notes: Linear multilevel random-effects model; controls: sex, age in months, migration background, and IQ (except for the IQ model) (level 1); type of school; school SES; and city (level 2); additional controls in IQ model: cognitive effort; CogEff and IQ are standardized across classrooms.

Table 3. Mediation analysis of SES on GPA via student ability using SEM (n=1,270)

	Coeff. (z-scores)	SE	% Total Effect	% Indirect Effect
Total Effect	0.467***	0.060	100	
Direct Effect	0.223***	0.059	47.8	
Indirect Effect	0.244***	0.036	52.2	100
via TpEff	0.109***	0.031	23.4	44.9
via CogEff	0.051***	0.011	11	21
via IQ	0.083***	0.014	17.8	34.1
TpEff - CogEff	+0.058+	0.033	+12.4	+23.8

Notes: p-value *** < 0.001 + < 0.1; SE=Standard Error; Controls: sex, age in months and migration background, type of school, school SES, and city; CogEff and IQ adjusted for measurement error (α = 0.7).

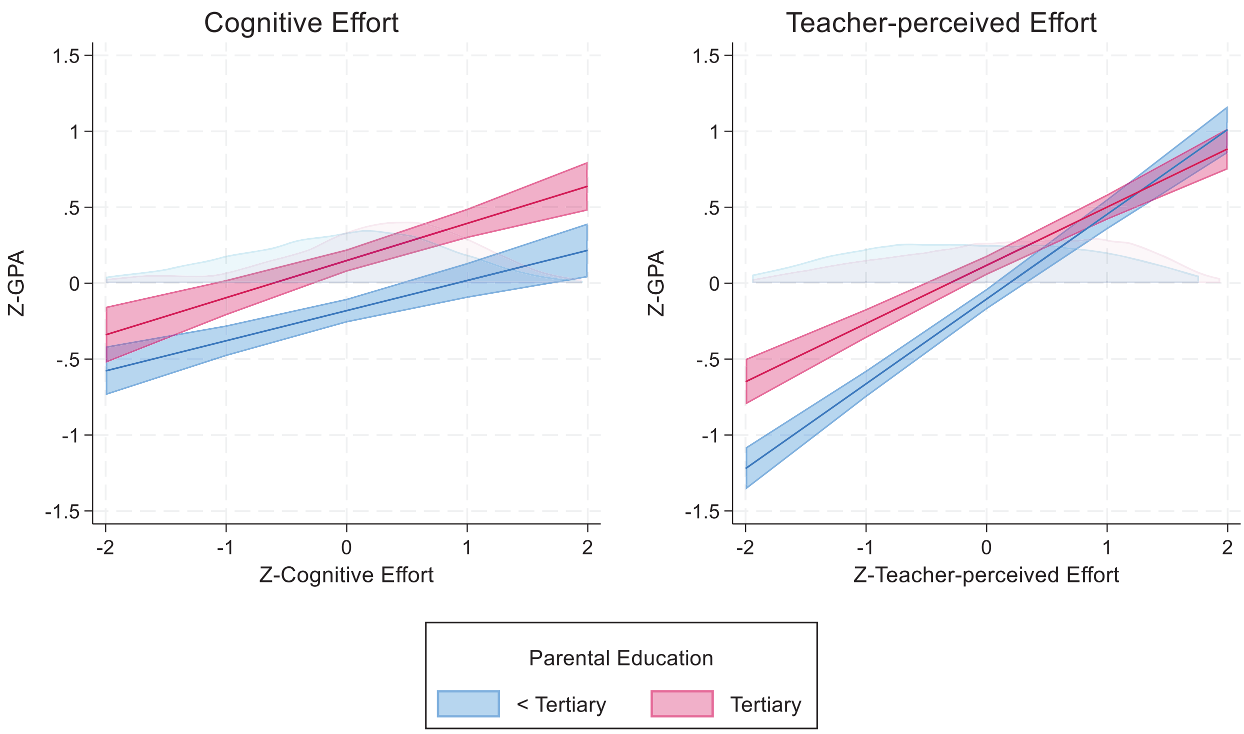
Next, once significant SES gaps in student ability have been validated, in Table 3, we present the SEM mediation analysis of parental education on GPA, identifying indirect effects via student ability (H2b) and direct effects of social background (H3). The indirect effect of parental SES via student observed ability and effort (TpEff + CogEff) accounts for 52.2% (p-value < 0.000) of the total effect. If we disentangle the role of each mediator in the indirect effect (0.244; p-value < 0.000), TpEff accounts for 44.9 % ($\beta = 0.109$; p-value < 0.001), IQ for 34.1% ($\beta = 0.083$; p-value < 0.001) and CogEff for 21 % ($\beta = 0.051$; p-value < 0.001) of the mediated paths. This supports H2b on the mediating role of student effort in explaining SES gaps in academic performance, with TpEff being a more substantial mediator than CogEff, accounting for 23.8 percentage points (p-value < 0.1) more of the indirect effect than CogEff. This higher relative contribution of TpEff aligns with its higher association with parental education (Figure 2) and GPA (Figure 3).

4.3. Moderation

In Table 2, M5-M6, we test H4, including the interaction terms between student effort (CogEff in M5; TpEff in M6) and parental education. On the one hand, contrary to H4 testing the compensatory advantage mechanism, the interaction between CogEff and tertiary parental education in M5 is of positive sign and statistically non-significant ($\beta = 0.062$; p-value > 0.1), close to a null effect. On the other hand, M6 shows that the interplay between TpEff and parental tertiary education is negative and statistically significant ($\beta = - 0.156$; p-value < 0.01), with a substantial effect size aligning with H4.

For better illustration, the marginal effects of CogEff (left-hand panel) and TpEff (right-hand panel) on GPA, by parental education, are shown in Figure 4. Regarding CogEff, its impact appears relatively independent of parental education, rejecting H4 since SES gaps in GPA seem fairly stable over the CogEff distribution. In turn, the right-hand panel on TpEff displays that, on average, low-SES students ($\beta = 0.562$; p-value < 0.001) get higher GPA returns to TpEff than high-SES students ($\beta = 0.371$; p-value < 0.001). When teachers perceive them to lack effort (from -2 to -1 SDs), low-SES students are much more penalized in grading than their high-SES peers. Thus, we find the largest SES gaps among low-effort students, at 0.572 (at TpEff = -2 SD; p-value < 0.001) and 0.397 (at TpEff = -1 SD; p-value < 0.001) compared to the average at 0.225 (at TpEff = 0; p-value < 0.001). This means that having parents with tertiary education partially buffers the negative impact of the teacher perceiving low effort on school grades, to some extent, supporting H4.

Figure 4. Marginal effects of the Effort*SES interactions in M5 and M6 (Table 2) with 95% CI



Notes: Effort distributions by parental education are shown as shades in the background.

5. Conclusion and Discussion

This study investigated the multifaceted impact of student effort on academic achievement and the reproduction of educational inequality. Effort is a central factor in academic merit, both normatively and empirically, but remains understudied due to measurement challenges, especially in sociological research. We implemented a novel design that draws on advances in cognitive psychology and experimental economics and features a behavioral measure with high internal validity employing three real-effort tasks, each tapping into different executive functions, in the presence of extrinsic incentives. The design also has high external validity, building on laboratory (in-the-field) data from balanced school-based samples of fifth graders in Madrid (Spain) and Berlin (Germany). Furthermore, we compared our objective measure of cognitive effort (CogEff) with subjective perceptions of teachers, the decisive evaluators of academic merit. Using and comparing both objective and subjective effort measures, we analyzed the predictive roles of effort vis-à-vis cognitive ability (IQ) in school grades (GPA), along with the attendant mediating and moderating roles in the association between parental SES and GPA, as well as residual GPA gaps by parental SES unaccounted for by merit factors. In this way, we scrutinized the contribution of effort to the intergenerational transmission of educational inequality.

We report four main findings that generally align with our hypotheses; we also discuss their theoretical implications and directions for future research. First, consistent with H1a, our laboratory effort measure is positively predictive of academic performance, net of IQ (fluid intelligence), with remarkable effect sizes comparable to those of IQ or parental education. This finding points to strong average returns to cognitive effort for educational performance that—largely due to measurement shortcomings—have remained underappreciated in previous research (Palacios-Abad 2021). This evidence also supports the internal and ecological validity (from the lab to the field) of our behavioral approach as a valid measure of student effort. Similarly, teacher-perceived effort (TpEff), net of both CogEff and IQ, is also a strong predictor of school grades, but with effect sizes more than twice those of CogEff, IQ, or parental education. This benchmarking further supports H1b, raising the question of whether TpEff is so consequential for grading because it best captures the set of student domain-specific knowledge and classroom behavior that, according to official criteria, teachers should evaluate, or whether it reflects biases. On its face, this finding suggests that teachers might conflate academic skills and pure effort. Nevertheless, the exact mechanism by which teacher effort perceptions affect grading remains unclear and is a worthy avenue for further research. Future research would also benefit from using task-based effort measures, ideally also coupled with standardized competence tests, to overcome the shortcomings of survey-based accounts.

Second, the distribution of CogEff and TpEff, net of IQ, is unequal by parental SES, with children from tertiary-educated families showing, on average, between 20% (CogEff) and 30% (TpEff) of a SD unit more effort, respectively, than their more disadvantaged peers. This evidence confirms a necessary condition for effort being a meaningful mediator in the intergenerational transmission of educational inequality, supporting H2a. This finding might also imply potential teacher bias in their (over)perception of student effort among high-SES students and/or other unobserved skills positively associated with both parental background and TpEff. Further validating H2b, both effort measures, and particularly the teachers' perceptions, significantly contribute to the intergenerational reproduction of educational inequality, as they mediate up to 34% of the total observed gaps in GPA by parental SES, accounting for up to 66 % (CogEff: 21 %; TpEff: 44.9 %, IQ: 34.1 %) of its indirect effect via merit factors. Our results align with and extend previous research on the role of motivational personality traits in educational stratification (Bowles and Gintis 2002; Farkas 2003; Shanahan et al. 2014). Consistent with evidence suggesting that such traits, like executive functioning and attention skills, are unequally distributed across

SES groups (Attanasio, De Paula, Toppeta and 2025), we find that effort is a significant mediator between parental SES and academic performance, with similar mediating shares as IQ.

Third, over and beyond students' cognitive ability and (objective and subjective) effort, a substantial residual association between parental SES and GPA remains, aligning with H3. Only about half (52.2 %) of the total effect of SES on academic performance is attributable to observed student effort, teacher perceptions and IQ, corrected for measurement error. Different complementary mechanisms might be behind the remaining half (47.8 %) that merit factors do not explain: e.g., greater parental support among high-SES students boosting curriculum learning; teacher bias, driven by psychological mechanisms such as competence stereotypes, or cultural capital.

Fourth, regarding the interaction between effort and parental SES, the empirical evidence gives mixed support to H4. While the academic returns to CogEff are relatively similar for both high- and low-SES students, privileged children are less penalized in their GPA when they are perceived as exerting low effort by their teachers, a pattern that aligns with the compensatory advantage and resource substitution mechanisms, as well as some previous studies (Damian et al. 2015; Liu 2019; Mele, Buchman, and Burger 2023). High-SES students receive better school grades than their less-advantaged peers when teachers perceive low effort, such that the largest educational inequalities (residual GPA gaps by SES) are observed among "lazy" students, even when cognitive skills (IQ) are equal. Unobserved competencies, attitudes, and behaviors predominantly found in advantaged families might be effectively boosting high-SES students' learning and knowledge of the curriculum, despite being perceived as lazy by their teachers. Thus, teachers might consider these factors as legitimately deserving of higher success expectations and grades, or they might display systematic favoritism for advantaged students due to competence stereotypes and/or parental pressure for higher grades.

Overall, the results hold up across several robustness and heterogeneity checks; however, we acknowledge three limitations of our research design that future studies should overcome. First, lacking longitudinal data, we could not rule out reciprocal effects or feedback loops between our measures of academic merit and school grades, which are retrospective reports. Yet, our objective measures of cognitive ability (IQ) and effort (executive functions), which we further corrected for measurement error, are arguably relatively stable personality traits that should be considered antecedents to both TpEff and GPA. Second, in the mediation analysis, since parental SES and the mediators are not randomly assigned, we could not claim causality due to the potential presence of unobserved confounders, particularly between the mediators (TpEff) and the outcome. Counterfactual sensitivity analysis showed that the indirect effect via TpEff would vanish only under moderate confounding, suggesting reasonable robustness but warranting cautious interpretation. Third, due to the limited sample size and power, we were unable to explore systematic heterogeneity between Spain and Germany, even though the overall similarity in findings from both subsamples appears to signal a limited role played by the educational system and national context.

Despite these limitations, our novel research design exploiting a robust multitask behavioral measure of effort offered a significant contribution to research on educational inequality. Overall, cognitive effort predicts academic success net of parental SES, an encouraging finding for student agency and equal opportunity, but also shapes educational (in)equality, primarily through teacher perceptions. TpEff seems to be shaped by unobserved factors beyond objective measures (IQ and CogEff), which systematically advantage students from higher-SES families. High-SES students show higher average effort, and even when perceived as lazy, avoid harsh grade penalties. In contrast, low-SES students show lower average effort and are even less likely to be perceived as hardworking by their teachers. Still, when that is the case, they can partially overcome the disadvantage of social origin and achieve disproportionately high grades. The varying returns to teacher-perceived effort uncovered in this study represent a new stimulus and their Janus-faced implications a challenge for normatively charged debates about meritocracy in education: should we celebrate that disadvantaged youth can overachieve if they manage, against all odds, to be recognized as hardworking, or should we condemn the unfairness of teachers' resistance to penalize privileged youth they see as lacking effort? This conundrum highlights the need to further scrutinize the gap between students' actual effort and how it is perceived, as teachers' subjective judgments—as gatekeepers of academic merit—may ultimately exert a greater influence on educational achievement than objective effort dispositions themselves.

Endnotes

1. Previous scholarship, particularly the economics literature, often uses the concept "noncognitive skills" as a catch-all term encompassing personality traits and behaviors expressed through thoughts, emotions, and patterns of conduct, including emotional regulation, self-control, executive functions, behavioral problems, and motivation. However, the term "noncognitive skills" is misleading, as many of the characteristics inherently rely on core cognitive processes like attention and self-regulation. Crucially, effort is not adequately thought of as a skill or ability according to theories of equality of opportunity (Roemer and Trannoy 2016). Therefore, we use the more accurate terms "socio-behavioral traits" or "personality traits" as umbrella concepts referring to these individual characteristics.

2. The order of the tasks varied across classes and experimental sessions to avoid order effects.

3. In the supplementary materials, we run a robustness check for the mediation analysis (Figure S.3.) and main models (Table S.6.) using teacher-perceived effort and grades normalized by the whole sample, respectively.

4. This path reflects potential differences in the propensities of high and low-ability students to exert effort. There is a legitimate argument that, additionally, effort determines ability through engagement in learning. However, since this effect would arguably only play out over time and could not be captured in cross-sectional data like ours, we prefer a parsimonious causal model that does not overtax the complex estimations.

5. When standardized between classrooms, the estimated effect of TpEff rises to 31% of an SD (p-value < 0.001) (see supplementary materials Figure S.3.).

References

Apascaritei, P., Demel, S., Radl, J. 2021. "The Difference Between Saying and Doing: Comparing Subjective and Objective Measures of Effort Among Fifth Graders." *American Behavioral Scientist* 65(11):1457-1479.

Asadullah, M.N., et al. 2021. "Measuring educational inequality of opportunity: pupil's effort matters." *World Development* 138, 105262.

Attanasio, O.P., De Paula, Á., Toppeta, A. 2025. "Intergenerational mobility in socio-emotional skills." *Journal of Public Economics 248*, 105423.

Batruch, A., et al. 2023. "Are Tracking Recommendations Biased? A Review of Teachers' Role in the Creation of Inequalities in Tracking Decisions." *Teaching and Teacher Education* 123, 103985.

Bernardi, F, Valdés, M.T. 2021. "Sticky Educational Expectations: A Cross-Country Comparison." *Research in Social Stratification and Mobility* 75, 100624.

Bernardi, F. 2014. "Compensatory Advantage as a Mechanism of Educational Inequality: A Regression Discontinuity Based on Month of Birth." *Sociology of Education* 87(2):74–88.

Borghans, L., Schils. T. 2012. "The Leaning Tower of Pisa. Decomposing achievement test scores into cognitive and noncognitive components". *JOLE 2012 Working Paper Proceedings*.

Borghans, L., et al. 2016. "What Grades and Achievement Tests Measure." *Proceedings of the National Academy of Sciences* 113(47):13354LP–13359.

Borgonovi, F., Ferrara, A. Piacentini, M. 2023. "From Asking to Observing. Behavioural Measures of Socio-Emotional and Motivational Skills in Large-Scale Assessments." *Social Science Research* 112(May):102874.

Boudon, R. 1974. *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*. New York: Wiley.

Bourdieu, P., Passeron, J.-P.. 1990. *Reproduction in Education, Society, and Culture*. London: Sage.

Bowers, A. J. 2011. "What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school." *Educational Research and Evaluation* 17(3):141-159.

Bowles S, Gintis H. 1976. *Schooling in Capitalist America*. New York: Basic Book

Bowles, S., Gintis, H. 2002. "Schooling in Capitalist America Revisited. *Sociology of Education*." 75:1-18

Burger, K., Brack, N. 2025. "Origins, belonging, and expectations: Assessing resource compensation and reinforcement in academic educational trajectories." *Social Forces 103*(3):905-927.

Calarco, Jessica M. 2014. "Coached for the Classroom: Parents' Cultural Transmission and Children's Reproduction of Educational Inequalities." *American Sociological Review* 79:1015–1037.

Calsamiglia, C., Loviglio, A. 2019. "Grading on a Curve: When Having Good Peers Is Not Good." *Economics of Education Review* 73:101916.

Carbonaro, W. (2005). Tracking, students' effort, and academic achievement. *Sociology of Education*, 78(1), 27-49.

Cespón, J., Galdo-Álvarez, S., Díaz, F. 2016. “Cognitive control activity is modulated by the magnitude of interference and pre-activation of monitoring mechanisms.” *Scientific reports* 6(1):1-11.

Chen, Y., Feng, S., Heckman, J.J., Kautz, T. 2020. “Sensitivity of Self-Reported Noncognitive Skills to Survey Administration Conditions.” *Proceedings of the National Academy of Sciences* 117(2): 931–35.

Conti, G., Heckman, J.J. 2014. “Understanding conscientiousness across the life course: An economic perspective.” *Developmental Psychology* 50(5):1451–1459.

Damian, R.I., et al. 2015. “Can personality traits and intelligence compensate for background disadvantage? Predicting status attainment in adulthood.” *Journal of Personality and Social Psychology* 109(3):473.

Diamond, A. 2013. “Executive functions.” *Annual Review of Psychology* 64:135-168.

DiPrete, T.A., Eirich, G.M. 2006. “Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments.” *Annual Review of Sociology* 32:271–297.

Duckworth, A.L., Kern, M.L. 2011. “A meta-analysis of the convergent validity of self-control measures.” *Journal of Research in Personality* 45(3):259-268.

Duckworth, A.L., Peterson, C., Matthews, M.D., Kelly, D.R. 2007. “Grit: perseverance and passion for long-term goals.” *Journal of Personality and Social Psychology* 92(6):1087.

Duncan, G.J., Magnuson, K. 2011. “The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems.” In: Duncan G., Duncan G., and Murnane R. (Eds.), *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (pp. 47-70). Russell Sage Foundation.

Farkas, G. 2003. “Cognitive skills and noncognitive traits and behaviors in stratification processes.” *Annual Review of Sociology* 29(1):541-562.

Fazio, R.H., et al. 2023. “Implicit Bias: What Is It?” In *The Cambridge Handbook of Implicit Bias and Racism*, edited by Jon A. Krosnick, H. Tobias, and A.L. Scott, Cambridge: Cambridge University Press.

Ferman, B., Fontes, L.F. 2022. “Assessing Knowledge or Classroom Behavior? Evidence of Teachers’ Grading Bias.” *Journal of Public Economics* 216:104773.

Fleming, K.A., Heintzelman, S.J., Bartholow, B.D. 2016. “Specifying Associations between Conscientiousness and Executive Functioning: Mental Set Shifting, Not Prepotent Response Inhibition or Working Memory Updating.” *Journal of Personality* 84(3):348-60.

Forster, A.G., van de Werfhorst, H.G. 2020. “Navigating Institutions: Parents’ Knowledge of the Educational System and Students’ Success in Education.” *European Sociological Review* 36(1):48–64.

Geven, S., et al. 2021. “How teachers form educational expectations for students: A comparative factorial survey experiment in three institutional contexts.” *Social Science Research 100*, 102599.

Gil-Hernández, C.J. 2021. “The (unequal) interplay between cognitive and noncognitive skills in early educational attainment.” *American Behavioral Scientist* 65(11):1577-1598.

Gill, D., Prowse, V. 2012. “A structural analysis of disappointment aversion in a real effort competition.” *American Economic Review* 102(1):469-503.

Gonthier, C., et al. 2016. “Inducing proactive control shifts in the AX-CPT.” *Frontiers in Psychology* 7:1822.

Gruijters, R.J., Raabe, I.J., Hübner, N. 2023. “Socio-emotional Skills and the Socioeconomic Achievement Gap.” *Sociology of Education 97*(2):120-147.

Hackman, D.A., Farah, M.J., Meaney, M.J. 2010. “Socioeconomic status and the brain: mechanistic insights from human and animal research.” *Nature Reviews/Neuroscience* 11:651-659.

Heckman, J.J., Kautz, T. 2012. “Hard evidence on soft skills.” *Labour Economics* 19(4):451–464.

Holm, A., Hjorth–Trolle, A., Jæger, M.M.. 2019. “Signals, Educational Decision–Making, and Inequality.” *European Sociological Review* 35(4):447–460.

Holtmann, A.C., Menze, L., Solga, H. 2021. “Intergenerational Transmission of Educational Attainment: How Important Are Children’s Personality Characteristics?” *American Be*havioral *Scientist* 65(11):1531-1554.

Hsin, A., Xie, Y. 2017. “Life-course changes in the mediation of cognitive and noncognitive skills for parental effects on children’s academic achievement.” *Social Science Research* 63:150-165.

Imai, K., Keele, L. Yamamoto, T. 2010. “Identification, inference, and sensitivity analysis for causal mediation effects.” *Statistical Science* 25:51-71.

Jackson, M. 2013. *Determined to Succeed? Performance Versus Choice in Educational Attainment*. Stanford: Stanford University Press.

Jæger, M.M. 2022. “Cultural Capital and Educational Inequality: An Assessment of the State of the Art.” Pp. 121–134 in *Handbook of Sociological Science*, edited by K. Gërxhani, N. D. de Graaf, and W. Raub. Edward Elgar Publishing.

Kaiser, T. 2016. “Socioeconomic status, parenting and conscientiousness: the unequal development of children's focus.” *Journal of Family Studies* 23(1):62–85

Kalil, A., Ryan, R., Corey, M. 2012. “Diverging destinies: Maternal education and the developmental gradient in time with children.” *Demograph*y 49(4):1361-1383.

Kelly, S. 2008. “What types of students’ effort are rewarded with high marks?” *Sociology of Education* 81(1):32-52.

Kröger, L.K., Palacios-Abad, A., Radl, J. 2024. “Chapter 30: Noncognitive Skills and Intergenerational Inequality: Reviewing the Power of Personality”. In *Research Handbook on Intergenerational Inequality*. Cheltenham, UK: Edward Elgar Publishing.

Krohn, G.A., O'Connor, C.M. 2005. “Student effort and performance over the semester.” *The Journal of Economic Education*, 36(1):3-28.

Kyllonen, P., Kell, H. 2018. "Ability Tests Measure Personality, Personality Tests Measure Ability: Disentangling Construct and Method in Evaluating the Relationship between Personality and Ability." *Journal of Intelligence* 6(3):32.

Kwon, H.W., Erola, J. 2023. "The limited role of personal goal striving in status attainment." *Social Science Research 112*:102797.

Langener, A.M., Kramer, W., Huizenga, H.M. 2022. "A shortened version of Raven's standard progressive matrices for children and adolescents." *British Journal of Developmental Psychology 40*(1):35-45.

Liu, A. 2019. "Can noncognitive skills compensate for background disadvantage? – the moderation of noncognitive skills on family socioeconomic status and achievement during early childhood and early adolescence." *Social Science Research* 83(April), 102306.

McCrae, R.R., Mõttus, R. 2019. "What Personality Scales Measure: A New Psychometrics and Its Implications for Theory and Assessment." *Current Directions in Psychological Science* 28(4):415–420.

Mcewen, C.A., Mcewen, B.S. 2017. "Social Structure, Adversity, Toxic Stress, and Intergenerational Poverty: An Early Childhood Model." *Annual Review of Sociology* (April), 1–28.

McGue, M., et al. 2020. "The Contribution of Cognitive and Noncognitive Skills to Intergenerational Social Mobility." *Psychological Science 31*(7):835-847

Mammadov, S. 2022. "Big Five personality traits and academic performance: A meta-analysis." *Journal of Personality 90*(2):222-255.

Melamed, D., et al. 2019. "Status Characteristics, Implicit Bias, and the Production of Racial Inequality." *American Sociological Review* 84(6):1013–1036.

Mele, F., Buchmann, M., Burger, K. 2023. "Making it to the academic path in a tracked education system: The interplay of individual agency and social origin in early educational transitions." *Journal of Youth and Adolescence* 52:2620–2635.

Miller, A.D., Ramirez, E.M., Murdock, T.B. 2017. “The influence of teachers’ self-efficacy on perceptions: Perceived teacher competence and respect and student effort and achievement.” *Teaching and Teacher Education* 64:260-269.

Owens, J. 2022. “Double Jeopardy: Teacher Biases, Racialized Organizations, and the Production of Racial/Ethnic Disparities in School Discipline.” *American Sociological Review* 87(6):1007–1048.

Palacios-Abad, A. 2021. “Strive to Succeed? The Role of Persistence in the Process of Educational Attainment.” *American Behavioral Scientist* 65(11):1555-1576.

Park, H., Byun, S.Y., Kim, K.K. 2011. “Parental involvement and students’ cognitive outcomes in Korea: Focusing on private tutoring.” *Sociology of Education* 84(1):3-22.

Radl, J., Miller, L. 2021. “Conceptual and Methodological Considerations on Effort: An Interdisciplinary Approach.” *American Behavioral Scientist* 65(11):1447-1456.

Randall, J., Engelhard, G. 2010. “Examining the grading practices of teachers.” *Teaching and Teacher Education* 26(7):1372-1380.

Roemer, J.E., Trannoy, A. 2016. “Equality of Opportunity: Theory and Measurement.” *Journal of Economic Literature* 54 (4):1288–1332.

Schoner, F., Mergele, L., Zierow, L. 2024. “Grading Student Behavior.” *Labour Economics* 90, 102570.

Schoon, I., Mele, F., Burger, K. 2025. “Academic agency and educational attainment: The intersection of SES, ethnicity and sex.” *Learning and Individual Differences 120*:102689.

Shanahan, M.J., et al. 2014. “Personality and the reproduction of social class.” *Social Forces* 93(1):209-240.

Skopek, J., Passaretta, G. 2021. “Socioeconomic Inequality in Children’s Achievement from Infancy to Adolescence: The Case of Germany.” *Social Forces 100*(1):86-112.

Smithers, L.G., et al. 2018. “A systematic review and meta-analysis of effects of early life noncognitive skills on academic, psychosocial, cognitive and health outcomes.” *Nature Human Behaviour* 2(11):867-880.

Südkamp, A., Kaiser, J., Möller, J. 2012. “Accuracy of teachers’ judgments of students’ academic achievement: A meta-analysis*.” Journal of Educational Psychology* 104(3), 743.

Tobisch, A., Dresel, M. 2017. “Negatively or positively biased? Dependencies of teachers’ judgments and expectations based on students’ ethnic and social backgrounds.” Social *Psychology of Education* 20(4):731-752.

Upadyaya, K., Eccles, J. 2015. “Do teachers’ perceptions of children’s math and reading related ability and effort predict children’s self-concept of ability in math and reading?.” *Educational Psychology* 35(1):110-127.

van Huizen, T., Jacobs, M., Oosterveen, H. 2024. “Teacher Bias or Measurement Error?” arXiv:2401.04200 [econ.EM]. DOI: 10.48550/arXiv.2401.04200.

van Iddekinge, et al. 2023. “Work Effort: A Conceptual and Meta-Analytic Review.” *Journal of Management* 49(1):125–57.

Wang, S., Rubie-Davies, C.M., Meissel, K. 2018. “A systematic review of the teacher expectation literature over the past 30 years.” *Educational Research and Evaluation* 24(3-5):124-179.

Watts, T.W., Duncan, G.J., Quan, H. 2018. “Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes.” *Psychological Science 29*(7):1159–1177.

Wenz, S. E., Hoenig, K. 2020. “Ethnic and social class discrimination in education: Experimental evidence from Germany.” *Research in Social Stratification and Mobility* 65, 100461.

Zamarro, G., Hitt, C., Mendez, I. 2019. “When students don’t care: Reexamining international differences in achievement and student effort.” *Journal of Human Capital* 13(4):519-552.

Zanga, G., De Gioannis, E. 2023. “Discrimination in Grading: A Scoping Review of Studies on Teachers’ Discrimination in School.” *Studies in Educational Evaluation* 78:101284.

ONLINE SUPPLEMENTARY MATERIALS

A. Robustness Checks

Several robustness checks replicate the main models and findings from Table 2. First, we examined heterogeneity in our results by estimating models separately for math and language grades (Table S.1.) to replicate the main findings on GPA. Second, we run the analyses by each city/country where the experiment was administered instead of using a pooled sample, and results are highly stable in Madrid and Berlin (Table S.2.), with Berlin displaying a stronger effort-SES interaction. Third, we tested the robustness of our findings to alternative measures of parental SES background, including the highest parental ISEI and a binary indicator for social class (salariat vs. other); results using parental ISEI and class mirrored those obtained with parental education (Table S.3.). Fourth, we constructed an alternative measure of CogEff by combining performance across all incentive conditions (no incentive, piece-rate, and tournament), yielding similar results to those from our preferred specification using only the piece-rate condition eliciting extrinsically motivated effort (Table S.4.). Fifth, to explore potential non-linearities in ability (*interflex* package in Stata) (Imai, Keele, and Yamamoto 2010), we ran models using fully flexible effort-SES interactions (linear, binning in tertiles, and kernel specifications); results were consistent with our main parametric findings (Figure S.1.). Sixth, we conducted additional checks by re-estimating M5 (CogEff-SES interaction) while controlling for TpEff and M6 (TpEff-SES interaction) without controlling for CogEff; in both cases, the direction and magnitude of the interaction terms remained highly stable (Table S.5.). Seventh, like other causal mediation approaches, SEM relies on the absence of unobserved confounding. To assess this assumption, we implemented a counterfactual-based sensitivity analysis (*medsens* package in Stata). Results indicate that the indirect effects would be reduced to zero only if an unobserved confounder induced a residual correlation (ρ) of 0.2 (IQ and CogEff) or 0.5 (TpEff) between the mediator and the outcome (Figure S.2.). This suggests moderate robustness of the estimated mediation, but results from single-mediator analysis should be interpreted with caution. Eight, we normalized GPA and TpEff by the whole sample instead of within classrooms (Table S.6.), which slightly increased SES coefficients but qualitatively replicated all the findings. Finally, we adopted a classroom fixed effects approach to fully account for teacher-specific evaluation practices and classroom composition (Table S.7.), and the results are highly stable compared to the main random effects specifications.

Table S.1. Linear multilevel RE models estimated separately for math and language grades

Z-Math Grade						
	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Parental Tertiary Edu. (SES)	0.424***	0.346***	0.299***	0.202***	0.297***	0.201***
	(0.064)	(0.062)	(0.062)	(0.058)	(0.062)	(0.058)
Z-IQ		0.249***	0.189***	0.113***	0.188***	0.116***
		(0.028)	(0.029)	(0.026)	(0.029)	(0.026)
Z-Cognitive Effort (CogEff)			0.208***	0.144***	0.178***	0.143***
			(0.029)	(0.026)	(0.039)	(0.026)
Z-Teacher-perceived Effort (TpEff)				0.422***		0.501***
				(0.025)		(0.035)
SES*Z-CogEff					0.066	
					(0.056)	
SES*Z-TpEff						-0.158**
						(0.049)
Observations	1,270	1,270	1,270	1,270	1,270	1,270
Number of clusters	66	66	66	66	66	66

Z-Language Grade						
	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Parental Tertiary Edu. (SES)	0.414***	0.350***	0.306***	0.208**	0.305***	0.207**
	(0.066)	(0.065)	(0.066)	(0.064)	(0.066)	(0.064)
Z-IQ		0.205***	0.151***	0.073**	0.150***	0.076**
		(0.028)	(0.029)	(0.027)	(0.029)	(0.027)
Z-Cognitive Effort (CogEff)			0.190***	0.125***	0.169***	0.124***
			(0.029)	(0.027)	(0.039)	(0.027)
Z-Teacher-perceived Effort (TpEff)				0.429***		0.492***
				(0.026)		(0.037)
SES*Z-CogEff					0.046	
					(0.058)	
SES*Z-TpEff						-0.126*
						(0.051)
Observations	1,270	1,270	1,270	1,270	1,270	1,270
Number of clusters	66	66	66	66	66	66

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05
Controls: sex, age in months and migration background (level 1); type of school; school SES; and city (level 2).

Table S.2. City-specific linear RE multilevel models for Madrid and Berlin

	M1	M2	M3	M4	M5	M6
Madrid						
Parental Tertiary Edu. (SES)	0.540***	0.441***	0.378***	0.215**	0.377***	0.211**
	(0.084)	(0.081)	(0.078)	(0.071)	(0.078)	(0.071)
Z-IQ		0.344***	0.267***	0.156***	0.266***	0.159***
		(0.036)	(0.036)	(0.032)	(0.036)	(0.032)
Z-Cognitive Effort (CogEff)			0.276***	0.203***	0.268***	0.200***
			(0.035)	(0.031)	(0.049)	(0.031)
Z-Teacher-perceived Effort (TpEff)				0.466***		0.522***
				(0.030)		(0.046)
SES*Z-CogEff					0.016	
					(0.067)	
SES*Z-TpEff						-0.102+
						(0.061)
Observations	789	789	789	789	789	789
Number of clusters	35	35	35	35	35	35
Berlin						
	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Parental Tertiary Edu. (SES)	0.388***	0.337***	0.310**	0.272**	0.308**	0.278**
	(0.101)	(0.100)	(0.101)	(0.094)	(0.102)	(0.093)
Z-IQ		0.148***	0.107*	0.055	0.106*	0.050
		(0.044)	(0.046)	(0.041)	(0.046)	(0.041)
Z-Cognitive Effort (CogEff)			0.148**	0.075+	0.115+	0.085+
			(0.049)	(0.045)	(0.062)	(0.045)
Z-Teacher-perceived Effort (TpEff)				0.454***		0.553***
				(0.042)		(0.057)
SES*Z-CogEff					0.086	
					(0.098)	
SES*Z-TpEff						-0.238**
						(0.087)
Observations	481	481	481	481	481	481
Number of clusters	31	31	31	31	31	31

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, +p<0.10
Controls: sex, age in months and migration background (level 1); type of school; and school SES (level 2).

Table S.3. Alternative parental SES measures: ISEI and social class

	Parental ISEI					
	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Parental High ISEI	0.282***	0.219***	0.188**	0.109*	0.186**	0.111*
	(0.063)	(0.062)	(0.060)	(0.054)	(0.060)	(0.054)
Z-IQ		0.264***	0.196***	0.108***	0.195***	0.111***
		(0.028)	(0.029)	(0.026)	(0.029)	(0.026)
Z-Cognitive Effort (CogEff)			0.231***	0.156***	0.204***	0.158***
			(0.029)	(0.026)	(0.038)	(0.025)
Z-Teacher-perceived Effort (TpEff)				0.476***		0.539***
				(0.025)		(0.035)
ISEI*Z-CogEff					0.060	
					(0.055)	
ISEI*Z-TpEff						-0.136**
						(0.051)
Observations	1,270	1,270	1,270	1,270	1,270	1,270
Number of clusters	66	66	66	66	66	66
	Parental Class					
	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Parental Salariat Social Class	0.239***	0.174**	0.139*	0.076	0.136*	0.078
	(0.062)	(0.060)	(0.059)	(0.052)	(0.059)	(0.052)
Z-IQ		0.266***	0.197***	0.109***	0.196***	0.113***
		(0.028)	(0.029)	(0.026)	(0.029)	(0.026)
Z-Cognitive Effort (CogEff)			0.232***	0.156***	0.208***	0.155***
			(0.029)	(0.026)	(0.037)	(0.025)
Z-Teacher-perceived Effort (TpEff)				0.478***		0.545***
				(0.025)		(0.035)
Class*Z-CogEff					0.061	
					(0.057)	
Class*Z-TpEff						-0.154**
						(0.054)
Observations	1,270	1,270	1,270	1,270	1,270	1,270
Number of clusters	66	66	66	66	66	66

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, +p<0.10
Controls: sex, age in months and migration background (level 1); type of school; school SES (level 2), and city.

Table S.4. Alternative cognitive effort measure combining intrinsic, extrinsic, and tournament performance

	Full Effort					
	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Parental Tertiary Edu. (SES)	0.461***	0.383***	0.340***	0.230***	0.337***	0.228***
	(0.064)	(0.062)	(0.064)	(0.059)	(0.063)	(0.059)
Z-IQ		0.250***	0.193***	0.110***	0.193***	0.113***
		(0.028)	(0.028)	(0.025)	(0.028)	(0.025)
Z-Full Cognitive Effort (CogEff)			0.219***	0.132***	0.183***	0.133***
			(0.029)	(0.027)	(0.040)	(0.027)
Z-Teacher-perceived Effort (TpEff)				0.467***		0.548***
				(0.025)		(0.035)
SES*Z-CogEff					0.072	
					(0.056)	
SES*Z-TpEff						-0.162***
						(0.048)
Observations	1,270	1,270	1,270	1,270	1,270	1,270
Number of clusters	66	66	66	66	66	66

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05
Controls: sex, age in months and migration background (level 1); type of school; school SES; and city (level 2).

Table S.5. Model specification variants: interaction models' controls

Interaction Controls			
	M5B	**M6B**	**M7**
Parental Tertiary Edu. (SES)	0.225***	0.253***	0.223***
	(0.059)	(0.058)	(0.058)
Z-IQ	0.102***	0.144***	0.105***
	(0.026)	(0.025)	(0.026)
Z-Full Cognitive Effort (CogEff)	0.129***		0.115***
	(0.034)		(0.034)
Z-Teacher-perceived Effort (TpEff)	0.468***	0.568***	0.551***
	(0.025)	(0.035)	(0.035)
SES*Z-CogEff	0.042		0.070
	(0.048)		(0.049)
SES*Z-TpEff		-0.158**	-0.167***
		(0.049)	(0.049)
Observations	1,270	1,270	1,270
Number of clusters	66	66	66

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05
Controls: sex, age in months and migration background (level 1); type of school; school SES; and city (level 2).

Table S.6. Models with GPA and teacher-perceived effort normalized across the full sample

Between-Classroom Standardization						
	M1	M2	M3	M4	M5	M6
Parental Tertiary Edu. (SES)	0.537***	0.468***	0.423***	0.295***	0.422***	0.294***
	(0.061)	(0.059)	(0.058)	(0.054)	(0.058)	(0.054)
Z-IQ		0.259***	0.196***	0.116***	0.195***	0.118***
		(0.026)	(0.026)	(0.024)	(0.026)	(0.024)
Z-Cognitive Effort (CogEff)			0.236***	0.166***	0.225***	0.166***
			(0.026)	(0.023)	(0.035)	(0.023)
Z-Teacher-perceived Effort (TpEff)				0.439***		0.505***
				(0.024)		(0.030)
SES*Z-CogEff					0.025	
					(0.050)	
SES*Z-TpEff						-0.150***
						(0.045)
Observations	1,270	1,270	1,270	1,270	1,270	1,270
Number of clusters	66	66	66	66	66	66

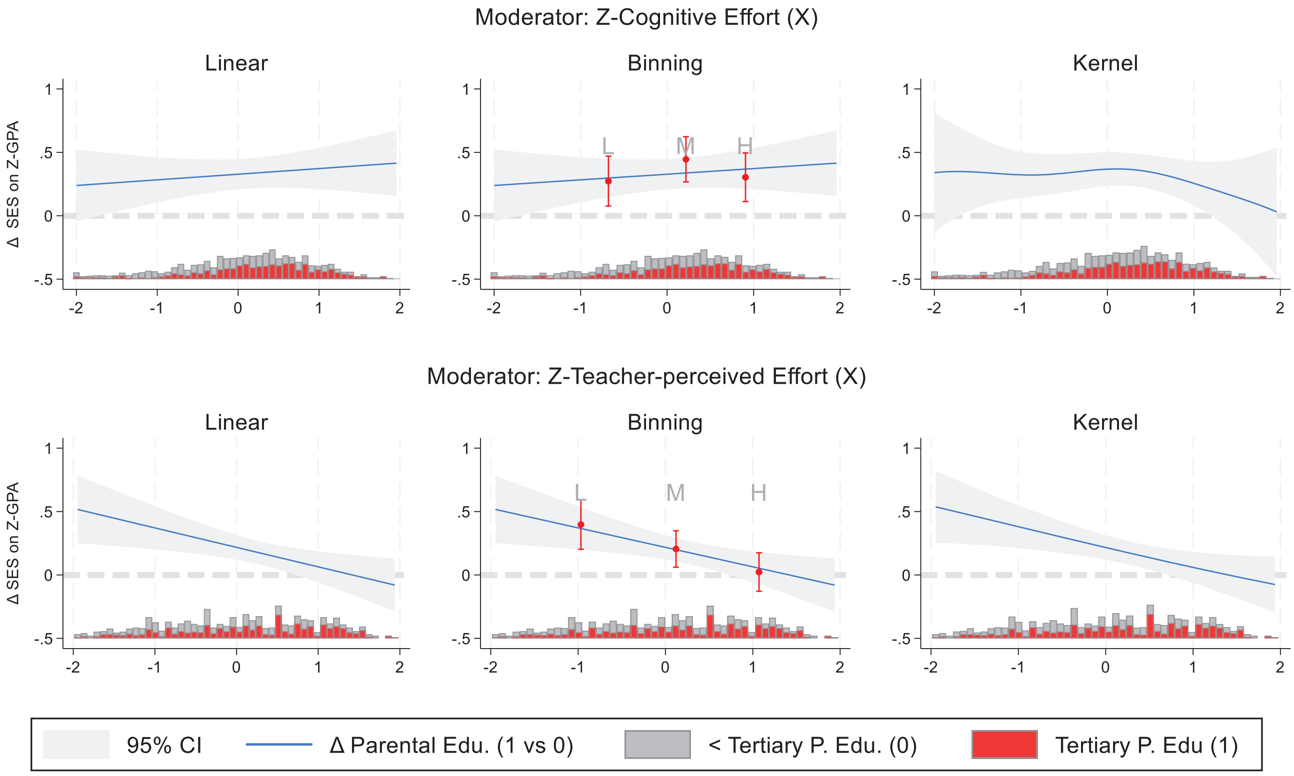
Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05
Controls: sex, age in months and migration background (level 1); type of school; school SES; and city (level 2).

Table S.7. Classroom fixed-effects models to control for teacher-level evaluation practices

	Classroom-FE					
	M1	**M2**	**M3**	**M4**	**M5**	**M6**
Parental Tertiary Edu. (SES)	0.573***	0.506***	0.464***	0.323***	0.463***	0.321***
	(0.071)	(0.068)	(0.067)	(0.065)	(0.067)	(0.064)
Z-IQ		0.296***	0.229***	0.133***	0.228***	0.136***
		(0.030)	(0.030)	(0.028)	(0.030)	(0.027)
Z-Cognitive Effort (CogEff)			0.259***	0.177***	0.231***	0.175***
			(0.030)	(0.027)	(0.040)	(0.027)
Z-Teacher-perceived Effort (TpEff)				0.446***		0.522***
				(0.025)		(0.036)
SES*Z-CogEff					0.061	
					(0.057)	
SES*Z-TpEff						-0.151**
						(0.050)
Classroom Fixed Effects	X	X	X	X	X	X
Observations	1,270	1,270	1,270	1,270	1,270	1,270
Number of clusters	66	66	66	66	66	66

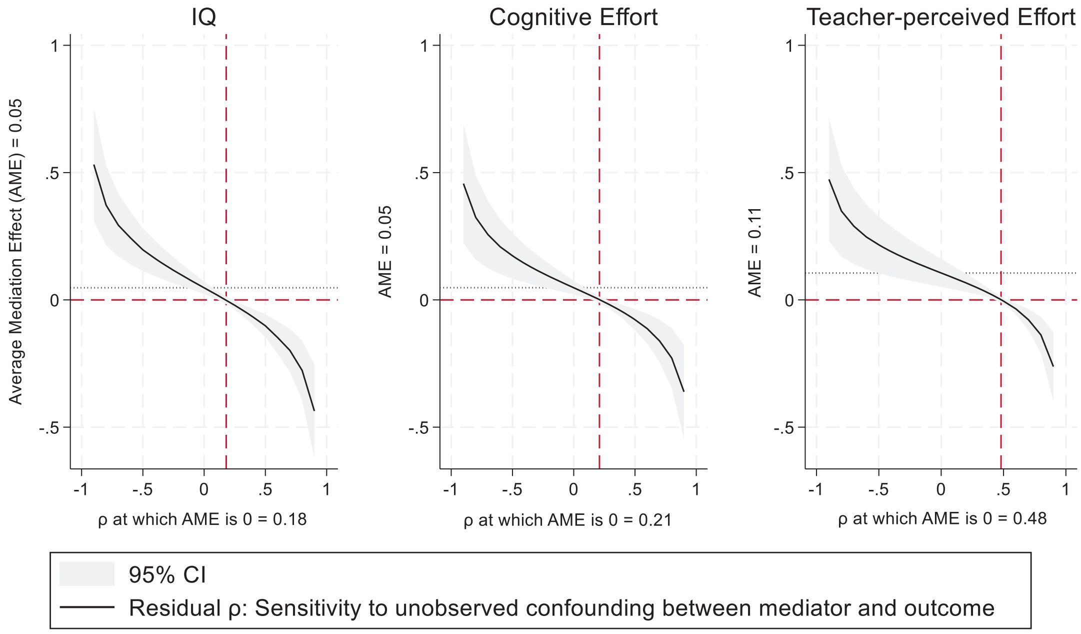
Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05
Controls: sex, age in months and migration background (level 1).

Figure S.1. SES-Effort interaction on GPA under different moderator specifications with 95% CI (n = 1,270): Linear (M5-M6, Table 2), Binning (low, medium, high), and Kernel



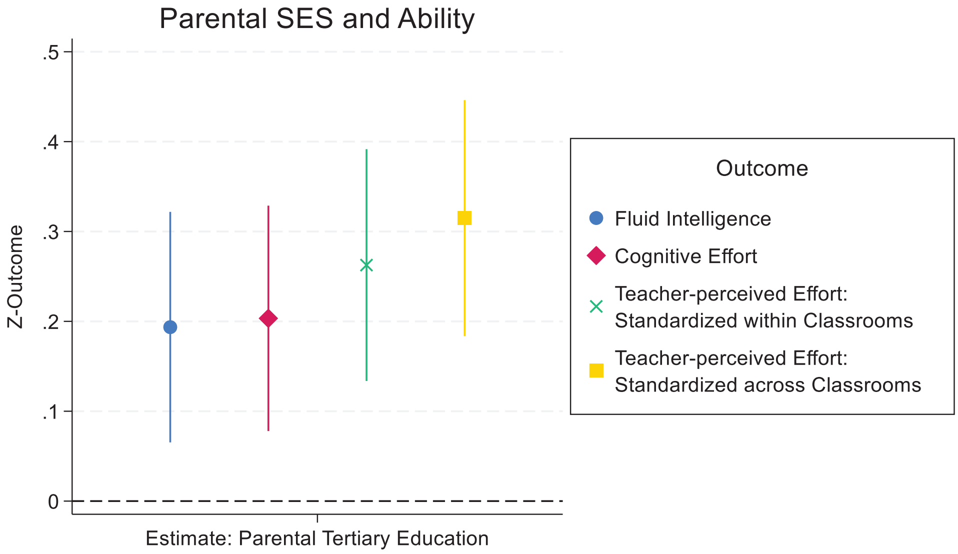
Notes: Controls: sex, age in months, migration background, and IQ (level 1); type of school; school SES; and city (level 2). Additional controls for teacher-perceived effort model: Cognitive effort. The p-value of the Wald statistic (CogEff = 0.323; TpEff = 0.881) is not statistically significant, providing no evidence against the null hypothesis that the linear interaction model and the three-bin model are statistically equivalent. The histograms at the bottom of each figure represent the effort distribution by parental education (low=grey; high=red).

Figure S.2. Robustness of the mediation effect to mediator-outcome unobserved confounding



Notes: n=1,270; Controls: sex, age in months, migration background, and IQ (except for IQ model) (level 1); type of school; school SES; and city (level 2). Additional controls for teacher-perceived effort model: Cognitive effort. 95% CI using 1,000 simulations with a precision step of 0.01.

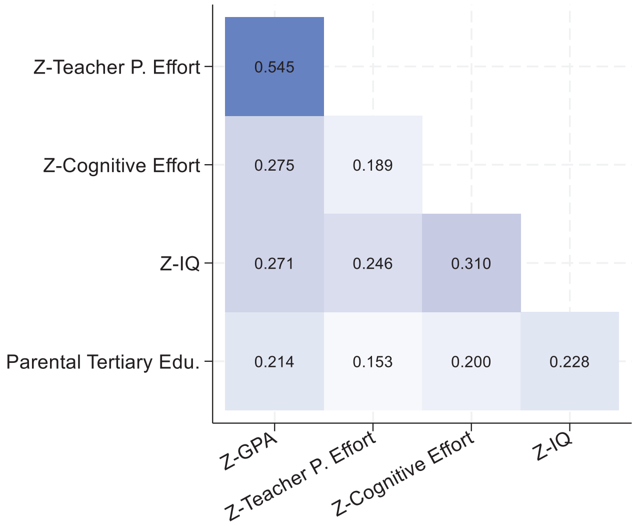
Figure S.3. Marginal effect of parental tertiary education on effort and ability with 95% CI



Notes: n = 1,270. Linear multilevel random-effects model; controls: sex, age in months, migration background, and IQ (except for the IQ model) (level 1); type of school; school SES; and city (level 2); additional controls in IQ model: cognitive effort; CogEff and IQ are standardized across classrooms.

B. Summary Statistics

Figure S.4. Pearson correlation matrix (n = 1,270)



Notes: all correlations are statistically significant at p<0.000.

C. Experimental Setup: Effort and Incentives

Table S.8. gives the rundown of the experimental sessions. At the beginning of the experiment, basic instructions were given to the students. During the experiment, the students carried out the tasks under different incentive conditions. The first one was the intrinsic condition, where the participants received no reward for doing the task. Afterwards, in the extrinsic condition, students received points (piece-rate) for each correct trial. They were informed that they could convert the points that they earned throughout the tasks into toys at the end of the day. Finally, the last condition was the tournament, where, besides still getting points for correct responses, the students were competing with their classmates for being the best in the class. As announced at the start of the tournament, the three best-performing students got a diploma as an extra reward, indicating their podium position. A "leisure task" was offered for students as an option for not doing the tasks and playing a computer game during that period. The purpose was to introduce an opportunity cost for doing the tasks, which makes the setup resemble real-life situations more closely, where distractions from learning or working are omnipresent.

Table S.8. Experimental setup

Task	Duration
Instructions + Leisure task	1 round each game of 1.5 min
Task 1	
Intrinsic condition	2 rounds of 2 min
Extrinsic condition	2 rounds of 2 min
Task 2	
Extrinsic condition	2 rounds of 2 min
Task 3	
Extrinsic condition	2 rounds of 2 min
Tournament condition	2 rounds of 2 min

Notes: own elaboration.

Table S.9. displays the proportion of students choosing the leisure task over the real-effort task. We can observe stark differences across conditions in the number of rounds in which the students choose to play games instead of doing the task. During the intrinsic condition, 20.9% of the sample carried out the tasks in all the rounds, 50.9% in less than half, and 28.2% played games in more than half of the rounds. However, during the piece-rate condition, over 92% of the students carried out the tasks in all the rounds, and only 7.8% played games at some point. Importantly, there are no significant differences in task engagement by parental SES. This means that students from different social classes were equally motivated by the piece-rate payoff, avoiding potential heterogeneity in the response to the extrinsic condition that might otherwise lead to biased results.

Table S.9. Task engagement by incentive condition

% of rounds in which students play games	Intrinsic condition			Extrinsic piece-rate condition		
	% of Low SES	% of High SES	% of Total	% of Low SES	% of High SES	% of Total
0	18.3	23.1	20.9	91.2	93.0	92.2
1-50	53.8	48.4	50.9	8.8	7.0	7.8
51-100	27.9	28.5	28.2	0	0	0
Total	100	100	100	100	100	100

Notes: own elaboration. Low SES = highest parental edu. < tertiary; High-SES = tertiary.