

Grades and Assessments – The Economists’ Perspective*

Ron Diris[†], Ingo E. Isphording[‡], Fanny Landaud[§] & Hans H. Sievertsen[¶]

Version: March 16, 2026

Forthcoming in ‘Bradley and Green, The Economics of Education,
3rd edition’

Abstract

This chapter reviews the economic literature on grading and assessment, focusing on how institutional design affects behavior and long-run outcomes. Grades and assessments serve as both informational devices and incentive mechanisms, shaping the decisions and beliefs of students, parents, teachers, and employers. We examine how key design choices—including who conducts assessments, the stakes attached to results, the coarseness of grading scales, and the format of performance feedback—generate trade-offs between informational accuracy, incentive provision, and equity. The evidence shows that seemingly minor institutional features, such as score thresholds, retake rules, and relative performance feedback, can substantially alter educational trajectories and labor market outcomes. Teacher-based assessments and standardized tests differ systematically in what they measure and whom they favor, with direct implications for sorting efficiency and inequality. We conclude by identifying promising directions for future research, including the welfare effects of alternative grading regimes, the rising importance of non-cognitive skills, and questions raised by AI-based assessment technologies.

*We thank Lex Borghans, Simon Burgess, Bart Golsteyn, Colin P. Green, Eric Maurin, Beatrice S. Rangvid, and Mikko Silliman for valuable comments.

[†]Leiden University, r.e.m.diris@law.leidenuniv.nl

[‡]Max Planck Institute for Behavioral Economics, isphording@coll.mpg.de

[§]CREST, CNRS, IPParis, fanny.landaud@ensae.fr

[¶]VIVE, hhs@vive.dk

1 Introduction

Grades and assessments — the institutional processes through which student performance is measured, translated into standardized evaluative scales, and communicated to stakeholders — are typically thought of as an indispensable part of any school-based education system. The resulting test scores and grade point averages (GPAs) have become the canonical measure of human capital in applied economics, despite growing evidence that they are far from a perfect measure — subject to teacher discretion, conflating skills and discipline, and affected by noise.

Furthermore, they are far more than passive measurements: through their design and use they translate complex learning processes into signals used by students, parents, teachers, schools, policymakers, and employers. By reducing informational frictions and shaping the incentives under which these actors operate, grades and assessments influence how human capital is formed. Moreover, grades and assessments have become central instruments of selection and ranking that assign individuals a position within educational and social hierarchies.

Yet, grading and assessing progress may take manifold forms, for example teacher-developed tests and judgmental (i.e. subjective) ratings, or externally graded standardized tests. The school questionnaire of PISA 2022 reveals substantial heterogeneity in type and frequency of assessment methods and grading systems among 15-year-old students (Figure 1), with the majority of students facing at least one teacher-based test per month and at least one standardized test per year.

Yet, such differences in the mode of assessment and grading systems—what they measure, how they measure it, and how results are communicated—may have far-reaching consequences for beliefs, effort provision, human capital accumulation, inequality, and life-course outcomes, making them institutional features of education systems worth economists' attention.

Economists' interest in grading and assessment can be further explained by two additional factors: first, the behavioral effects of grades speak directly to foundational microeconomic principles: they shape incentives, guide decisions under uncertainty, and influence belief formation in environments with imperfect and asymmetric information. Second, beyond this theoretical importance, institutional features such as score thresholds, retake rules, and external grading practices strongly serve economists' core methodological interest of clear causal inference.

In this chapter, we review the economic literature on grades and assessments, highlighting their diverse implications for individual decision-making and their consequences

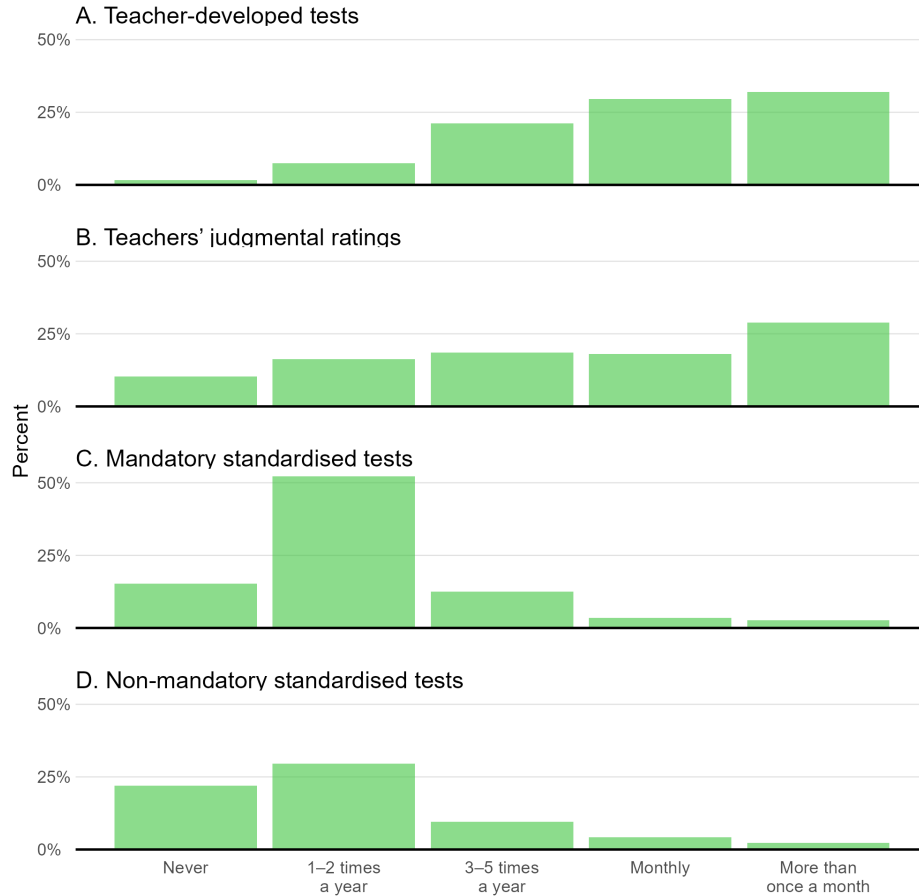


Figure 1: Distribution of assessment types across countries

Notes: The chart shows the distribution of country averages responses in the school questionnaire. Data source: PISA 2022.

for educational and labor market outcomes. We view grades and assessments as economic institutions that can be designed by policymakers and practitioners, with design choices involving important trade-offs and associated costs and benefits.

We structure the chapter as follows. We first discuss the economic rationale for assessing and grading students in Section 2. We then examine how assessments are designed and graded, and how their results are communicated in Section 3. Section 4 concludes with a discussion of the current research frontier.

2 Functions of grading and assessment: why we assess

2.1 Motives for Grading: Information and incentives

Education involves both the development of skills and the certification of those skills. Assessments and the grades they produce therefore perform two fundamental and interconnected functions. First, they serve as *informational devices*: they transform complex and imperfectly observed learning processes into observable signals, and reduce both informational imperfections and asymmetries between students, parents, teachers, schools, policymakers, and employers. Second, they serve as *incentives*: When assessments carry consequences for progression, tracking, scholarships, access to selective programs, or future employment, they shape the effort and behavior of students, teachers, and institutions.

These informational and incentive functions are inseparable. The same assessment that produces a grade which communicates a student's mastery also affects how that mastery is produced; the same test score that guides instructional decisions may influence how intensely students prepare; the same metric used for school accountability may redirect teachers' attention and pedagogy. The balance between these two functions differs across recipients and settings, and design choices that strengthen one dimension may weaken the other.

2.2 Recipients of grades and assessments

2.2.1 Students: learning about ability and incentives for effort

Students use grades and assessments both to learn about their own ability and to guide their effort. As informational signals, grades and test scores prompt students to update their beliefs about their academic ability, and prospects. Such feedback affects how students allocate study time and adjust learning strategies, ultimately shaping learning progress and affecting enrollment decisions (Zafar, 2011, Stinebrickner and Stinebrickner, 2012, Bobba and Frisancho, 2022). Grades may also be informative about perceived comparative advantages across skill dimensions, for example by shaping students' relative assessment of their strengths in STEM versus non-STEM subjects (Breda and Napp, 2019, Goulas, Griselda and Megalokonomou, 2024, Elsner, Isphording and Zölitz, 2021). Empirical evidence establishes causal links between receiving grades and decisions to apply for or drop out of college (Hyman, 2017, Stinebrickner and Stinebrickner, 2012), college selectivity (Bond et al., 2018) and field of study (Zafar, 2011, Avery et al., 2018, McEwan, Rogers and Weerapana, 2021, Li and Xia, 2024), also when these grades do not

directly affect eligibility. These effects are not necessarily unambiguous or symmetric: [Pistoletti \(2017\)](#) find no effects from receiving positive signals tied to score thresholds, but a substantial negative effect on study field enrollment from receiving a negative signal. Furthermore, [Foote, Schulkind and Shapiro \(2015\)](#) find no effect from marginally passing the “college-readiness” threshold of the ACT test on college enrollment, and [Main and Ost \(2014\)](#) find no effect of higher letter grades in introductory economics on undergraduate course-taking or on decisions to major in economics, indicating that the institutional context may matter.

Additionally, grades and assessment results can impact student effort, as educational programs or employers use them as sorting devices. In such cases, grades provide students with strong incentives to perform well: they make achievement salient, tie current effort to future opportunities, and thus reinforce goal-oriented behavior. [Hvidman and Sievertsen \(2021\)](#) and [Main and Ost \(2014\)](#) both find that receiving a lower grade for the same performance induces a compensating increase in effort and higher subsequent grades. [Hvidman and Sievertsen \(2021\)](#) further find that this translates into higher university enrollment. In contrast, [Tan \(2023\)](#) find that higher letter grades have a positive effect on grades in subsequent semesters, and lead to more difficult courses being chosen, and higher earnings.

2.2.2 Parents: reducing uncertainty

For parents, grades serve mainly as an information signal. They provide a standardized summary of their child’s learning progress and help identify strengths, weaknesses, and the need for support. Such information shapes parental beliefs and responses, influencing how much encouragement, praise, or pressure they apply ([Dizon-Ross, 2019](#)). Parents are not only interested in their own offspring’s assessments. Evaluations of class- or grade-mates offer a reference point for interpreting their child’s grades ([Barrera-Osorio et al., 2020](#)). Parents also value aggregated assessments at higher levels, for example school-wide indicators, especially when making school-choice decisions ([Burgess et al., 2015](#)).

2.2.3 Teachers: diagnostic inputs and accountability pressures

Teachers occupy a dual position in the assessment system: they generate a large share of grades and assessments, yet they also respond to these signals when making instructional, evaluative, and administrative decisions. Much of the economics literature focuses on teachers as producers of assessments, e.g., highlighting heterogeneity in grading standards, the role of non-blind evaluations, and the potential for implicit bias (e.g., [Betts and](#)

Grogger, 2003, Figlio and Lucas, 2004, Cornwell, Mustard and Van Parys, 2013, Terrier, 2020).

However, teachers also rely heavily on assessment outcomes as inputs into their own decision-making. Assessments provide teachers with a diagnostic signal about students' achievement, effort, and behavioral skills. Teachers may use these signals to adjust effort (Aucejo, Romano and Taylor, 2022, Virudachalam, Savin and Steinberg, 2024), allocate attention across students (Neal and Schanzenbach, 2010), or recommend placements into ability groups (Betts, 2011). Teachers may also rely on past assessments, which shape their expectations and beliefs about students (Burgess and Greaves, 2013).

Teachers are also directly affected by assessments through accountability systems (Deming and Figlio, 2016). While such systems have shown ambiguous effects on student achievement (e.g., Hanushek and Raymond, 2005), existing evidence documents behavioral responses such as teaching to the test (Jacob, 2005, Neal and Schanzenbach, 2010) or even manipulation (Jacob and Levitt, 2003).

2.2.4 Policymakers: performance measurement and system governance

Grades provide aggregate indicators for monitoring school performance, evaluating reforms, and guiding evidence-based policy. In this role, grades function as standardized data that make learning outcomes comparable across schools, cohorts, and regions, also providing the basis for evaluation and accountability of schools (Deming and Figlio, 2016). Their motivational effect arises when such metrics feed into accountability systems that shape the behavior of schools and teachers (Dee and Wyckoff, 2015).

2.2.5 Educational programs and selection committees: allocating access to opportunities

Assessments and resulting grades are widely used as a direct sorting device in educational programs. In primary and secondary education, grades guide decisions concerning educational tracks and ability-groups, grade retention and promotion, or eligibility for summer schools (Dustmann, Puhani and Schönberg, 2017, Dahl, Rooth and Stenberg, 2023, Jacob and Lefgren, 2004, 2009, Schwerdt, West and Winters, 2017, Manacorda, 2012). Grades may also affect the quality of attended school. Findings on the consequences of such quality effects in the context of exam schools in the US are mixed (Dobbie, Fryer et al., 2011, Abdulkadiroğlu, Angrist and Pathak, 2014, Angrist and Rokkanen, 2015), which also extends to non-US contexts (Jackson, 2013, Clark and Bono, 2016, Landaud, Ly and Maurin, 2020, Bütikofer et al., 2023). A potential explanation is that they may induce

opposing treatment effects: marginal grade increases that lead to attendance of higher-quality schools also lead to reduced parental effort, and lower self-ranking compared to one's peers (Pop-Eleches and Urquiola, 2013).

Grades can have especially large impacts on eligibility for post-secondary enrollment, as these are more regularly based on meeting pre-defined thresholds. The literature on these “threshold-crossing effects” of grades is mixed when it comes to the long-term consequences of marginally passing the threshold for college eligibility, with some studies finding small or no effects (Öckert, 2010, Clark and Martorell, 2014, Canaan and Mouganie, 2018) while others identify substantial returns (Zimmerman, 2014, Machin, McNally and Ruiz-Valenzuela, 2020). Such effects of grades on college eligibility may also operate through thresholds that determine whether students may *remain* enrolled. For example, Ost, Pan and Webber (2018) exploit cutoffs for college dismissal, identifying negative long-run consequences of dropout that is induced by marginal differences in grades. In addition to potentially impacting the years spent in education, threshold-crossing effects can also strongly impact the quality of education attended (Black, Denning and Rothstein, 2023, Canaan and Mouganie, 2018, Hoekstra, 2009, Hastings, Neilson and Zimmerman, 2013), and choices for field of study (Kirkeboen, Leuven and Mogstad, 2016, Heinesen, 2018, Hastings, Neilson and Zimmerman, 2013, Bleemer and Mehta, 2022). They also impact educational investments in post-secondary education through remedial programs (De Paola and Scoppa, 2014) and enrichment programs (Gong and Pan, 2023, Pugatch and Thompson, 2025).

External stakeholders such as scholarship committees rely on grades as well, e.g., in selection processes and to hold students accountable for study effort. Grades convey information about prior achievement and learning potential and help reduce uncertainty in admission decisions, and can provide a measure of ongoing effort provision. For example, Georgia's intensively analyzed HOPE Scholarship is granted to students who meet minimum GPA requirements, and requires students to maintain a minimum GPA to keep the scholarship (Cornwell, Mustard and Sridhar, 2006). The awareness that grades are used in such evaluations may strengthen their motivational effect for students (Hvidman and Sievertsen, 2021).

2.2.6 Employers: screening in environments of limited information

For employers, grades serve primarily as information signals. Especially in early-career hiring, where work experience is limited, grades act as proxies for skills, discipline, and trainability: they reduce uncertainty about productivity and fit, but tend to lose importance once employers learn about workers on the job (Hansen, Hvidman and Sievertsen,

2024). The credibility of these signals depends on stable and rigorous grading standards (Zubrickas, 2015, Ehlers and Schwager, 2016). Grade inflation—now well documented across secondary and higher education (Buckley, Letukas and Wildavsky, 2018, Denning et al., 2022, OFS, 2022)—counteracts this credibility by compressing the distribution of grades and reducing their informational content (Schwager, 2012).

3 The design of assessments: how we assess

Evaluating students' skills can be thought of consisting of three subsequent steps, each offering design choices that shape behavior. *Assessment* concerns the measurement of students' knowledge, skills, or performance. *Grading* translates these assessments into a communicable scale—such as pass/fail thresholds, letter grades, or continuous scores—introducing choices about granularity and scaling. *Communication* then transmits the information to recipients. Grades can be communicated in absolute form or benchmarked against past performance or peer groups; this can occur publicly or privately; and grades can be attached to specific awards or honor labels. Evidence shows that design choices at each stage can meaningfully alter how students, teachers, and families interpret and respond to grades, making the structure of assessment and grading systems a central component of their behavioral impact.

3.1 Assessment

3.1.1 What is assessed: (Non-)Cognitive Skills and Behavior

A first fundamental design decision about the process of measuring student achievement concerns *what* an assessment aims to capture. In principle, assessments can measure a wide range of student attributes, including domain-specific knowledge, general cognitive skills, non-cognitive traits such as motivation or conscientiousness, and observable behaviors such as participation or persistence. Teacher-based and standardized assessments, for example, tend to capture different mixtures of these dimensions. Empirically, Borghans et al. (2016) show that teacher-given grades and standardized test scores load differently on cognitive proficiency and non-cognitive characteristics, with teacher assessments placing more weight on behavioral and personality-related traits. This aligns with evidence that teacher grades combine cognitive performance with components of effort, engagement, and self-regulation, whereas standardized tests place greater weight on cognitive skills (see Almlund et al., 2011). Economists have further shown interest in “behav-

ioral grades”, which assess students’ social and work behavior in the classroom (Schoner, Mergele and Zierow, 2024), factors that have been shown to be highly predictive for future success (Jackson, 2018).

3.1.2 Who is assessing: Teachers and external examiners

A second influential decision is *who* is assessing. Two options seem to be most important: either the assessment can be undertaken by teachers themselves, or it can be outsourced to external parties, especially in the form of standardized tests. The merits of, and concerns about, teacher-based assessments and standardized tests have received considerable attention in the economic literature. The decision about who assesses has consequences for the *content* of the assessment, as discussed above. On the one hand, even highly standardized exams pick up non-cognitive skills such as endurance (Brown et al., 2025), personality (Segal, 2012), and test-taking motivation (Gneezy et al., 2019). On the other hand, standardized test scores almost entirely miss out behavioral dimensions—reflected, for example, in absences and suspensions and known to strongly predict long-run outcomes (Jackson, 2018).

Who assesses also has consequences for the *accuracy* of the measurement, with particular concerns that standardized tests may provide a relatively noisy measure of a student’s overall learning and effort compared to repeated teacher assessments spread over many tasks and periods. Because standardized tests are costly to organize, they are typically administered less frequently over a student’s educational career than teacher-based assessments. As a result, individual outcomes become more sensitive to transitory shocks on the exam day—such as air pollution, allergens, heat, time of day, welfare benefit timing, or even major football tournaments (e.g., Sievertsen, Gino and Piovesan, 2016, Ebenstein, Lavy and Roth, 2016, Amanzadeh, Vesal and Fatemi Ardestani, 2020, Bond et al., 2022, Gaggero and Tommasi, 2023, Park, 2022, Metcalfe, Burgess and Proud, 2019). These idiosyncratic shocks to performance on the test day can in turn have persistent consequences for educational attainment and earnings. Moreover, standardized exams cover only a narrow subset of skills, making performance sensitive to whether test content aligns with a student’s specific strengths rather than reflecting their broader knowledge (Landaud et al., 2024).

Lastly, debates exist about the relative *fairness* of standardized exams and teacher-based assessments. Preparation opportunities for standardized exams vary widely across schools and even classrooms, meaning that exam outcomes—though often interpreted as meritocratic—can partly reflect disparities in preparation rather than differences in ability or effort. For example, small institutional features—such as the option to retake ex-

ams or criteria for school-track assignment—can have sizeable and unequal effects on who obtains high scores and who accesses selective programs (e.g., [Vigdor and Clotfelter, 2003](#), [Hyman, 2017](#), [Goodman, Gurantz and Smith, 2020](#), [Landaud and Maurin, 2023](#), [Bizopoulou, Megalokonomou and Simion, 2024](#), [Kang et al., 2024](#), [Landaud and Maurin, 2025](#)). Regarding teacher-assessment, the main concern relates to the comparability of teacher-based grades, and to their potential biases in grading, which we discuss in more detail in the section [3.2.4](#).

3.1.3 High or Low Stakes of assessment

Assessments may differ substantially in their *stakes*. High-stakes assessments determine access to selective school or tracks, higher education programs, scholarships, or employment. Low-stakes assessments, by contrast, are typically used for feedback, diagnostics, formative purposes, international comparisons, or research.¹ These differences generate important trade-offs regarding incentives and informational content.

Because low-stakes tests carry few or no consequences, the effort students exert when taking them might vary systematically. This may reduce the informational value of observed scores for teachers, policymakers, or external evaluators. This issue is well illustrated by PISA: since the test is low-stakes, cross-country differences in PISA scores partly reflect differences in test effort rather than differences in underlying skills ([Gneezy et al., 2019](#), [Zamarro, Hitt and Mendez, 2019](#)).

By contrast, the use of high-stakes assessments may strengthen students' incentives to learn, stimulating sustained effort and generate genuine learning gains. . In line with this, several recent papers document positive effects of high-stakes testing on achievement and human capital accumulation across diverse settings (e.g., [Hvidman and Sievertsen, 2021](#), [Leschnig, Schwerdt and Zigova, 2022](#), [Fidjeland, 2023](#), [Bergbauer, Hanushek and Woessmann, 2024](#), [Molin, 2024](#)).

However, the same incentives that induce students to work harder may also generate important distortions. One aspect of this is learning to the test and strategic course selection (e.g., selecting leniently graded courses) (e.g., [Bar, Kadiyali and Zussman, 2009](#), [Hernandez-Julian and Looney, 2016](#)). In these cases, the informational value of assess-

¹Systems that eliminate grades altogether can be viewed as an extreme form of low-stakes assessment. When no formal grade is provided, the assessment conveys information to the student but carries no external consequences. As such, the trade-offs associated with “no grades” versus grades are similar to those of low-versus high-stakes assessments. In practice, however, many such policy experiments substitute numerical grades with written assessments and descriptive feedback, so that these reforms often change the mode of grading and communication rather than eliminate information altogether. We point to the need for research on these regimes in our outlook in Section 4.

ment results is diminished, as scores also reflect students' strategic choices and not only their educational skills. Another distortion effect of higher stakes is its potential psychological cost. Whereas low-stakes assessments tend to be more health-preserving, high-stakes assessments can increase stress and anxiety and negatively affect students' mental health (e.g., [Bach and Fischer, 2020](#), [Heissel et al., 2021](#), [Högberg and Horn, 2022](#), [Linder, Gerdtham and Heckley, 2025](#)). Apart from being harmful by themselves, stress effects of high-stakes tests may also reduce their informational value.

Distortions generated by high-stakes setting are not limited to students. High-stakes systems can also affect teachers, as their own career prospects or school resources may depend on test performance. When this is the case, teachers may prioritize maximizing test scores over promoting broader learning through practices such as teaching to the test or reallocating attention toward students near proficiency thresholds (e.g., [Jacob, 2005](#), [Wikström and Wikström, 2005](#), [Reback, 2008](#), [Carrell and West, 2010](#), [Dee and Jacob, 2010](#), [Neal and Schanzenbach, 2010](#)). In more extreme cases, high stakes can even lead to outright manipulation or cheating (e.g., [Jacob and Levitt, 2003](#), [Dee et al., 2019](#)).

Beyond the incentives-information trade-offs, an additional consideration when choosing between high- and low-stakes assessments is fairness. Higher stakes can amplify inequalities when student groups respond differently to pressure or test anxiety. A growing literature documents gender differences in performance under pressure, with girls often performing worse than their prior achievement would predict in high-stakes settings (e.g., [Azmat, Calsamiglia and Iriberry, 2016](#), [Cai et al., 2019](#), [Iriberry and Rey-Biel, 2019](#), [Angelo and Reis, 2021](#)).

3.1.4 The Format of Assessments

The economic literature has further documented far-reaching and often unexpected consequences of specific design choices in assessment. In particular, the format of assessment questions—such as multiple-choice versus open-ended items, or the use of penalties for incorrect answers—can influence test results and have fairness implications.

Multiple-choice formats tend to systematically disadvantage certain groups—most notably girls in mathematics and boys in reading ([Griselda, 2024](#)). Additionally, gender differences in willingness to guess and risk preferences contribute to gender performance gaps under scoring rules that penalize incorrect responses (e.g., [Baldiga, 2014](#), [Pekkarinen, 2015](#), [Coffman and Klinowski, 2020](#), [Iriberry and Rey-Biel, 2021](#)). Time pressure similarly affects groups of students differently, with evidence showing that it disproportionately lowers girls' performance ([De Paola and Gioia, 2016](#)). A recent paper by [Banerjee et al. \(2025\)](#) shows that item wording can matter: students from a low-income background perform

better when math questions are applied to real-world contexts (e.g. market exchanges) compared to the more abstract math problems that are typically prevalent in achievement tests.

A further design element concerns the *order* in which questions are presented. Recent work by [Anaya et al. \(2022\)](#) shows that question difficulty order can meaningfully influence performance: students perform worse when relatively difficult questions appear early in an exam. Finally, test length matters. Longer tests can reduce measurement error as the test result is based on more information, but may also conflate subject knowledge with the ability to sustain effort. Evidence shows that performance decline along tests is substantial, and has important implications for estimated performance differences by gender or by country ([Borghans and Schils, 2018](#), [Balart and Oosterveen, 2019](#)).

3.2 Grading

3.2.1 Coarse vs. fine grading scales.

We now turn to design decisions regarding how assessments are being summarized — graded — to be then communicated to stake holders. According grading systems may thereby differ in their *coarseness*. Coarse grading systems—such as pass/fail evaluations or broad letter grades—may obscure underlying performance differences ([Tan, 2023](#)) and generate threshold effects. Students who narrowly meet a cutoff receive the same grade as peers with substantially higher performance, whereas those just below may face disproportionate consequences. Empirical evidence shows that such threshold effects can meaningfully shape educational trajectories: receiving a higher letter grade or a more favorable performance label at a cutoff increases the likelihood of majoring in a subject or pursuing further study in the field (e.g., [Owen, 2010](#), [Papay, Murnane and Willett, 2016](#), [Avery et al., 2018](#)).

By contrast, finer-grained grading scales convey more precise information and can better distinguish variation in student achievement. However, increased granularity can also heighten performance pressure and potentially discourage students who receive the lowest marks. Recent evidence from a natural experiment in Sweden shows that introducing a more detailed grading scale reduced the likelihood of graduation from high school and decreased enrollment in STEM courses at university, with effects consistent with a discouragement mechanism ([Collins and Lundstedt, 2024](#)). Thus, while finer grading scales improve informational content, they can come at a psychological cost.

3.2.2 Grading standards

Putting assessments on a common scale asks for *grading standards*. In the absence of clear guidelines and standardization, individuals may differ substantially in what they consider to be excellent or insufficient performance. Heterogeneity in grading standards is well documented empirically: several studies find systematic differences in grading standards across teachers and institutions (e.g., [Betts and Grogger, 2003](#), [Figlio and Lucas, 2004](#), [Bar, Kadiyali and Zussman, 2009](#), [Hernandez-Julian and Looney, 2016](#)). Furthermore, it is well documented that teachers tend to grade “to the curve”: they adjust evaluations relative to the performance distribution within their class or cohort, and can adhere to implicit quotas. As a result, teacher-based grades are not directly comparable across contexts, which raises fairness concerns when they are used to allocate scarce opportunities, such as places in selective secondary or higher education programs.

3.2.3 Grade inflation

Closely tied to heterogeneity in grading standards, the concept of *grade inflation* has received considerable attention by economists. Depending on the institutional setting, teachers and schools may face incentives to adopt increasingly lenient grading practices in order to attract students ([Bar, Kadiyali and Zussman, 2009](#)), meet accountability targets ([Dee et al., 2019](#)), or motivate lower-performing students ([Chan, Hao and Suen, 2007](#), [Zubrickas, 2015](#)). These incentives, however, are counterbalanced by reputational concerns ([Ehlers and Schwager, 2016](#)). At the micro level, studies document teachers’ tendencies toward lenient grading ([Diamond and Persson, 2016](#)), and the presence of grade inflation—where grades rise without corresponding improvements in actual performance—is well established ([Buckley, Letukas and Wildavsky, 2018](#), [Nordin, Heckley and Gerdtham, 2019](#), [Denning et al., 2022](#)). As most grading scales have a ceiling, grade inflation ultimately leads to “grade compression” ([Babcock, 2010](#)).

3.2.4 Teacher biases

Teacher assessments are typically graded non-blind, which can introduce potential biases. Even when performance signals are similar, teachers’ prior beliefs about students’ abilities or the quality of the student-teacher relationship may influence the grades awarded. Such teacher biases have been analyzed extensively in settings where both teacher-based grades and externally graded standardized exams are available. Comparisons between these independent sources of information reveal systematic grading gaps by gender, socioeconomic background, or ethnicity (e.g., [Burgess and Greaves, 2013](#), [Falch and Naper, 2013](#), [Breda](#)

and Ly, 2015, Lindahl, 2016, Lavy and Sand, 2018, Terrier, 2020, Ferman and Fontes, 2022).

However, as emphasized by Delaney and Devereux (2021), part of the observed grading gaps between teacher-based and standardized evaluations may reflect differences in how students perform under the specific conditions of standardized tests rather than pure grading bias, and recent work highlights methodological challenges in comparing blind and non-blind assessments (Delaney and Devereux, 2025).

Interestingly, experimental and quasi-experimental evidence generally finds that while teacher assessments can be influenced by subjective beliefs, teachers are rarely systematically biased against a specific group (e.g., Hinnerich, Höglin and Johannesson, 2011, Hanna and Linden, 2012, Burgess and Greaves, 2013, Rangvid, 2019, Avitzour et al., 2020).

3.3 Communication and Feedback

3.3.1 Absolute versus relative feedback

Finally, we turn to design decisions about how graded assessments are *communicated* to recipients. A large body of evidence shows that information about past performance affects subsequent achievement and schooling decisions (e.g., Azmat and Iriberry, 2010, Bandiera, Larcinese and Rasul, 2015, Andersen and Nielsen, 2020, Beuchert, Eriksen and Krægpøth, 2020, Hakimov, Schmacker and Terrier, 2022, Fischer and Wagner, 2023). The effectiveness of this feedback depends on *how* it is framed and communicated. A first key factor is whether feedback includes information about *relative performance*. An extensive literature shows that relative rankings matter for the formation of student beliefs (e.g., Elsner and Ispording, 2017, Murphy and Weinhardt, 2020). It is therefore not surprising that the degree of emphasis in feedback on relative performance mediates its effects. In a natural experiment in Spanish high schools, Azmat and Iriberry (2010) show that adding relative performance information to report cards increases students' grades by about 5%. In contrast, Azmat et al. (2019) study Spanish university students and find that providing information on students' rank in the grade distribution decreases the number of exams passed and GPA in the short run, while increasing self-reported satisfaction. Goulas and Megalokonomou (2021) show that relative performance feedback increases effort, future achievement, ambition and earnings for high-ability students, but has detrimental effects for low-ability students. Hermes et al. (2021) show that communicating feedback in terms of learning gains relative to past performance significantly raises motivation, effort and future achievement for low-ability students. Overall, the effects of relative feedback seem

to largely depend on students' initial beliefs about their relative ability, and on the time horizon considered.

3.3.2 Public versus private feedback

Public provision of feedback can lower effort through peer pressure and social norms. [Bursztyn and Jensen \(2015\)](#) show that using public school achievement leaderboards substantially reduces students' future performance. The effect is driven by high-ability students, who reduce effort to avoid being at the top of the public leaderboard, a fact that suggests certain peer cultures that stigmatize student effort [Bursztyn, Egorov and Jensen \(2019\)](#).

A related literature examines effects of secondary signals of academic distinctions that are based on specific grade thresholds, e.g. honors and awards. Findings on the consequences of such distinctions are mixed. [Freier, Schumann and Siedler \(2015\)](#) finds a substantial effect of graduating with honors for law students (14% five to six years after graduation), while [Khoo and Ost \(2018\)](#) identify positive earnings effects for the first two years after graduation but fade out by year three, consistent with a model of employer learning. [Atay, Asik and Tumen \(2024\)](#) identify a modest effect of around 4% for economics students in Turkey, but only for males and those at non-elite institutions. [Feng and Graetz \(2017\)](#), [Baert and Verhaest \(2021\)](#) identify substantial positive effects of university degree classes on labor market outcomes. Relatedly, [Busso, Montañó and Muñoz-Morales \(2025\)](#) show that distinctions based on specific-skill thresholds can also act as salient labor-market signals, affecting job placement and early career outcomes.

4 Conclusion: the current research frontier

Economic research increasingly views grades and assessments not simply as measurements of ability, but as informational and incentive devices that influence students, teachers, parents, institutions, and employers. As shown throughout this chapter, even small features of assessment and grading design can meaningfully shape beliefs, effort, stress, sorting, and long-run outcomes. In this conclusion, we highlight emerging trends in recent research and key gaps that remain open for future work.

4.1 Emerging trends

1. Institutional rules as policy tools A substantial share of recent contributions exploits discrete institutional features—retake policies, grading scale reforms, and threshold rules—to study causal effects. Eligibility thresholds for programs or diplomas have been shown to affect attainment and labor-market outcomes (Zimmerman, 2014, Machin, McNally and Ruiz-Valenzuela, 2020). Research on retake opportunities highlights how “second chances” alter persistence and high-stakes performance (Goodman, Gurantz and Smith, 2020, Kang et al., 2024). Grading-scale reforms show that increasing granularity can reduce graduation rates and STEM enrollment by inducing discouragement (Collins and Lundstedt, 2024). These findings demonstrate that seemingly small institutional design choices can have far-reaching consequences.

2. What assessments and test scores miss An important trend highlights how assessments capture not only knowledge but also personality traits, motivation, and effort. Teacher-based grades load more strongly on behavioral traits than standardized exams (Borghans et al., 2016), while test-taking motivation and effort have been shown to vary widely across individuals and countries (Gneezy et al., 2019, Zamarro, Hitt and Mendez, 2019). Research on performance decline within tests (Balart and Oosterveen, 2019) and on cognitive endurance (Brown et al., 2025) further illustrates that assessments depend on psychological and behavioral inputs, challenging interpretations of test scores as pure measures of cognitive ability. Finally, recent research highlights what test scores miss, and shows that the information they fail to capture is highly predictive of subsequent outcomes (Jackson, 2018, Brown et al., 2025). This might be of increasing importance in light of growing labor market returns to higher-order skills, such as social skills (Castex and Kogan Dechter, 2014, Hermo et al., 2022, Woessmann, 2025, Deming and Silliman, 2025). These shifts in skill demand imply that traditional grades and standardized test scores — which primarily capture domain-specific knowledge and basic cognitive proficiency — are becoming less sufficient as measures of the human capital that matters for economic success. From an economist’s perspective, this suggests to broaden measurements: developing and validating assessments that reliably capture higher-order and non-cognitive skills for more accurate human capital measurement.

3. Assessment, grading and mental health A small, but growing number of studies address the link between assessments, grades and mental health (Bach and Fischer, 2020, Heissel et al., 2021, Högberg and Horn, 2022, Bütikofer et al., 2023, Linder, Gerdtham and Heckley, 2025). While identifying causal effect remains empirically challenging, there

is some evidence pointing to direct negative effects, and that stress induced by high-stakes tests may further weaken the informational content of the scores they generate. An important avenue for future research is isolating the separate effects of assessment format and of the grades received. Moreover, in studying the causal effect of grades and assessments on mental health, it is important to be clear about the counterfactual and the long-run consequences of these.

4. Efficiency and equity Assessment and grading policies have direct effects for labor market sorting (Schwager, 2012, Ehlers and Schwager, 2016, Hansen, Hvidman and Sievertsen, 2024). Designing grading and assessment systems that provide accurate signals therefore matters for labor market efficiency (Zubrickas, 2015). Moreover, assessment and grading policies have direct implications for inequality and access to degrees (Hyman, 2017).

Beyond these emerging trends, our review has left us with the feeling that the applied economics literature, to which the authors of this review have also contributed, focuses on narrowly defined institutional features and single design choices, prioritizing settings and questions that allow for clean identification of causal effects. While this approach has generated valuable insights into the behavioral consequences of specific components of grading and assessment systems, it risks losing sight of broader questions.

In particular, we know little about interactions between system components, their long-run effects, or general equilibrium effects, including how assessment policies shape skill distributions over time. We also lack evidence on how grading and assessment interact with other institutional settings such as school choice and college admissions, which themselves rely heavily on school-based grades, and on how these systems affect stakeholders beyond students and teachers, notably parents' beliefs and investment decisions. An “optimal” grading system — maximizing informational accuracy and incentives for learning while minimizing fairness concerns and distortions in the learning process — may be impossible to achieve, as relevant trade-offs are numerous.

Moreover, by working largely from the status quo, the literature may overlook fundamentally different assessment regimes without grading, which may also have economic rationales given the opportunity costs of grading, including elevated stress among students (Heissel et al., 2019) and time burdens for teachers. Relatedly, rapid developments in digital technologies, automation, and AI-based assessment tools raise new questions about transparency, bias, and the welfare consequences of algorithmic grading that remain largely unexplored.

Taken together, these considerations point to several promising directions for future research that move beyond isolated design features and instead study assessment systems in a more integrated, institutionally grounded, and welfare-oriented way, including their potential implications for inequality and long-run economic outcomes.

References

- Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak.** 2014. “The elite illusion: Achievement effects at Boston and New York exam schools.” *Econometrica*, 82(1): 137–196.
- Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz.** 2011. “Personality psychology and economics.” In *Handbook of the Economics of Education*. Vol. 4, 1–181. Elsevier.
- Amanzadeh, Naser, Mohammad Vesal, and Seyed Farshad Fatemi Ardestani.** 2020. “The impact of short-term exposure to ambient air pollution on test scores in Iran.” *Population and Environment*, 41(3): 253–285.
- Anaya, Lina, Nagore Iriberry, Pedro Rey-Biel, and Gema Zamarro.** 2022. “Understanding performance in test taking: The role of question difficulty order.” *Economics of Education Review*, 90(C).
- Andersen, Søren C., and Helena Skyt Nielsen.** 2020. “Learning from performance information.” *Journal of Public Administration Research and Theory*, 30(3): 415–431.
- Angelo, Catarina, and Ana Balcão Reis.** 2021. “Gender gaps in different grading systems.” *Education Economics*, 29(1): 105–119.
- Angrist, Joshua D, and Miikka Rokkanen.** 2015. “Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff.” *Journal of the American Statistical Association*, 110(512): 1331–1344.
- Atay, Salim, Gunes A Asik, and Semih Tumen.** 2024. “Impact of Graduating with Honours on Entry Wages of Economics Majors.” *Oxford Bulletin of Economics and Statistics*, 86(3): 606–640.
- Aucejo, Esteban, Teresa Romano, and Eric S Taylor.** 2022. “Does evaluation change teacher effort and performance? Quasi-experimental evidence from a policy of retesting students.” *Review of Economics and Statistics*, 104(3): 417–430.
- Avery, Christopher, Oded Gurantz, Michael Hurwitz, and Jonathan Smith.** 2018. “Shifting College Majors in Response to Advanced Placement Exam Scores.” *Journal of Human Resources*, 53(4): 918–956.

- Avitzour, Eliana, Adi Choen, Daphna Joel, and Victor Lavy.** 2020. “On the Origins of Gender-Biased Behavior: The Role of Explicit and Implicit Stereotypes.” National Bureau of Economic Research 27818.
- Azmat, Ghazala, and Nagore Iriberry.** 2010. “The importance of relative performance feedback information: Evidence from a natural experiment using high school students.” *Journal of Public Economics*, 94(7-8): 435–452.
- Azmat, Ghazala, Caterina Calsamiglia, and Nagore Iriberry.** 2016. “Gender Differences in Response to Big Stakes.” *Journal of the European Economic Association*, 14(6): 1372–1400.
- Azmat, Ghazala, Manuel Bagues, Antonio Cabrales, and Nagore Iriberry.** 2019. “What You Don’t Know. . . Can’t Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education.” *Management Science*, 65(8): 3714–3736.
- Babcock, Philip.** 2010. “Real costs of nominal grade inflation? New evidence from student course evaluations.” *Economic inquiry*, 48(4): 983–996.
- Bach, Maximilian, and Mira Fischer.** 2020. “Understanding the Response to High-Stakes Incentives in Primary Education.” Institute of Labor Economics (IZA) 13845.
- Baert, Stijn, and Dieter Verhaest.** 2021. “Work hard or play hard? Degree class, student leadership and employment opportunities.” *Oxford Bulletin of Economics and Statistics*, 83(4): 1024–1047.
- Balart, Pau, and Matthijs Oosterveen.** 2019. “Females show more sustained performance during test-taking than males.” *Nature communications*, 10(1): 1–11.
- Baldiga, Katherine.** 2014. “Gender Differences in Willingness to Guess.” *Management Science*, 60(2): 434–448.
- Bandiera, Oriana, Valentino Larcinese, and Imran Rasul.** 2015. “Blissful ignorance? A natural experiment on the effect of feedback on students’ performance.” *Labour Economics*, 34: 13–25.
- Banerjee, Abhijit V, Swati Bhattacharjee, Raghavendra Chattopadhyay, Esther Duflo, Alejandro J Ganimian, Kailash Rajah, and Elizabeth S Spelke.** 2025. “Children’s arithmetic skills do not transfer between applied and academic mathematics.” *Nature*, 1–9.

- Barrera-Osorio, Felipe, Kathryn Gonzalez, Francisco Lagos, and David J Deming.** 2020. "Providing performance information in education: An experimental evaluation in Colombia." *Journal of Public Economics*, 186: 104185.
- Bar, Talia, Vrinda Kadiyali, and Asaf Zussman.** 2009. "Grade Information and Grade Inflation: The Cornell Experiment." *Journal of Economic Perspectives*, 23(3): 93–108.
- Bergbauer, Annika B., Eric A. Hanushek, and Ludger Woessmann.** 2024. "Testing." *Journal of Human Resources*, 59(2): 349–388.
- Betts, Julian R.** 2011. "The economics of tracking in education." In *Handbook of the Economics of Education*. Vol. 3, 341–381. Elsevier.
- Betts, Julian R, and Jeff Grogger.** 2003. "The impact of grading standards on student achievement, educational attainment, and entry-level earnings." *Economics of Education Review*, 22(4): 343–352.
- Beuchert, Louise, Tine Louise Mundbjerg Eriksen, and Morten Visby Krægpøth.** 2020. "The impact of standardized test feedback in math: Exploiting a natural experiment in 3rd grade." *Economics of Education Review*, 77.
- Bizopoulou, Aspasia, Rigissa Megalokonomou, and Ștefania Simion.** 2024. "Do second chances pay off? Evidence from a natural experiment with low-achieving students." *Journal of Public Economics*, 239.
- Black, Sandra E., Jeffrey T. Denning, and Jesse Rothstein.** 2023. "Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes." *American Economic Journal: Applied Economics*, 15(1): 26–67.
- Bleemer, Zachary, and Aashish Mehta.** 2022. "Will studying economics make you rich? A regression discontinuity analysis of the returns to college major." *American Economic Journal: Applied Economics*, 14(2): 1–22.
- Bobba, Matteo, and Veronica Frisancho.** 2022. "Self-perceptions about academic achievement: Evidence from Mexico City." *Journal of Econometrics*, 231(1): 58–73.
- Bond, Timothy N, George Bulman, Xiaoxiao Li, and Jonathan Smith.** 2018. "Updating human capital decisions: Evidence from SAT score shocks and college applications." *Journal of Labor Economics*, 36(3): 807–839.

- Bond, Timothy N, Jillian B Carr, Analisa Packham, and Jonathan Smith.** 2022. “Hungry for Success? SNAP Timing, High-Stakes Exam Performance, and College Attendance.” *American Economic Journal: Economy Policy*, 14(4): 51–79.
- Borghans, Lex, and Trudie Schils.** 2018. “Decomposing achievement test scores into measures of cognitive and noncognitive skills.” *Available at SSRN 3414156*.
- Borghans, Lex, Bart HH Golsteyn, James J Heckman, and John Eric Humphries.** 2016. “What grades and achievement tests measure.” *Proceedings of the National Academy of Sciences*, 113(47): 13354–13359.
- Breda, Thomas, and Clotilde Napp.** 2019. “Girls’ comparative advantage in reading can largely explain the gender gap in math-related fields.” *Proceedings of the National Academy of Sciences*, 116(31): 15435–15440.
- Breda, Thomas, and Son Thierry Ly.** 2015. “Professors in Core Science Fields Are Not Always Biased against Women: Evidence from France.” *American Economic Journal: Applied Economics*, 7(4): 53–75.
- Brown, Christina, Supreet Kaur, Geeta Kingdon, and Heather Schofield.** 2025. “Cognitive endurance as human capital.” *The Quarterly Journal of Economics*, 140(2): 943–1002.
- Buckley, Jack, Lynn Letukas, and Ben Wildavsky.** 2018. “Measuring Success: Testing, Grades, and the Future of College Admissions.” *Johns Hopkins University Press*.
- Burgess, Simon, and Ellen Greaves.** 2013. “Test scores, subjective assessment, and stereotyping of ethnic minorities.” *Journal of Labor Economics*, 31(3): 535–576.
- Burgess, Simon, Ellen Greaves, Anna Vignoles, and Deborah Wilson.** 2015. “What parents want: School preferences and school choice.” *The Economic Journal*, 125(587): 1262–1289.
- Bursztyn, Leonardo, and Robert Jensen.** 2015. “How does peer pressure affect educational investments?” *The quarterly journal of economics*, 130(3): 1329–1367.
- Bursztyn, Leonardo, Georgy Egorov, and Robert Jensen.** 2019. “Cool to be Smart or Smart to be Cool? Understanding Peer Pressure in Education.” *The Review of Economic Studies*, 86(4): 1487–1526. RePEc:oup:restud:v:86:y:2019:i:4:p:1487-1526.

- Busso, Matias, Sebastián Montaña, and Juan Muñoz-Morales.** 2025. “Signaling Specific Skills and the Labor Market of College Graduates.” *The Review of Economics and Statistics*, 1–46.
- Bütikofer, Aline, Rita Ginja, Katrine V Løken, and Fanny Landaud.** 2023. “Higher-achievement schools, peers and mental health.” *The Economic Journal*, 133(655): 2580–2613.
- Cai, Xiqian, Yi Lu, Jessica Pan, and Songfa Zhong.** 2019. “Gender Gap under Pressure: Evidence from China’s National College Entrance Examination.” *The Review of Economics and Statistics*, 101(2): 249–263.
- Canaan, Serena, and Pierre Mouganie.** 2018. “Returns to education quality for low-skilled students: Evidence from a discontinuity.” *Journal of Labor Economics*, 36(2): 395–436.
- Carrell, Scott E., and James E. West.** 2010. “Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors.” *Journal of Political Economy*, 118(3): 409–432.
- Castex, Gonzalo, and Evgenia Kogan Dechter.** 2014. “The Changing Roles of Education and Ability in Wage Determination.” *Journal of Labor Economics*, 32(4): 685–710.
- Chan, William, Li Hao, and Wing Suen.** 2007. “A signaling theory of grade inflation.” *International Economic Review*, 48(3): 1065–1090.
- Clark, Damon, and Emilia Del Bono.** 2016. “The Long-Run Effects of Attending an Elite School: Evidence from the United Kingdom.” *American Economic Journal: Applied Economics*, 8(1): 150–176.
- Clark, Damon, and Paco Martorell.** 2014. “The signaling value of a high school diploma.” *Journal of Political Economy*, 122(2): 282–318.
- Coffman, Katherine B., and David Klinowski.** 2020. “The impact of penalties for wrong answers on the gender gap in test scores.” *Proceedings of the National Academy of Sciences*, 117(16): 8794–8803.
- Collins, Matthew, and Jonas Lundstedt.** 2024. “The effects of more informative grading on student outcomes.” *Journal of Economic Behavior & Organization*, 218: 514–549.

- Cornwell, Christopher, David B Mustard, and Deepa J Sridhar.** 2006. “The enrollment effects of merit-based financial aid: Evidence from Georgia’s HOPE program.” *Journal of Labor Economics*, 24(4): 761–786.
- Cornwell, Christopher, David B. Mustard, and Jessica Van Parys.** 2013. “Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School.” *Journal of Human Resources*, 48(1): 236–264.
- Dahl, Gordon B, Dan-Olof Rooth, and Anders Stenberg.** 2023. “High school majors and future earnings.” *American Economic Journal: Applied Economics*, 15(1): 351–382.
- Dee, Thomas S., and Brian A. Jacob.** 2010. “The Impact of No Child Left Behind on Students, Teachers, and Schools.” *Brookings Papers on Economic Activity*, 41(2): 149–207.
- Dee, Thomas S, and James Wyckoff.** 2015. “Incentives, selection, and teacher performance: Evidence from IMPACT.” *Journal of Policy Analysis and Management*, 34(2): 267–297.
- Dee, Thomas S, Will Dobbie, Brian A Jacob, and Jonah Rockoff.** 2019. “The causes and consequences of test score manipulation: Evidence from the New York regents examinations.” *American Economic Journal: Applied Economics*, 11(3): 382–423.
- Delaney, Judith M., and Paul J. Devereux.** 2021. “The Economics of Gender and Educational Achievement: Stylized Facts and Causal Evidence.” *Oxford Research Encyclopedia of Economics and Finance*.
- Delaney, Judith M, and Paul J Devereux.** 2025. “Teacher Bias and Evaluation Differences in Test Scores: Different Methods for Different Questions.” *Oxford Bulletin of Economics and Statistics*.
- Deming, David J, and David Figlio.** 2016. “Accountability in US education: Applying lessons from K–12 experience to higher education.” *Journal of Economic Perspectives*, 30(3): 33–56.
- Deming, David J., and Mikko I. Silliman.** 2025. “Skills and Human Capital in the Labor Market.” In *Handbook of Labor Economics*. Vol. 6, 115–152.
- Denning, Jeffrey T, Eric R Eide, Kevin J Mumford, Richard W Patterson, and Merrill Warnick.** 2022. “Why have college completion rates increased?” *American Economic Journal: Applied Economics*, 14(3): 1–29.

- De Paola, Maria, and Francesca Gioia.** 2016. “Who performs better under time pressure? Results from a field experiment.” *Journal of Economic Psychology*, 53: 37–53.
- De Paola, Maria, and Vincenzo Scoppa.** 2014. “The effectiveness of remedial courses in Italy: a fuzzy regression discontinuity design.” *Journal of Population Economics*, 27(2): 365–386.
- Diamond, Rebecca, and Petra Persson.** 2016. “The long-term consequences of teacher discretion in grading of high-stakes tests.” National Bureau of Economic Research.
- Dizon-Ross, Rebecca.** 2019. “Parents’ beliefs about their children’s academic ability: Implications for educational investments.” *American Economic Review*, 109(8): 2728–65.
- Dobbie, Will, Roland G Fryer, et al.** 2011. “Exam high schools and academic achievement: Evidence from New York City.” National Bureau of Economic Research.
- Dustmann, Christian, Patrick A Puhani, and Uta Schönberg.** 2017. “The long-term effects of early track choice.” *The Economic Journal*, 127(603): 1348–1380.
- Ebenstein, Avraham, Victor Lavy, and Sefi Roth.** 2016. “The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution.” *American Economic Journal: Applied Economics*, 8(4): 36–65.
- Ehlers, Tim, and Robert Schwager.** 2016. “Honest grading, grade inflation, and reputation.” *CESifo Economic Studies*, 62(3): 506–521.
- Elsner, Benjamin, and Ingo E Isphording.** 2017. “A big fish in a small pond: Ability rank and human capital investment.” *Journal of Labor Economics*, 35(3): 787–828.
- Elsner, Benjamin, Ingo E Isphording, and Ulf Zölitz.** 2021. “Achievement rank affects performance and major choices in college.” *The Economic Journal*, 131(640): 3182–3206.
- Falch, Torberg, and Linn Renée Naper.** 2013. “Educational evaluation schemes and gender gaps in student achievement.” *Economics of Education Review*, 36: 12–25.
- Feng, Andy, and Georg Graetz.** 2017. “A question of degree: the effects of degree class on labor market outcomes.” *Economics of Education Review*, 61: 140–161.
- Ferman, Bruno, and Luiz Felipe Fontes.** 2022. “Assessing knowledge or classroom behavior? Evidence of teachers’ grading bias.” *Journal of Public Economics*, 216.

- Fidjeland, Andreas.** 2023. “Using high-stakes grades to incentivize learning.” *Economics of Education Review*, 94.
- Figlio, David N, and Maurice E Lucas.** 2004. “Do high grading standards affect student performance?” *Journal of Public Economics*, 88(9-10): 1815–1834.
- Fischer, Mira, and Valentin Wagner.** 2023. “Do timing and reference frame of feedback influence high-stakes educational outcomes?” *Economics of Education Review*, 94.
- Foote, Andrew, Lisa Schulkind, and Teny M Shapiro.** 2015. “Missed signals: The effect of ACT college-readiness measures on post-secondary decisions.” *Economics of Education Review*, 46: 39–51.
- Freier, Ronny, Mathias Schumann, and Thomas Siedler.** 2015. “The earnings returns to graduating with honors—Evidence from law graduates.” *Labour Economics*, 34: 39–50.
- Gaggero, Alessio, and Denni Tommasi.** 2023. “Time of Day and High-Stake Cognitive Assessments.” *The Economic Journal*, 133(652): 1407–1429.
- Gneezy, Uri, John A. List, Jeffrey A. Livingston, Xiangdong Qin, Sally Sadoff, and Yang Xu.** 2019. “Measuring Success in Education: The Role of Effort on the Test Itself.” *American Economic Review: Insights*, 1(3): 291–308.
- Gong, Jie, and Jessica Pan.** 2023. “The returns to an additional year of education for college graduates.” *Journal of Public Economics*, 218: 104796.
- Goodman, Joshua, Oded Gurantz, and Jonathan Smith.** 2020. “Take Two! SAT Retaking and College Enrollment Gaps.” *American Economic Journal: Economic Policy*, 12(2): 115–158.
- Goulas, Sofoklis, and Rigissa Megalokonomou.** 2021. “Knowing who you actually are: The effect of feedback on short- and longer-term outcomes.” *Journal of Economic Behavior & Organization*, 183: 589–615.
- Goulas, Sofoklis, Silvia Griselda, and Rigissa Megalokonomou.** 2024. “Comparative advantage and gender gap in STEM.” *Journal of Human Resources*, 59(6): 1937–1980.
- Griselda, Silvia.** 2024. “Gender gap in standardized tests: What are we measuring?” *Journal of Economic Behavior & Organization*, 221: 191–229.

- Hakimov, Rustamdjan, Renke Schmacker, and Camille Terrier.** 2022. “Confidence and college applications: Evidence from a randomized intervention.” WZB Berlin Social Science Center SP II 2022-209.
- Hanna, Rema N, and Leigh L Linden.** 2012. “Discrimination in grading.” *American Economic Journal: Economic Policy*, 4(4): 146–68.
- Hansen, Anne Toft, Ulrik Hvidman, and Hans Henrik Sievertsen.** 2024. “Grades and employer learning.” *Journal of Labor Economics*, 42(3): 659–682.
- Hanushek, Eric A., and Margaret E. Raymond.** 2005. “Does school accountability lead to improved student performance?” *Journal of Policy Analysis and Management*, 24(2): 297–327.
- Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman.** 2013. “Are some degrees worth more than others? Evidence from college admission cutoffs in Chile.” National Bureau of Economic Research.
- Heinesen, Eskil.** 2018. “Admission to higher education programmes and student educational outcomes and earnings—Evidence from Denmark.” *Economics of Education Review*, 63: 1–19.
- Heissel, Jennifer A., Emma K. Adam, Jennifer L. Doleac, David N. Figlio, and Jonathan Meer.** 2019. “Testing, Stress, and Performance: How Students Respond Physiologically to High-Stakes Testing.” *Education Finance and Policy*, 16(2): 183–208.
- Heissel, Jennifer A., Emma K. Adam, Jennifer L. Doleac, David N. Figlio, and Jonathan Meer.** 2021. “Testing, Stress, and Performance: How Students Respond Physiologically to High-Stakes Testing.” *Education Finance and Policy*, 16(2): 183–208.
- Hermes, Henning, Martin Huschens, Franz Rothlauf, and Daniel Schunk.** 2021. “Motivating low-achievers—Relative performance feedback in primary schools.” *Journal of Economic Behavior & Organization*, 187: 45–59.
- Hermo, Santiago, Miika Päällysaho, David Seim, and Jesse M. Shapiro.** 2022. “Labor Market Returns and the Evolution of Cognitive Skills: Theory and Evidence.” *The Quarterly Journal of Economics*, 137(4): 2309–2361.
- Hernandez-Julian, Rey, and Adam Looney.** 2016. “Measuring inflation in grades: An application of price indexing to undergraduate grades.” *Economics of Education Review*, 55: 220–232.

- Hinnerich, Björn Tyrefors, Erik Höglin, and Magnus Johannesson.** 2011. “Are boys discriminated in Swedish high schools?” *Economics of Education Review*, 30(4): 682–690.
- Hoekstra, Mark.** 2009. “The effect of attending the flagship state university on earnings: A discontinuity-based approach.” *The review of economics and statistics*, 91(4): 717–724.
- Högberg, Björn, and Daniel Horn.** 2022. “National High-Stakes Testing, Gender, and School Stress in Europe: A Difference-in-Differences Analysis.” *European Sociological Review*, 38(6): 975–987.
- Hvidman, Ulrik, and Hans Henrik Sievertsen.** 2021. “High-Stakes Grades and Student Behavior.” *Journal of Human Resources*, 56(3): 821–849.
- Hyman, Joshua.** 2017. “ACT for All: The Effect of Mandatory College Entrance Exams on Postsecondary Attainment and Choice.” *Education Finance and Policy*, 12(3): 281–311.
- Iriberry, Nagore, and Pedro Rey-Biel.** 2019. “Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-stage Competition in Mathematics.” *The Economic Journal*, 129(620): 1863–1893.
- Iriberry, Nagore, and Pedro Rey-Biel.** 2021. “Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large scale natural field experiment.” *European Economic Review*, 131.
- Jackson, C. Kirabo.** 2013. “Can higher-achieving peers explain the benefits to attending selective schools? Evidence from Trinidad and Tobago.” *Journal of Public Economics*, 108(C): 63–77.
- Jackson, C Kirabo.** 2018. “What do test scores miss? The importance of teacher effects on non-test score outcomes.” *Journal of Political Economy*, 126(5): 2072–2107.
- Jacob, Brian A.** 2005. “Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools.” *Journal of Public Economics*, 89(5-6): 761–796.
- Jacob, Brian A, and Lars Lefgren.** 2004. “Remedial education and student achievement: A regression-discontinuity analysis.” *Review of economics and statistics*, 86(1): 226–244.
- Jacob, Brian A, and Lars Lefgren.** 2009. “The effect of grade retention on high school completion.” *American Economic Journal: Applied Economics*, 1(3): 33–58.

- Jacob, Brian A, and Steven D Levitt.** 2003. “Rotten apples: An investigation of the prevalence and predictors of teacher cheating.” *The Quarterly Journal of Economics*, 118(3): 843–877.
- Kang, Le, Ziteng Lei, Yang Song, and Peng Zhang.** 2024. “Gender Differences in Reactions to Failure in High-Stakes Competition: Evidence from the National College Entrance Exam Retakes.” *Journal of Political Economy Microeconomics*, 2(2): 355–397.
- Khoo, Pauline, and Ben Ost.** 2018. “The effect of graduating with honors on earnings.” *Labour Economics*, 55: 149–162.
- Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad.** 2016. “Field of study, earnings, and self-selection.” *The Quarterly Journal of Economics*, 131(3): 1057–1111.
- Landaud, Fanny, and Eric Maurin.** 2023. “Tracking when Ranking Matters.” THEMA 2023-12.
- Landaud, Fanny, and Eric Maurin.** 2025. “The Choice of Retaking Competitive Exams.” THEMA 2025-06.
- Landaud, Fanny, Éric Maurin, Barton Willage, and Alexander Willén.** 2024. “The value of a high school gpa.” *Review of Economics and Statistics*, 1–24.
- Landaud, Fanny, Son Thierry Ly, and Éric Maurin.** 2020. “Competitive schools and the gender gap in the choice of field of study.” *Journal of Human Resources*, 55(1): 278–308.
- Lavy, Victor, and Edith Sand.** 2018. “On the origins of gender gaps in human capital: Short- and long-term consequences of teachers’ biases.” *Journal of Public Economics*, 167: 263–279.
- Leschnig, Lisa, Guido Schwerdt, and Katarina Zigova.** 2022. “Central exams and adult skills: Evidence from PIAAC.” *Economics of Education Review*, 90.
- Li, Hongyan, and Xing Xia.** 2024. “Grades as signals of comparative advantage: How letter grades affect major choices.” *Journal of Economic Behavior & Organization*, 227: 106717.
- Lindahl, Erica.** 2016. “Are teacher assessments biased? – evidence from Sweden.” *Education Economics*, 24(2): 224–238.

- Linder, Anna, Ulf-G. Gerdtham, and Gawain Heckley.** 2025. “Adolescent Mental Health: Impact of Introducing Earlier Compulsory School Grades.” *Health Economics*, 34(9): 1731–1746.
- Machin, Stephen, Sandra McNally, and Jenifer Ruiz-Valenzuela.** 2020. “Entry through the narrow door: The costs of just failing high stakes exams.” *Journal of Public Economics*, 190: 104224.
- Main, Joyce B, and Ben Ost.** 2014. “The impact of letter grades on student effort, course selection, and major choice: A regression-discontinuity analysis.” *The Journal of Economic Education*, 45(1): 1–10.
- Manacorda, Marco.** 2012. “The cost of grade retention.” *Review of Economics and Statistics*, 94(2): 596–606.
- McEwan, Patrick J, Sheridan Rogers, and Akila Weerapana.** 2021. “Grade sensitivity and the economics major at a women’s college.” Vol. 111, 102–106, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
- Metcalf, Robert, Simon Burgess, and Steven Proud.** 2019. “Students’ effort and educational achievement: Using the timing of the World Cup to vary the value of leisure.” *Journal of Public Economics*, 172: 111–126.
- Molin, Elin.** 2024. “Long-Term Impacts of Achievement-Based School Assignment.” Working paper.
- Murphy, Richard, and Felix Weinhardt.** 2020. “Top of the class: The importance of ordinal rank.” *The Review of Economic Studies*, 87(6): 2777–2826.
- Neal, Derek, and Diane Whitmore Schanzenbach.** 2010. “Left Behind by Design: Proficiency Counts and Test-Based Accountability.” *The Review of Economics and Statistics*, 92(2): 263–283.
- Nordin, Martin, Gawain Heckley, and Ulf Gerdtham.** 2019. “The impact of grade inflation on higher education enrolment and earnings.” *Economics of Education Review*, 73: 101936.
- Öckert, Björn.** 2010. “What’s the value of an acceptance letter? Using admissions data to estimate the return to college.” *Economics of Education Review*, 29(4): 504–516.
- OFS.** 2022. “Analysis of degree classifications over time - Changes in graduate attainment from 2010-11 to 2020-21.” Office for Students.

- Ost, Ben, Weixiang Pan, and Douglas Webber.** 2018. “The returns to college persistence for marginal students: Regression discontinuity evidence from university dismissal policies.” *Journal of Labor Economics*, 36(3): 779–805.
- Owen, Ann L.** 2010. “Grades, Gender, and Encouragement: A Regression Discontinuity Analysis.” *The Journal of Economic Education*, 41(3): 217–234.
- Papay, John P, Richard J Murnane, and John B Willett.** 2016. “The impact of test score labels on human-capital investment decisions.” *Journal of Human Resources*, 51(2): 357–388.
- Park, R. Jisung.** 2022. “Hot Temperature and High-Stakes Performance.” *Journal of Human Resources*, 57(2): 400–434.
- Pekkarinen, Tuomas.** 2015. “Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations.” *Journal of Economic Behavior & Organization*, 115: 94–110.
- Pistolesi, Nicolas.** 2017. “Advising students on their field of study: Evidence from a French university reform.” *Labour Economics*, 44: 106–121.
- Pop-Eleches, Cristian, and Miguel Urquiola.** 2013. “Going to a better school: Effects and behavioral responses.” *American Economic Review*, 103(4): 1289–1324.
- Pugatch, Todd, and Paul Thompson.** 2025. “Excellence for all? University honors programs and human capital formation.” *Educational Evaluation and Policy Analysis*, 47(2): 577–603.
- Rangvid, Beatrice Schindler.** 2019. “Gender Discrimination in Exam Grading? Double Evidence from a Natural Experiment and a Field Experiment.” *The B.E. Journal of Economic Analysis & Policy*, 19(2): 1–23.
- Reback, Randall.** 2008. “Teaching to the rating: School accountability and the distribution of student achievement.” *Journal of Public Economics*, 92(5-6): 1394–1415.
- Schoner, Florian, Lukas Mergele, and Larissa Zierow.** 2024. “Grading student behavior.” *Labour economics*, 90: 102570.
- Schwager, Robert.** 2012. “Grade inflation, social background, and labour market matching.” *Journal of Economic Behavior & Organization*, 82(1): 56–66.

- Schwerdt, Guido, Martin R West, and Marcus A Winters.** 2017. “The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida.” *Journal of Public Economics*, 152: 154–169.
- Segal, Carmit.** 2012. “Working When No One Is Watching: Motivation, Test Scores, and Economic Success.” *Management Science*, 58(8): 1438–1457.
- Sievertsen, Hans Henrik, Francesca Gino, and Marco Piovesan.** 2016. “Cognitive fatigue influences students’ performance on standardized tests.” *Proceedings of the National Academy of Sciences*, 113(10): 2621–2624.
- Stinebrickner, Todd, and Ralph Stinebrickner.** 2012. “Learning about academic ability and the college dropout decision.” *Journal of Labor Economics*, 30(4): 707–748.
- Tan, Brandon Joel.** 2023. “The Consequences of Letter Grades for Labor Market Outcomes and Student Behavior.” *Journal of Labor Economics*, 41(3): 565 – 588.
- Terrier, Camille.** 2020. “Boys lag behind: How teachers’ gender biases affect student achievement.” *Economics of Education Review*, 77.
- Vigdor, Jacob L., and Charles T. Clotfelter.** 2003. “Retaking the SAT.” *Journal of Human Resources*, 38(1).
- Virudachalam, Vanitha, Sergei Savin, and Matthew P Steinberg.** 2024. “Too much information: When does additional testing benefit schools?” *Management Science*, 70(9): 6220–6233.
- Wikström, Christina, and Magnus Wikström.** 2005. “Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools.” *Economics of education Review*, 24(3): 309–322.
- Woessmann, Ludger.** 2025. “Skills and Earnings: A Multidimensional Perspective on Human Capital.” *Annual Review of Economics*, 17: 397–425.
- Zafar, Basit.** 2011. “How do college students form expectations?” *Journal of Labor Economics*, 29(2): 301–348.
- Zamarro, Gema, Collin Hitt, and Ildefonso Mendez.** 2019. “When Students Don’t Care: Reexamining International Differences in Achievement and Student Effort.” *Journal of Human Capital*, 13(4): 519–552.

Zimmerman, Seth D. 2014. “The returns to college admission for academically marginal students.” *Journal of Labor Economics*, 32(4): 711–754.

Zubrickas, Robertas. 2015. “Optimal grading.” *International Economic Review*, 56(3): 751–776.