NBER WORKING PAPER SERIES

WHAT DOES A GRADE MEAN?
INFORMATIVENESS AND STRATEGIC MANIPULATION OF GRADING SYSTEMS

Joshua S. Gans
ⓡ
Scott Duke Kominers

Working Paper 35183
http://www.nber.org/papers/w35183

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2026

This paper subsumes most of the results of an earlier version circulated under the title "Informative Grading Requires Cross-Course Comparability." We are grateful to Susan Athey, James Bland, Ralph Boleslavsky, Jiafeng Chen, Jason Furman, Michael Herron, David Laibson, Paul Milgrom, Jeff Miron, Colin Percival, Jesse Shapiro, Alex Tabarrok, and especially Yannai Gonczarowski, Jack Hirsch, Daniel Kornbluth, and Shengwu Li (who, to our knowledge, originated the term "eigengrade") for helpful conversations. We also thank Refine.ink, ChatGPT 5.5, Claude Opus 4.7 , and Gemini 3 for research assistance. Responsibility for all errors remains our own. Kominers is both a faculty member and an alum of Harvard University, whose policy conversation around grade inflation in part inspired this work; and he has expressed views on this question within Harvard based on the analysis found herein. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

What Does A Grade Mean? Informativeness and Strategic Manipulation of Grading Systems
Joshua S. Gans ⓡ Scott Duke Kominers
NBER Working Paper No. 35183
May 2026
JEL No. D81, D82

ABSTRACT

When do university grades permit informative comparisons across courses, and how does transcript adjustment affect student and instructor incentives? A raw grade mixes student performance with course-specific conditions, so grade-only comparisons fail whenever course effects are large enough to reverse ability rankings at grade cutoffs. We show that full transcripts can recover comparable student signals through what we call eigengrades: course-adjusted reports that use common or externally anchored grading standards and enrollment overlap to identify centered student effects. In the scalar additive benchmark, row-mean, affinity-spectral, and graph-Laplacian methods recover the same object. Eigengrades are, therefore, not a separate source of identification; they are a representation of fixed-effect adjustment. The framework also clarifies limits: ordinary letter grades with unanchored course-specific cutoffs do not separate course difficulty from grading standards, and multidimensional transcripts identify a skill-match subspace rather than a unique universal ranking unless the institution specifies a benchmark. Finally, exact difficulty adjustment removes the direct report-mediated incentive to choose easier courses and eliminates a competitive enrollment channel behind grade inflation, while leaving other strategic and governance margins intact.

Joshua S. Gans
University of Toronto
Rotman School of Management
and NBER
joshua.gans@rotman.utoronto.ca

ⓡ

Scott Duke Kominers
Harvard University
Harvard Business School
kominers@fas.harvard.edu

1 Introduction

A university transcript is expected to summarize many heterogeneous classroom experiences in a few symbols. A student may receive an "A" in an introductory course with many weak classmates, a "B" in an advanced course taught by a strict instructor, and a "PASS" in a seminar whose main output is a research paper. Employers and graduate schools, as well as fellowship and award committees, nevertheless use these grading records to compare students. The transcript is therefore asked to transform local classroom outcomes into a portable signal of student quality.

This pressure is visible in policy debates over grade inflation. A natural administrative response to grade inflation is to cap the fraction of high grades in each course. Such a cap appears to solve the inflation problem by construction: if only a fixed share of students can receive an A, then average grades cannot drift upward too far. But the cap also changes the meaning of the grade: an A becomes a statement about position within a course-specific cohort rather than a statement about an absolute performance level. If one course contains unusually strong students and another contains weaker students—or one course is less difficult than the other—then the same letter grade can have different meanings.

In this paper, we ask when grades can provide informative cross-course comparisons of student performance. We argue that the central issue is not grade inflation alone; rather, it is an identification problem. A raw grade is a one-dimensional report generated in a setting with at least two relevant dimensions: a student component and a course component. A high grade can reflect high ability, a favorable course environment, a lenient grading standard, or some combination thereof. Unless the observer has additional information beyond just the grade, these components are not separately identified.

We build the analysis in three steps. We first study the strongest possible interpretation of raw grades. A universally calibrated grading rule uses the same performance thresholds in every course. Ability sufficiency requires that whenever one student receives a higher grade than another, the first student is at least as able as the second. We show that this pointwise interpretation is equivalent to ruling out grade-boundary reversals: a lower-ability student in a sufficiently favorable course must never cross above a higher-ability student in an unfavorable course at any grade cutoff. In additive environments, the relevant primitive condition is transparent. A reversal across a cutoff occurs exactly when the easy-hard course-effect gap exceeds the ability gap and the cutoff lies between the two induced performance levels. Thus the impossibility result is best read as a curriculum-span characterization: grade-only comparability fails when course heterogeneity is large enough to swamp ability differences at a grade boundary. Throughout the paper, we return to a five-student, eight-course

running example that makes these three steps explicit: first a grade-boundary reversal, then transcript-level eigengrades, and finally the course-choice incentive effect.

This formulation clarifies what the negative result does and does not say. The result is not that grades contain no information, nor is it that every observed curriculum must contain a reversal. Rather, it says that a scalar report cannot uniformly identify a one-dimensional student effect when performance also depends on a course effect. The same grade boundary cuts through the ability–difficulty plane; moving along that boundary, student ability substitutes against course difficulty. Once the realized curriculum contains enough variation to cross the boundary in opposite directions, grade-only comparisons fail to be informative about student ability.

We next ask what survives once the pointwise informativeness requirement is weakened. If an outside observer sees only a grade and sets a wage or priority score equal to the posterior mean of ability, a weaker requirement is that higher grades imply higher posterior expected ability. This property can hold under exogenous course assignment and a monotone–likelihood-ratio condition on the difficulty-marginalized performance density. The assumption is meaningful: exogeneity fixes the mixture over courses, and MLRP makes higher performance better news about ability. By contrast, when students sort into courses endogenously, the mixture over course difficulty can depend on ability. We show that such selection can reverse the posterior ordering.

The constructive part of the paper turns from single grades to full transcripts and deliberately recasts the problem as a two-way fixed-effect measurement problem. Suppose latent cardinal performance of student i in course c is

$$P_{ic} = a_i - d_c + \varepsilon_{ic},$$

where a_i is a student effect and d_c is a course effect. In this additive benchmark, a student's own average grades cannot generally remove arbitrary course effects—but the full student-course matrix can. With complete noiseless data, row means net of the grand mean recover centered student effects. The same object can also be represented as a scaled leading eigenvector of a centered student affinity matrix. With incomplete noiseless data, connected enrollment overlap is necessary and sufficient for identification up to a common additive constant, and the exact recovery operator is the pseudoinverse of the bipartite graph Laplacian. The spectral language is therefore not a new source of identification beyond fixed-effect connectedness; it is a representation and implementation layer that makes the transcript-adjustment target explicit.

The additive benchmark is deliberately stylized. The parameter a_i should be read as

the one-dimensional student component relevant to the maintained measurement model, not as a complete description of a person. The parameter d_c is likewise a composite course intercept. It may capture intrinsic difficulty, grading standards, assessment design, instructor strictness, or other additive course-level shifters. The model is useful because it isolates the fixed-effect confounding problem cleanly. When the analyst has only ordinal grades or noisy data, additional assumptions are required. We, therefore, state an ordered-response identification theorem, an anchoring interpretation for ordinal transcripts, and a noisy-cardinal precision bound rather than treating ordinary letter grades as if they were structural cardinal observations.

The same logic extends to multidimensional student skills and course demands, but the object recovered from transcripts changes. In a Multi-Dimensional Additive Interaction (MDAI) model, performance has a scalar baseline component plus a low-rank match term between student skill profiles and course skill demands. Double-centering removes the scalar student and course baselines and leaves the interaction matrix. Spectral methods can then recover the comparative-advantage subspace, but that subspace is not itself a universal ranking of students. A scalar summary requires an explicit benchmark, such as expected performance on a shared core curriculum.

Finally, we study strategic response. Students choose courses partly for their educational value and partly for their transcript consequences. Meanwhile, instructors may choose grading leniency partly to attract enrollment or avoid complaints. We define a transcript mechanism's strategic susceptibility as the spread across courses in expected market valuation, holding latent student effects fixed and using a report-only market benchmark. Exact difficulty adjustment makes the spread zero in the additive benchmark: once the public report is a function of a_i rather than of $a_i - d_c$, there is no direct signaling return to taking an easier course. For instructors, we connect this same exactness condition to a reduced-form demand semi-elasticity by treating that parameter as the pass-through from course-level mean-grade shifts into the market valuation in the student choice model. Under report-mediated demand, exact adjustment makes that pass-through zero, eliminating the competitive enrollment component of leniency while leaving direct preferences for high grades intact.

The policy implication is a two-signal architecture. For external reporting, the institution should use course-context information and transcript overlap to construct an adjusted comparison signal (or at least provide such information on the transcript, so that those reading it can extract the signal). For internal governance, the institution should separately monitor the raw input process: course means, dispersion, assessment design, and the preservation of within-course variation. A distributional grade cap conflates these tasks—it may restrain average grades, but it also makes grade thresholds cohort-dependent and can destroy precisely

the within-course information needed for the grading system to be informative.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 introduces the environment. Section 4 proves the grade-only impossibility theorem. Section 5 studies weaker Bayesian informativeness. Section 6 develops transcript-level fixed-effect identification and spectral implementations. Section 7 studies student course choice. Section 8 studies instructor leniency. Section 9 discusses grading architecture. Section 10 collects limitations and extensions. Section 11 presents concluding remarks.

2 Related Work

The closest literature studies grade inflation, grading standards, and course choice. Sabot and Wakeman-Linn (1991) show that grading standards can affect student course selection. Bar et al. (2009) study the Cornell grade-information experiment. Hernández-Julián and Looney (2016) provide an empirical price-indexing decomposition of rising grades into student characteristics, course choices, and residual inflation; their evidence that course choices explain a substantial part of grade growth is a direct empirical counterpart to our separation of student effects, course effects, and residual grade drift. Herron and Markovich (2017) model student sorting between ability-concealing and ability-revealing departments and show that high grades can reflect endogenous sorting rather than lower standards alone. Chan et al. (2007) analyze grades, course evaluations, and academic choice, while Popov and Bernhardt (2013) study university competition and grading standards. Boleslavsky and Cotton (2015) study strategic grading as Bayesian persuasion with endogenous investment in education quality, showing that less-informative grading can raise school investment and, in some cases, evaluator welfare. Rojstaczer and Healy (2012) document the long-run rise of average grades, and Johnson (2003) discusses the institutional consequences of grade inflation. In contrast to this prior work, our analysis centers on an identification question: even without strategic behavior, raw grades generally cannot separate student effects from course effects.

The paper is also related to information economics and signaling. In the classic setting of Spence (1973), costly actions may signal type because costs differ by type. Here course choice and instructor leniency are valuable in part because raw grades load on course-specific conditions. The remedy is therefore not only to interpret actions as signals, but to change the measurement system so that the course component is removed. The connection to Blackwell comparisons is similar: a richer transcript can be more informative than a single grade (Blackwell, 1953), but its value depends on how students and instructors respond to the reporting rule; related incentive issues arise in manipulable-signal environments such as Kominers ⓡ Gans (2026).

A matching-market connection follows Ostrovsky and Schwarz (2010), who study schools that choose transcript structures and may strategically compress information. We ask whether the raw signal being disclosed already confounds ability with course difficulty, and whether internal strategic behavior distorts that signal before the matching market observes it. Once the fixed-effect confounding has been removed, a school may still choose how much information to disclose.

Finally, the spectral grading rule constructions we present belong to a broader family of ranking and comparison methods. PageRank, paired-comparison models, and rank centrality all extract latent rankings from relational data (Page et al., 1999; Bradley and Terry, 1952; Luce, 1959; Negahban et al., 2017). The common theme is that the pattern of comparisons can contain information not present in any single observation. Our primitive data are bipartite student–course outcomes, and the main task is to remove course fixed effects while preserving student effects. This also links the paper to fixed-effect identification in connected bipartite graphs, as in Abowd et al. (1999). The noisy and incomplete transcript problem is also adjacent to matrix-completion and interactive-fixed-effects methods for panel data; Athey et al. (2021), for example, use nuclear-norm regularization to impute missing potential outcomes in causal panel settings and relate matrix completion to synthetic control, unconfoundedness, and interactive fixed effects. Our graph-Laplacian and eigengrade results are instead identification results for a student–course measurement problem, but the shared statistical theme is that partially observed matrices can reveal latent row and column structure once enough overlap is available.

3 Environment

There is a finite set of students $S = \{1, \dots, n\}$ and a finite set of courses $C = \{1, \dots, m\}$. Student i has a one-dimensional latent student effect $a_i \in \mathbb{R}$. Course c has a one-dimensional latent course effect $d_c \in \mathbb{R}$. Higher a_i improves performance; higher d_c lowers performance. We call a_i *ability* and d_c *difficulty* for brevity, but the constructive results concern the composite course effect rather than a literal difficulty parameter. Table 1 collects the notation used in the model setup, the running example, and the transcript-adjustment and incentive sections that follow.

Table 1: Notation

Symbol	Meaning
S, C	Finite sets of students and courses.
n, m	Number of students and courses.
i, j; c, k	Student indices; course indices.
a_i	Scalar student effect, or baseline ability, for student i.
$\bar{a}$	Average scalar student effect, $n^{-1}\sum_i a_i$.
d_c	Scalar course effect for course c; higher d_c lowers performance and is interpreted as greater difficulty or stricter course conditions.
$\pi(a, d)$	Latent performance function, increasing in ability and decreasing in difficulty.
P_{ic}	Latent cardinal performance of student i in course c.
ε_{ic}	Idiosyncratic performance shock.
$G = \{g_1 < \cdots < g_K\}$	Ordered set of grade labels.
$g(\cdot)$	Grading rule mapping latent performance into a grade label.
t_k; τ_k	Common grade thresholds; ordered-response cutpoints in the ordinal model.
$\mathcal{R}$; $Y_i^{\mathcal{R}}$	Transcript mechanism; public information reported about student i under that mechanism.
$\hat{a}_i^{\mathcal{R}}$	Observer's posterior assessment of student i's ability under report $\mathcal{R}$.
E	Observed enrollment edge set in the student–course bipartite graph.
B, L, L^+	Signed incidence matrix, graph Laplacian $B^\top B$, and its Moore–Penrose pseudoinverse.
M_n, M_m	Student and course centering matrices.
$\alpha_i^\star$	Scalar eigengrade target, $a_i - \bar{a}$.
$\hat{\alpha}^{AE}$; $\hat{\alpha}^{\mathrm{GL}}$	Complete-data affinity-eigenvector score; graph-Laplacian adjusted score.
R_{ic}	Raw cardinal score in the running example.
q_i, v_i; Q_c, V_c	Student-specific quantitative and verbal strengths; course-specific quantitative and verbal demands in the running example.
u_i, v_c; U, V	Multidimensional student skill profile and course demand profile; the corresponding factor matrices.
$m_i(c; \mathcal{R})$	Expected market valuation for student i after choosing course c under mechanism $\mathcal{R}$.
$b_i(c; \mathcal{R})$; $\Omega(\mathcal{R})$	Course-choice bias and strategic susceptibility of transcript mechanism $\mathcal{R}$.
$\bar{G}_c$; $\eta_{\mathcal{R}}$	Mean raw grade in course c; report-mediated mean-grade demand semi-elasticity.

3.1 Performance

A performance function maps ability and difficulty into a latent performance score.

Definition 3.1 (Performance function)**.** A function $\pi : \mathbb{R}^2 \to \mathbb{R}$ is a *performance function* if $\pi(a, d)$ is strictly increasing in a and strictly decreasing in d.

Theorem 3.1 only encodes the basic comparative statics of course performance. Our grade-only impossibility theorem (Theorem 4.5) requires that course conditions can matter enough to offset ability differences.

Assumption 3.2 (Technology-level overlap)**.** The performance technology satisfies *technology-level overlap* if, for every pair of attainable performance levels $p_H > p_L$ in the range of π, there exist ability levels $a_H > a_L$ and difficulties d_H, d_L such that

$$\pi(a_L, d_L) \geq p_H \qquad \text{and} \qquad \pi(a_H, d_H) \leq p_L.$$

Theorem 3.2 is a technology-level richness condition. It says that easy-course/hard-course reversals are feasible in the underlying technology. It is stronger than what is needed for any fixed realized curriculum or any fixed grading rule. The realized-curriculum condition that matters for a particular transcript is the presence of a grade-boundary reversal, characterized in Theorem 4.4 below. For example, the additive model $\pi(a, d) = a - d$ satisfies the technology-level condition on any domain of difficulties wide enough to offset the relevant performance gaps.

For constructive results, we specialize to an additive latent-performance benchmark.

Assumption 3.3 (Additive latent performance)**.** When student i takes course c, latent cardinal performance is *additive* if

$$P_{ic} = a_i - d_c + \varepsilon_{ic}, \tag{1}$$

where $\mathbb{E}[\varepsilon_{ic}] = 0$. Unless otherwise stated, the idiosyncratic shocks are independent across student-course pairs and independent of (a_i, d_c).

Theorem 3.3 is restrictive, but it isolates the fixed-effect confounding problem sharply because it is exactly the structure under which the course component enters as an additive course intercept. Later sections discuss what changes when grades are ordinal, transcripts are incomplete, or the latent student effect is multidimensional.

3.2 Grades and Transcript Mechanisms

A grading rule maps performance into a finite ordered set of grade labels $G = \{g_1 < \cdots < g_K\}$. The most transparent class is that of threshold rules, which reflect canonical finite-valued monotone grading maps.[1]

Definition 3.4 (Threshold grading rule)**.** A grading rule $g : \mathbb{R} \to G$ is a *threshold grading rule* if there exist cutoffs

$$-\infty = t_0 < t_1 < \cdots < t_{K-1} < t_K = \infty$$

such that

$$g(p) = g_k \quad \Longleftrightarrow \quad p \in [t_{k-1}, t_k).$$

While we start by considering just reading rules, eventually we will look at more general mechanisms based on transcripts, i.e., grade aggregation mechanisms that can use the full matrix of student-course outcomes, course labels, enrollment overlap, and other institutional information.

Definition 3.5 (Transcript mechanism)**.** A *transcript mechanism* $\mathcal{R}$ maps transcript data into public information. We write $Y_i^{\mathcal{R}}$ for the information about student i reported under $\mathcal{R}$. An outside observer forms $\hat{a}_i^{\mathcal{R}}(Y_i^{\mathcal{R}}) = \mathbb{E}[a_i \mid Y_i^{\mathcal{R}}, \mathcal{R}]$.

Raw grade reporting is a transcript mechanism, as is a grade-point average (GPA). So are difficulty-adjusted scores, course-context reports, and the spectral transcript scores we describe in the sequel.

3.3 Desiderata

We introduce two baseline objectives for a grading system.

Axiom 3.6 (Universal calibration)**.** A grading system is *universally calibrated* if the same grading rule applies in every course. Equivalently, there is a single function g such that the grade in course c is $g(\pi(a_i, d_c))$ for all i and c.

Axiom 3.7 (Ability sufficiency)**.** A grading system is *ability-sufficient* if, for all students i, j and all courses c, k,

$$g(\pi(a_i, d_c)) > g(\pi(a_j, d_k)) \quad \Longrightarrow \quad a_i \geq a_j.$$

[1]Note that here we pick a specific choice of convention for the threshold boundaries, but the chosen convention is immaterial for the results.

Universal calibration is the standard meaning of grade comparability: the same grade label is intended to correspond to the same performance standard across courses. The axiom is also what creates the identification problem. If the same performance threshold applies in every course, then the ability required to reach that threshold varies with the course effect.

Ability sufficiency, meanwhile, is a strong pointwise interpretation requirement. It says that a higher grade is weak evidence of higher ability no matter which courses generated the two grades.

3.4 A Running Example

We will use the same illustrative example throughout the paper to make the analysis concrete. There are five students and eight courses. Each student takes exactly four courses. The raw cardinal score is reported on a grade-point scale and is generated by

$$R_{ic} = 2.5 + a_i + q_i Q_c + v_i V_c - d_c. \tag{2}$$

The scalar a_i is the general student effect that the main additive model tries to recover. The variables q_i and v_i allow student ability to vary across course families, while Q_c and V_c describe the quantitative and verbal demands of each course. Thus, the example is intentionally a little richer than the scalar additive benchmark: setting $q_i = v_i = 0$ gives the exact scalar case, while the displayed version foreshadows the multidimensional extension in Section 6. The common intercept 2.5 only puts scores on a familiar grade-point scale.

Letter grades are assigned using common institutional thresholds:

$$A : R_{ic} \geq 4, \qquad B : 3 \leq R_{ic} < 4, \qquad C : 2 \leq R_{ic} < 3, \qquad D : 1 \leq R_{ic} < 2, \qquad F : R_{ic} < 1;$$

these are common cutoffs, not course-specific curves. The point of the example is therefore not that the grading rule is arbitrary, it is that even a common grading rule cannot by itself make cross-course grades comparable once courses differ.

Table 2: Student and course primitives in the running example

Student	a_i	q_i	v_i
Ava	1.10	0.30	0.10
Ben	0.60	-0.25	0.50
Cara	0.20	0.55	-0.20
Diego	-0.40	-0.10	0.35
Eli	-1.50	0.10	-0.10

Course	d_c	Q_c	V_c
Econ 101	-1.20	0.00	0.80
Film	-0.70	0.40	0.20
Statistics	0.00	0.90	0.00
Data Science	0.60	0.80	0.10
Algorithms	0.20	0.10	0.80
Moral Philosophy	1.30	1.00	-0.10
History of Dinosaurs	-1.10	0.00	0.90
Organic Chemistry	0.90	0.50	0.10

Table 3: Observed transcripts in the running example

Student	Course 1	Course 2	Course 3	Course 4	Letter GPA
Ava	Statistics 3.87 (B)	Data Science 3.25 (B)	Moral Philosophy 2.59 (C)	Organic Chemistry 2.86 (C)	2.50
Ben	Econ 101 4.70 (A)	Film 3.80 (B)	Algorithms 3.28 (B)	History of Dinosaurs 4.65 (A)	3.50
Cara	Film 3.58 (B)	Statistics 3.20 (B)	Data Science 2.52 (C)	Organic Chemistry 2.06 (C)	2.50
Diego	Econ 101 3.58 (B)	Algorithms 2.17 (C)	Moral Philosophy 0.66 (F)	History of Dinosaurs 3.52 (B)	2.00
Eli	Econ 101 2.12 (C)	Film 1.72 (D)	Algorithms 0.73 (F)	History of Dinosaurs 2.01 (C)	1.25

The enrollment graph in Table 3 is connected. Cara links Film to Statistics and Data Science; Diego links Econ 101 and Algorithms to Moral Philosophy; and the remaining courses attach through shared students. This connectedness is what will later let the graph-Laplacian adjustment propagate comparisons through the transcript network rather than requiring every student to take every course.

4 The Grade-Only Impossibility

A single scalar grade cannot solve cross-course comparability on a rich curriculum. The logic has two parts. A reversal that crosses a grade boundary violates ability sufficiency. Technology-level overlap and non-triviality guarantee that such a reversal is feasible; the boundary-overlap characterization identifies exactly when a realized curriculum contains one.

Axiom 4.1 (Non-triviality)**.** A *non-trivial* grading rule separates at least two attainable performance levels. That is, there exist $p_H > p_L$ in the range of π such that $g(p_H) > g(p_L)$.

Non-triviality rules out the degenerate rule that assigns everyone the same grade.

Definition 4.2 (Grade-separated reversal)**.** Fix a universally calibrated grading rule g. A realized curriculum contains a *grade-separated reversal* if there exist performance levels $p_H > p_L$ with $g(p_H) > g(p_L)$, students i_H, i_L, and courses c_H, c_L such that

$$a_{i_H} > a_{i_L}, \qquad \pi(a_{i_L}, d_{c_L}) \geq p_H, \qquad \pi(a_{i_H}, d_{c_H}) \leq p_L.$$

Under a grade-separated reversal, the lower-ability student reaches the higher performance region because the course is sufficiently favorable. The higher-ability student falls into the lower performance region because the course is sufficiently unfavorable.

The next lemma states that once such a reversal occurs, the grade can no longer serve as a reliable cross-course ranking of ability.

Lemma 4.3 (Grade-separated reversals violate ability sufficiency)**.** *Let g be a universally calibrated threshold grading rule. If a realized curriculum contains a grade-separated reversal under g, then ability sufficiency fails.*

Proof. By monotonicity of the threshold rule,

$$g(\pi(a_{i_L}, d_{c_L})) \geq g(p_H) > g(p_L) \geq g(\pi(a_{i_H}, d_{c_H})).$$

The lower-ability student receives a strictly higher grade than the higher-ability student; this contradicts Axiom 3.7. □

Theorem 4.3 isolates the basic obstruction to informative grading: once a lower-ability student crosses above a higher-ability student at a grade boundary, grade-based comparison fails. Our next proposition shows that this is not merely one way failure can occur. For threshold grading rules, such boundary overlap is exactly the condition that breaks ability sufficiency on a realized curriculum.

Proposition 4.4 (Boundary-overlap characterization)**.** *Fix a realized curriculum and a universally calibrated threshold rule g with finite cutoffs $t_1, \ldots, t_{K-1}$. Ability sufficiency holds on that curriculum if and only if there is no tuple (i_L, i_H, c_L, c_H, r) such that*

$$a_{i_H} > a_{i_L} \qquad \textit{and} \qquad \pi(a_{i_L}, d_{c_L}) \geq t_r > \pi(a_{i_H}, d_{c_H}).$$

In the additive case $\pi(a, d) = a - d$, such a boundary overlap across cutoff t_r is equivalently

$$a_{i_L} - d_{c_L} \geq t_r > a_{i_H} - d_{c_H},$$

or, in gap form,

$$d_{c_H} - d_{c_L} > a_{i_H} - a_{i_L} \qquad \textit{with} \qquad t_r \in (a_{i_H} - d_{c_H},\, a_{i_L} - d_{c_L}].$$

Proof. If such a tuple exists, the lower-ability student i_L receives a grade weakly above the grade interval beginning at t_r, while the higher-ability student i_H receives a grade strictly below that interval. Hence, the lower-ability student receives a strictly higher grade, violating ability sufficiency. Conversely, if ability sufficiency fails on the realized curriculum, then there are students and courses with $a_i < a_j$ but

$$g(\pi(a_i, d_c)) > g(\pi(a_j, d_k)).$$

For a threshold rule, some finite cutoff t_r must separate the two latent performance values:

$$\pi(a_i, d_c) \geq t_r > \pi(a_j, d_k).$$

Setting $(i_L, i_H, c_L, c_H) = (i, j, c, k)$ gives the displayed boundary overlap. The additive equivalence follows by substituting $\pi(a, d) = a - d$ and rearranging. □

Theorem 4.4 moves the discussion from intuition to a realized-curriculum test: ability sufficiency fails exactly when a common grade cutoff is crossed in opposite directions by student-course pairs with reversed ability ordering. The remaining question is whether such an obstruction is a rare accident or a generic design problem. The theorem shows that, under Theorem 3.2, any non-trivial universally calibrated threshold rule is vulnerable to precisely this kind of failure.

Theorem 4.5 (Impossibility of grade-only comparability)**.** *Let g be a universally calibrated non-trivial threshold grading rule, and suppose the performance technology satisfies Theorem 3.2. Then there exist ability levels $a_H > a_L$ and difficulties d_H, d_L that generate a grade-separated reversal. Consequently, no universally calibrated non-trivial threshold rule can guarantee ability-sufficient cross-course comparisons on every realized curriculum rich enough to contain the reversal witness.*

Proof. By Theorem 4.1, choose attainable performance levels $p_H > p_L$ such that $g(p_H) > g(p_L)$. By Theorem 3.2, there exist $a_H > a_L$ and d_H, d_L such that

$$\pi(a_L, d_L) \geq p_H \qquad \text{and} \qquad \pi(a_H, d_H) \leq p_L.$$

Any realized curriculum containing the corresponding students and courses contains a grade-separated reversal. Theorem 4.3 then implies that ability sufficiency fails. □

Our theorem is a uniform design result. It does not say that every observed curriculum contains a reversal; rather, it says that once the curriculum is rich enough to contain even one boundary-crossing easy-course/hard-course reversal, a single universally calibrated scalar grade cannot rule out cross-course misordering. The remedy must use more information than the grade alone.

It is useful to see the force of the result in the additive special case. Suppose $\pi(a, d) = a - d$ and there is a relevant cutoff t. A lower-ability student in the favorable course receives the high grade when $a_L - d_L \geq t$, while a higher-ability student in the unfavorable course receives the low grade when $a_H - d_H < t$. Such a pair exists whenever the course-effect gap exceeds the ability gap, $d_H - d_L > a_H - a_L$. The impossibility is therefore not driven by an unusual performance technology; it is driven by the interaction of course heterogeneity and coarse grade thresholds.

Running example, first pass: a grade-boundary reversal. The observed transcript in Table 3 contains the obstruction in Theorem 4.4. Ben has lower general ability than Ava, $a_{\text{Ben}} = 0.60 < 1.10 = a_{\text{Ava}}$, but Ben receives an A in Econ 101 while Ava receives a C in Organic Chemistry:

$$R_{\text{Ben,Econ 101}} = 4.70 \geq 4 \qquad \text{and} \qquad R_{\text{Ava,Organic}} = 2.86 < 3.$$

Thus a lower-ability student in an easier course receives a strictly higher grade than a higher-ability student in a harder course. The additive intuition is visible directly from the primitives in Table 2: the course-difficulty gap between Organic Chemistry and Econ 101 is

$$d_{\text{Organic}} - d_{\text{Econ 101}} = 0.90 - (-1.20) = 2.10,$$

which is much larger than the general-ability gap

$$a_{\text{Ava}} - a_{\text{Ben}} = 1.10 - 0.60 = 0.50.$$

The course gap easily swamps the ability gap, and Ben's verbal match with Econ 101 strengthens the reversal further. The example also shows why equal grade labels are not enough. Ava receives a C in Moral Philosophy, while Eli receives a C in Econ 101, even though Ava's general ability is 1.10 and Eli's is -1.50. The same letter grade is therefore not a portable comparison of student ability across courses.

The result also clarifies the informational cost of policies such as grade caps that try to constrain the distribution of ex post grades. A cap can preserve within-course monotonicity—

for example, by assigning top grades to the top fraction of students in a class—but once it binds, the effective performance threshold for a given grade becomes course- and cohort-dependent. The policy therefore abandons universal calibration in the literal sense: there is no longer a single fixed mapping from performance to grade labels across courses. At the same time, the underlying ability–course confounding remains. In that sense, a cap may restrain average grades while leaving the core measurement problem unresolved and, in some cases, making the meaning of a grade label even less stable across contexts.

5 What Raw Grades Can Still Say

Our impossibility theorem concerns pointwise cross-course comparisons. A weaker Bayesian interpretation can survive under stronger assumptions. This section separates two ideas: grade-only comparability is impossible in general, but at the same time, raw grades need not be statistically meaningless.

Let A and D denote random variables for ability and course difficulty, respectively. Suppose D is independent of A and that the latent performance density marginalized over course difficulty exists:

$$f(p \mid a) = \int f_{P|A,D}(p \mid a, d)\, dF_D(d).$$

Let the prior density of A be μ, with finite first moment. For the grade interval

$$I_k = [t_{k-1}, t_k),$$

define the conditional grade probability

$$q_k(a) = \Pr(G = g_k \mid A = a) = \int_{I_k} f(p \mid a)\, dp.$$

The key shape restriction is a monotone likelihood ratio in the *difficulty-marginalized* performance density. We state it in a division-free form so that zero-density regions do not create unnecessary regularity problems.

Assumption 5.1 (Marginalized MLRP)**.** For any $a_H > a_L$ and any performance levels $p \geq q$, the marginalized densities satisfy the cross-product inequality

$$f(p \mid a_H) f(q \mid a_L) \geq f(p \mid a_L) f(q \mid a_H) \tag{3}$$

for almost every (p, q).

Theorem 5.1 is imposed on the marginalized density $f(p \mid a)$, not derived automatically from exogenous assignment. Exogeneity is needed so that the difficulty mixture is the same across ability types. The monotone-likelihood-ratio property is an additional substantive restriction. When the two densities are strictly positive on a common support, (3) is equivalent to the usual statement that $f(p \mid a_H)/f(p \mid a_L)$ is weakly increasing in p.

Lemma 5.2 (Threshold coarsening preserves ordered likelihood ratios)**.** *Under Theorem 5.1, for any grades $g_k > g_\ell$ and any abilities $a_H > a_L$,*

$$q_k(a_H)q_\ell(a_L) \geq q_k(a_L)q_\ell(a_H).$$

Equivalently, whenever the denominators are positive, the ratio $q_k(a)/q_\ell(a)$ is weakly increasing in a.

Proof. Write $I_k = [t_{k-1}, t_k)$ and $I_\ell = [t_{\ell-1}, t_\ell)$ with $k > \ell$, so every point in I_k weakly exceeds every point in I_ℓ. For $a_H > a_L$,

$$q_k(a_H)q_\ell(a_L) - q_k(a_L)q_\ell(a_H) = \int_{I_k}\int_{I_\ell} \big[f(p \mid a_H)f(q \mid a_L) - f(p \mid a_L)f(q \mid a_H)\big]\, dq\, dp.$$

For every $(p, q) \in I_k \times I_\ell$, we have $p \geq q$. Theorem 5.1 therefore makes the integrand weakly nonnegative almost everywhere, so the double integral is weakly nonnegative. The ratio statement follows by dividing the displayed inequality by the positive product $q_\ell(a_H)q_\ell(a_L)$. □

Theorem 5.2 identifies the key statistical step: thresholding latent performance into grades preserves the relevant likelihood-ratio ordering. The next result uses that fact to show why higher grades remain better news about ability under exogenous assignment.

Proposition 5.3 (Monotone posterior means under exogenous assignment)**.** *Suppose D is independent of A, the prior density μ has finite first moment, Theorem 5.1 holds, and each realized grade level has positive unconditional probability. Then the posterior mean $\mathbb{E}[A \mid G = g_k]$ is weakly increasing in the grade level g_k.*

Proof. Let

$$Z_k = \int q_k(a)\mu(a)\, da,$$

so $Z_k > 0$ by assumption and the posterior density of A given $G = g_k$ is

$$\mu_k(a) = \frac{q_k(a)\mu(a)}{Z_k}.$$

Fix $k > \ell$. Then

$$Z_k Z_\ell \left(\mathbb{E}[A \mid G = g_k] - \mathbb{E}[A \mid G = g_\ell]\right) = \iint (a-b) q_k(a) q_\ell(b) \mu(a) \mu(b)\, da\, db$$
$$= \frac{1}{2} \iint (a-b) \big[q_k(a) q_\ell(b) - q_k(b) q_\ell(a)\big] \mu(a) \mu(b)\, da\, db.$$

On the region $\{a > b\}$, Theorem 5.2 implies that

$$q_k(a) q_\ell(b) - q_k(b) q_\ell(a) \geq 0.$$

On the region $\{a < b\}$, the same inequality with (a, b) swapped implies $q_k(a) q_\ell(b) - q_k(b) q_\ell(a) \leq 0$, while $a - b < 0$. Hence, the product in the symmetrized expression is weakly nonnegative everywhere. Therefore

$$\mathbb{E}[A \mid G = g_k] \geq \mathbb{E}[A \mid G = g_\ell]. \qquad \square$$

The practical implications of Theorem 5.3 are limited. Exogenous assignment fixes the difficulty mixture across ability types. Once students choose courses strategically, the difficulty mixture can depend on ability or preferences. Ambitious high-ability students may sort into harder courses; transcript-maximizing students may sort into easier ones. Both empirical decompositions of grade growth and formal sorting models emphasize this point: course-choice responses can change measured average grades and can make raw grade distributions hard to interpret independently of who selected into which courses (Hernández-Julián and Looney, 2016; Herron and Markovich, 2017). Either pattern can break the grade-ability ordering. Monotone posterior means are, therefore, a benchmark property, not a design solution.

Endogenous course choice breaks the fixed-mixture logic in a particularly transparent way. The next proposition gives a two-type example. It should be read as an existence result: it shows what can go wrong once the distribution of course difficulty is selected as a function of ability.

The logic of Theorem 5.3 is therefore conditional on a fixed assignment environment. Once course choice is itself informative about student type, the observer is no longer comparing grades generated from a common difficulty mixture. The next proposition shows that this is not just a technical caveat: endogenous selection can overturn the Bayesian ordering entirely.

Proposition 5.4 (Selection can reverse Bayesian informativeness)**.** *Suppose Theorem 3.2 holds, and let g be a universally calibrated non-trivial threshold grading rule. There exists a two-type endogenous course-selection environment in which higher-ability students take harder courses and a higher observed grade implies lower posterior expected ability.*

Proof. By non-triviality, choose attainable performance levels $p_H > p_L$ with $g(p_H) > g(p_L)$.

By Theorem 3.2, there exist $a_H > a_L$ and d_H, d_L such that

$$\pi(a_L, d_L) \geq p_H, \qquad \pi(a_H, d_H) \leq p_L.$$

Because π is strictly increasing in ability and strictly decreasing in the course component, these inequalities imply $d_H > d_L$; otherwise $a_H > a_L$ and $d_H \leq d_L$ would imply $\pi(a_H, d_H) > \pi(a_L, d_L)$, contradicting $p_H > p_L$.

Now construct an environment with $A \in \{a_L, a_H\}$, each type occurring with probability $1/2$. The lower type takes the lower course component d_L, while the higher type takes the higher course component d_H. Thus, higher-ability students take harder courses. The lower type's grade is at least $g(p_H)$, while the higher type's grade is at most $g(p_L)$. Hence, the lower type receives a strictly higher grade. Conditional on the higher observed grade, posterior expected ability is a_L; conditional on the lower observed grade, it is a_H. The grade ordering is therefore reversed. □

The contrast between Theorem 5.3 and Theorem 5.4 identifies the role of selection. Under exogenous assignment, the grade is a coarsened signal drawn from a fixed difficulty mixture. Under endogenous assignment, the grade is also a signal of the course chosen. The observer must either condition on course identity or use transcript-level structure to estimate course effects.

6 Transcript-Level Eigengrades and Fixed-Effect Adjustment

The prior section refined the basic obstruction to grade-based comparison: a single observed grade mixes the student component with the course component, and so cannot support universally valid inference about ability. Transcripts contain additional cross-student and cross-course information through overlap in enrollments. We now show that, in the additive benchmark, this relational information is enough to recover student and course effects.

The algebra in this section is intentionally close to standard two-way fixed-effect identification. Connectedness of the student–course graph identifies additive student and course effects up to a common shift; the graph-Laplacian pseudoinverse is the corresponding least-squares recovery operator. The contribution here is to put that familiar object to work as a grading mechanism: to state the transcript-adjustment estimand, to show how the complete-data row-mean and affinity-eigenvector formulas represent the same target, to make the ordinal and noisy-data requirements explicit, and to use exact recovery as the benchmark for the

incentive analysis that follows.

Definition 6.1 (Scalar eigengrade)**.** In the scalar additive benchmark $P_{ic} = a_i - d_c$, the *eigengrade* of student i is the centered student effect

$$\alpha_i^\star = a_i - \bar{a}, \qquad \alpha^\star = a - \bar{a}\,\mathbf{1}_n.$$

Theorem 6.1 separates the object from the operator used to compute it. The object of interest is the course-adjusted student component $\alpha^\star$, not any single matrix formula. The term *eigengrade* is appropriate because one exact construction writes $\alpha^\star$ as a scaled leading eigenvector of a student–affinity matrix. But that eigenvector formula is only one representation of the same estimand. In complete cardinal data, $\alpha^\star$ is row mean minus grand mean; in incomplete cardinal data, the same vector is recovered from the student block of a graph-Laplacian pseudoinverse solution. The ordinal and noisy-data subsections then explain what extra measurement structure is needed to infer this same target when grades are not exact cardinal performance observations.

6.1 Why Own-Transcript Averages Are Not Enough

A weighted average of a student's own grades cannot remove arbitrary course fixed effects unless it also removes all information. The missing information is cross-sectional. To see this, we start with a linear score based only on student i's own course outcomes:

$$T_i = \sum_{c=1}^{m} w_{ic} P_{ic}. \tag{4}$$

Proposition 6.2 (No non-trivial own-transcript invariant score)**.** *In the* noiseless additive model $P_{ic} = a_i - d_c$*, suppose T_i is given by* (4)*. If T_i is independent of the difficulty vector $d \in \mathbb{R}^m$ for every possible d, then $w_{ic} = 0$ for every course c.*

Proof. Under the noiseless additive model,

$$T_i = a_i \sum_{c=1}^{m} w_{ic} - \sum_{c=1}^{m} w_{ic} d_c.$$

Independence from d for every $d \in \mathbb{R}^m$ requires

$$\sum_{c=1}^{m} w_{ic} d_c = 0 \qquad \text{for every } d \in \mathbb{R}^m,$$

which is possible only if $w_{ic} = 0$ for all c. □

Theorem 6.2 explains why the adjustment problem cannot be solved student-by-student. If a score is constructed only from a student's own grades, then eliminating arbitrary course effects forces the score to collapse to the trivial zero rule. The missing information is therefore relational: to recover a student's course-adjusted component, the mechanism must compare outcomes across students and courses jointly.

6.2 Full-Matrix Requirements

Allow the institution to compute student i's score from the full performance matrix:

$$T_i(P) = \sum_{j=1}^{n} \sum_{c=1}^{m} L_{i,jc} P_{jc}. \tag{5}$$

Proposition 6.3 (Linear adjustment requirements)**.** *In the noiseless additive model*

$$P = a\mathbf{1}_m^\top - \mathbf{1}_n d^\top,$$

the linear score (5) *is invariant to arbitrary additive course shifts*

$$P \mapsto P - \mathbf{1}_n h^\top \qquad \text{for all } h \in \mathbb{R}^m$$

if and only if

$$\sum_{j=1}^{n} L_{i,jc} = 0 \qquad \text{for every course } c. \tag{6}$$

Moreover, $T_i(P) = a_i - \bar{a}$ *for every* (a, d) *if and only if* (6) *holds and*

$$\sum_{c=1}^{m} L_{i,jc} = \mathbf{1}\{j = i\} - \frac{1}{n} \qquad \text{for every student } j. \tag{7}$$

Proof. For any P and h,

$$T_i(P - \mathbf{1}_n h^\top) - T_i(P) = -\sum_{c=1}^{m} h_c \left(\sum_{j=1}^{n} L_{i,jc} \right);$$

this equals 0 for every h if and only if (6) holds. Under the additive model,

$$T_i(P) = \sum_{j=1}^{n} a_j \left(\sum_{c=1}^{m} L_{i,jc} \right) - \sum_{c=1}^{m} d_c \left(\sum_{j=1}^{n} L_{i,jc} \right).$$

The second term vanishes for every d exactly under (6). Conditional on that, recovering $a_i - \bar{a}$ for every a requires the coefficient on each a_j to equal $\mathbf{1}\{j = i\} - 1/n$, which is (7). □

In words, Theorem 6.3 says that a valid linear adjustment cannot be built by "correcting" each student's grades in isolation. The weights must pool information across students so that each course's common intercept cancels out, and the resulting combination must leave exactly student i's ability relative to the average student.

Theorem 6.3 moreover identifies the algebraic structure any exact linear adjustment must satisfy. To remove course heterogeneity, the score must assign weights that sum to zero within each course, so that pure course-level shifts cancel out. To recover student i's latent performance component, the remaining weights must then aggregate the student dimension in exactly the pattern that produces $a_i - \bar{a}$.

The conditions we have derived describe the structure any exact linear adjustment must satisfy, but they do not yet tell us when such an adjustment is available in closed form. The next step is therefore to study benchmark data environments in which identification is transparent, beginning with the complete noiseless matrix.

6.3 Complete Data

In complete noiseless data, row means net of the grand mean recover centered student effects, and hence recover the scalar eigengrade vector $\alpha^\star$.

Proposition 6.4 (Complete-data identification)**.** *Suppose $P_{ic} = a_i - d_c$ is observed for every student-course pair. Then the centered student effects $a_i - \bar{a}$ and centered course effects $d_c - \bar{d}$ are identified from P.*

Proof. The row mean is $\bar{P}_{i\cdot} = a_i - \bar{d}$, and the grand mean is $\bar{P}_{\cdot\cdot} = \bar{a} - \bar{d}$, so

$$\bar{P}_{i\cdot} - \bar{P}_{\cdot\cdot} = a_i - \bar{a}.$$

The column mean is $\bar{P}_{\cdot c} = \bar{a} - d_c$, so

$$\bar{P}_{\cdot\cdot} - \bar{P}_{\cdot c} = d_c - \bar{d}.$$

Thus, the centered course effects are identified as well. □

Theorem 6.4 already gives the core identification result in the most direct possible way: with complete cardinal data, centered row means recover the student object of interest. The spectral construction that follows does not improve on that benchmark in terms of identification; rather, it rewrites the same object in a form that extends more naturally to relational and incomplete-data settings.

Spectral language represents the same scalar eigengrade in eigenvector form. Let

$$M_n = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$$

be the student-centering matrix, and define the student affinity matrix

$$A = M_n P P^\top M_n. \tag{8}$$

Definition 6.5 (Scaled affinity score)**.** Let $\lambda_1(A)$ be the largest eigenvalue of A. If $\lambda_1(A) = 0$, define $\hat{\alpha}^{AE} = 0$. If $\lambda_1(A) > 0$, let $v_1(A)$ be a corresponding unit eigenvector. When the leading eigenspace is one-dimensional, orient $v_1(A)$ so that

$$v_1(A)^\top M_n P \mathbf{1}_m \geq 0.$$

The scaled affinity score is

$$\hat{\alpha}^{AE} = \sqrt{\frac{\lambda_1(A)}{m}}\, v_1(A).$$

In applications in which the leading eigenspace is not one-dimensional, this formula selects a leading spectral direction rather than a uniquely defined scalar score; the exact recovery result below uses the rank-one case in which the direction is unique unless the target vector is zero.

Proposition 6.6 (Complete-data spectral recovery)**.** *Suppose* $P = a\mathbf{1}_m^\top - \mathbf{1}_n d^\top$. *Then*

$$A = m(M_n a)(M_n a)^\top, \qquad \hat{\alpha}^{AE} = M_n a = a - \bar{a}\,\mathbf{1}_n.$$

If $M_n a \neq 0$*, then additionally* $\lambda_1(A) = m\|M_n a\|^2$ *and the leading eigenspace is one-dimensional.*

Proof. Because $M_n\mathbf{1}_n = 0$,

$$M_n P = (M_n a)\mathbf{1}_m^\top.$$

Therefore

$$A = (M_n P)(M_n P)^\top = m(M_n a)(M_n a)^\top.$$

If $M_n a = 0$, then $A = 0$, so Definition 6.5 sets $\hat{\alpha}^{AE} = 0 = M_n a$.

Now suppose $M_n a \neq 0$. The matrix $m(M_n a)(M_n a)^\top$ is rank one with unique nonzero eigenvalue $m\|M_n a\|^2$ and unit leading eigenvector $(M_n a)/\|M_n a\|$ up to sign. The sign convention in Definition 6.5 selects the positive orientation, because

$$\left(\frac{M_n a}{\|M_n a\|}\right)^\top M_n P \mathbf{1}_m = m\|M_n a\| > 0.$$

Hence,

$$\hat{\alpha}^{AE} = \sqrt{\frac{m\|M_n a\|^2}{m}} \cdot \frac{M_n a}{\|M_n a\|} = M_n a = a - \bar{a}\,\mathbf{1}_n. \qquad \square$$

Theorem 6.6 confirms that, in complete data, the eigengrade is not a new estimand created by spectral machinery; it is exactly the same centered student effect recovered by row means. The reason to introduce the spectral representation is forward-looking. Once the matrix is no longer complete, direct row averaging is no longer available in the same way, and the relevant structure becomes the overlap graph generated by observed enrollments.

6.4 Incomplete Transcripts

The analysis in the prior section shows that it is possible to identify student effects when we can compare the performance of all students in all courses. But in practice, student transcripts are typically incomplete, in the sense that not all students take all courses. Thus, we have to operationalize the spectral grading idea for the observed bipartite graph of student-course pairs $\mathcal{G} = (S, C, E)$, where $E \subseteq S \times C$ is the edge set of observed enrollments.[2]

Proposition 6.7 (Connected-graph identification)**.** *In the noiseless additive model with observations on E, the vectors $(a_1, \ldots, a_n)$ and $(d_1, \ldots, d_m)$ are identified up to a common additive constant if and only if $\mathcal{G}$ is connected.*

Proof. Suppose two parameter vectors (a, d) and (a', d') generate the same observed outcomes. Then for every observed edge $(i, c) \in E$,

$$a_i - d_c = a'_i - d'_c,$$

or equivalently,

$$a_i - a'_i = d_c - d'_c;$$

[2]This is conceptually related to matrix-completion methods for incomplete panel data, especially the emphasis in Athey et al. (2021) on low-rank and interactive-fixed-effects structure under nontrivial missing-data patterns. Our noiseless graph result is deliberately more primitive: it asks when the additive student–course effects are identified at all, before adding regularization or distributional assumptions.

on a connected graph, this equality propagates along any alternating student-course path, so all student and course differences must equal the same constant κ. Thus $a'_i = a_i - \kappa$ for all i and $d'_c = d_c - \kappa$ for all c. If the graph is disconnected, one may apply different constants on different connected components without changing any observed edge value, so identification fails beyond componentwise shifts. □

Theorem 6.7 shows that overlap is enough precisely when it links the transcript network into a single connected component. But identification alone is not yet a grading rule. To turn connectedness into an explicit adjustment operator, we need a matrix object that encodes those overlap comparisons and solves for the student and course effects simultaneously. The natural object is the *bipartite graph Laplacian*.

Index edges arbitrarily and stack observed outcomes into $p \in \mathbb{R}^{|E|}$. Let $B \in \mathbb{R}^{|E|\times(n+m)}$ be the signed incidence matrix with one row per edge: the row for edge (i, c) has a 1 in student column i, a -1 in course column $n + c$, and 0s elsewhere. For $x = (\alpha, \delta) \in \mathbb{R}^{n+m}$,

$$(Bx)_{ic} = \alpha_i - \delta_c.$$

Define the Laplacian

$$L = B^\top B.$$

For any $x = (\alpha, \delta)$, let $\bar{\alpha}(x) = n^{-1} \sum_{i=1}^{n} \alpha_i$ denote the average of the student block, and define the student-zero-sum–normalization operator

$$N_s(x) = x - \bar{\alpha}(x)\,(\mathbf{1}_n, \mathbf{1}_m).$$

Subtracting the same constant from both blocks leaves all fitted edge differences unchanged and sets the student block mean to zero.

Definition 6.8 (Graph-Laplacian–adjusted score)**.** Suppose $\mathcal{G}$ is connected. Define

$$x^{\mathrm{GL}} = L^+ B^\top p,$$

where L^+ is the Moore–Penrose pseudoinverse of L. Let

$$N_s(x^{\mathrm{GL}}) = (\hat{\alpha}^{\mathrm{GL}}, \hat{\delta}^{\mathrm{GL}}).$$

The student block $\hat{\alpha}^{\mathrm{GL}}$ is the graph-Laplacian–adjusted score.

Theorem 6.9 (Incomplete-graph graph-Laplacian recovery)**.** *Suppose $p_{ic} = a_i - d_c$ for every*

observed edge and $\mathcal{G}$ is connected. Then

$$\hat{\alpha}^{\mathrm{GL}} = a - \bar{a}\,\mathbf{1}_n.$$

Proof. Let $x^\star = (a, d) \in \mathbb{R}^{n+m}$. Because $p = Bx^\star$,

$$x^{\mathrm{GL}} = L^+ B^\top p = L^+ L x^\star.$$

For any vector $z = (\alpha, \delta)$, $z \in \ker(L)$ if and only if $Bz = 0$, which means $\alpha_i = \delta_c$ on every observed edge (i, c). Because the transcript graph is connected, these equalities propagate along every alternating path, so $\ker(L) = \mathrm{span}\{(\mathbf{1}_n, \mathbf{1}_m)\}$. Therefore L^+L is the orthogonal projector onto $\ker(L)^\perp$, so

$$x^{\mathrm{GL}} = x^\star - \kappa(\mathbf{1}_n, \mathbf{1}_m)$$

for some scalar κ. The student block of x^{GL} is thus $a - \kappa\mathbf{1}_n$. Applying N_s subtracts the average of that student block, namely $\bar{a} - \kappa$, from every component. Hence, the normalized student block is

$$\hat{\alpha}^{\mathrm{GL}} = (a - \kappa\mathbf{1}_n) - (\bar{a} - \kappa)\mathbf{1}_n = a - \bar{a}\,\mathbf{1}_n. \qquad \square$$

Theorem 6.9 is the constructive counterpart to the earlier impossibility result. A single grade cannot separate student and course components, but a connected transcript network can. This means that Theorem 6.9 resolves the grade-only impossibility inside the additive noiseless benchmark: once the transcript graph is connected, the institution can recover centered student effects without requiring every student to take every course. What changes under incomplete transcripts is not the estimand but the informational route by which it is recovered: identification now propagates through paths in the overlap graph rather than through direct comparison in a complete matrix. In the noiseless model, this entails no loss in exact recovery, but in noisy data, sparse or weakly connected overlap can substantially reduce precision, making estimation less stable and, in statistical terms, slower to concentrate around the target.

Running example, second pass: eigengrades. Now compute the graph-Laplacian score on the observed cardinal entries in Table 3. Stack one observation for each student-course enrollment into the vector p, form the signed incidence matrix B, and solve

$$\hat{x} = L^+ B^\top p, \qquad L = B^\top B.$$

The student block of the normalized solution is the eigengrade vector $\hat{\alpha}^{\mathrm{GL}}$. Because the running example includes the small match terms $q_i Q_c + v_i V_c$, exact equality with a_i is not guaranteed; in the exact scalar version obtained by setting those terms to zero, Theorem 6.9 applies exactly. The point of the richer example is to show how the adjustment behaves when the transcript contains both course difficulty and genuine subject-specific fit.

Table 4: Raw GPA, raw-score averages, and eigengrades in the running example

Student	True a_i	Letter GPA	Avg. raw score	Eigengrade $\hat{\alpha}_i^{\mathrm{GL}}$
Ava	1.10	2.50	3.14	1.23
Ben	0.60	3.50	4.11	0.65
Cara	0.20	2.50	2.84	0.41
Diego	-0.40	2.00	2.48	-0.47
Eli	-1.50	1.25	1.65	-1.81

The raw GPA ranking has Ben above Ava and Cara, even though Ava has the highest general ability. The eigengrade ranking is instead

$$\text{Ava} > \text{Ben} > \text{Cara} > \text{Diego} > \text{Eli},$$

which matches the ranking of the scalar student effects in Table 2. In this small example, the correlation with true a_i is 0.812 for letter GPA, 0.837 for average raw score, and 0.996 for the graph-Laplacian eigengrade. These numbers should not be over-interpreted as statistics from a population; they are arithmetic for the toy transcript. But the arithmetic shows the measurement logic: GPA rewards the courses in which the raw scores were produced, while eigengrades use overlap to estimate and remove the course component.

The same calculation also recovers relative course effects. Because (2) contains the common intercept 2.5, the course block of $\hat{x}$ absorbs that intercept; the relative course-difficulty estimate is therefore the centered course block, $\hat{\delta}_c - \bar{\hat{\delta}}$.

Table 5: Estimated relative course difficulties in the running example

Course	True d_c	Estimated centered difficulty
Econ 101	-1.20	-1.32
Film	-0.70	-0.60
Statistics	0.00	-0.03
Data Science	0.60	0.62
Algorithms	0.20	0.09
Moral Philosophy	1.30	1.44
History of Dinosaurs	-1.10	-1.25
Organic Chemistry	0.90	1.05

Negative estimates correspond to easier courses and positive estimates to harder courses. The estimates track the intended ordering: Econ 101 and History of Dinosaurs are easy, Moral Philosophy and Organic Chemistry are hard, and the other courses sit between them. This is the "sausage" made visible: the transcript network supplies comparisons that are not present in any single grade.

Taken together, the scalar eigengrade is the same object throughout the cardinal analysis: in complete data, it is the vector of row means net of the grand mean; equivalently it is the scaled affinity score $\hat{\alpha}^{AE}$; in incomplete connected data, it is the normalized student block $\hat{\alpha}^{\mathrm{GL}}$ of the Laplacian solution. The next subsection does not change the target of interest; it changes the measurement assumptions under which that target can be inferred from observed grades.

6.5 Ordinal Grades and Noisy Cardinal Data

Our exact results so far recover the eigengrade target from latent cardinal performance. But real transcripts often report ordinal grades. We, therefore, need an explicit measurement model.

Assumption 6.10 (Ordered-response measurement with known thresholds)**.** On each observed edge $(i, c) \in E$, the reported grade $G_{ic} \in \{1, \ldots, K\}$, with $K \geq 2$, satisfies

$$G_{ic} = k \quad \Longleftrightarrow \quad \tau_{k-1} < a_i - d_c + u_{ic} \leq \tau_k,$$

where $-\infty = \tau_0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K = \infty$. The finite cutpoints $\tau_1, \ldots, \tau_{K-1}$ and the shock distribution are known; the shocks u_{ic} have strictly increasing cumulative distribution function F, and the parameters (a, d) are fixed. No cross-edge independence restriction is needed for

identification from marginal grade laws; the likelihood expression below additionally assumes conditional independence across observed edges.

Under Theorem 6.10, the edge-specific grade probabilities are

$$\Pr(G_{ic} = k \mid a_i, d_c) = q_k(a_i - d_c), \qquad q_k(z) = F(\tau_k - z) - F(\tau_{k-1} - z).$$

Lemma 6.11 (Injective ordered-response grade law)**.** *Under Theorem 6.10, the map*

$$z \longmapsto (q_1(z), \ldots, q_K(z))$$

is injective.

Proof. Because $K \geq 2$, the first finite threshold τ_1 exists. Because $\tau_0 = -\infty$,

$$q_1(z) = F(\tau_1 - z).$$

Since F is strictly increasing and τ_1 is known and finite, $q_1(z)$ is strictly decreasing in z. Therefore, equality of the entire grade-probability vector at two values z and z' implies $q_1(z) = q_1(z')$, hence $z = z'$. □

Theorem 6.11 provides the missing bridge from ordinal grades back to the additive latent index. It shows that the distribution of grades on a single observed edge pins down the underlying scalar $a_i - d_c$. Once that edge-level index is identified, the ordinal problem collapses to the same connected-graph fixed-effect problem as in Theorem 6.7: transcript overlap can then propagate those pairwise differences into student and course effects, up to a common additive normalization.

Theorem 6.12 (Ordinal identification on a connected transcript graph)**.** *Suppose the observed transcript graph $\mathcal{G}$ is connected and grades satisfy Theorem 6.10. Then the joint distribution of the grade vector $\{G_{ic} : (i, c) \in E\}$ identifies (a, d) up to a common additive constant.*

Proof. Suppose two parameter vectors (a, d) and (a', d') induce the same joint distribution of the observed grade vector. Then, in particular, they induce the same marginal distribution on every observed edge $(i, c) \in E$. Hence

$$q_k(a_i - d_c) = q_k(a'_i - d'_c) \qquad \text{for every } k \in \{1, \ldots, K\}.$$

By Theorem 6.11,

$$a_i - d_c = a'_i - d'_c \qquad \text{for every } (i, c) \in E;$$

we then see from Theorem 6.7 that (a, d) and (a', d') differ only by a common additive constant. □

Remark 6.13 (Identification versus finite-sample recovery). Theorem 6.12 is an identification-in-distribution result. It states that the ordered-response model is injective in the underlying parameters on a connected graph—it does not claim exact finite-sample recovery from one noisy ordinal realization per edge. Estimation in practice requires a specified likelihood or Bayesian procedure, and the quality of that estimation depends on the amount of data and the overlap structure. The known-link and known-cutpoint assumptions are part of the identification statement. If there is only one grade category, or if the link and cutpoints are allowed to shift freely without a normalization, the injectivity argument fails.

Theorem 6.12 shows that ordinal reporting is not by itself fatal to transcript adjustment. What matters is not cardinality per se, but whether the ordinal grading law places all courses on a common latent scale. The next proposition identifies the failure mode when that common scale is abandoned.

Proposition 6.14 (Course-specific cutpoints confound standards and course effects)**.** *Consider an ordered-response model with course-specific cutpoints:*

$$G_{ic} = k \quad \Longleftrightarrow \quad \tau_{c,k-1} < a_i - d_c + u_{ic} \leq \tau_{c,k},$$

where $-\infty = \tau_{c,0} < \tau_{c,1} < \cdots < \tau_{c,K-1} < \tau_{c,K} = \infty$ *and the link distribution* F *is known. If the finite cutpoints* $\tau_{c,1}, \ldots, \tau_{c,K-1}$ *are unrestricted by course, then the joint distribution of ordinal grades cannot separately identify the course effect* d_c *from the course-specific grading standard. Specifically, for any vector* $h \in \mathbb{R}^m$*, the transformed parameters*

$$d'_c = d_c + h_c, \qquad \tau'_{c,k} = \tau_{c,k} - h_c \quad (k = 1, \ldots, K-1)$$

produce exactly the same grade probabilities on every observed edge, with the same student-effect vector a*.*

Proof. For any edge (i, c) and grade k, the probability of receiving a grade weakly below k is

$$\Pr(G_{ic} \leq k \mid a_i, d_c, \tau_c) = F(\tau_{c,k} - a_i + d_c).$$

Under the transformed parameters,

$$\tau'_{c,k} - a_i + d'_c = (\tau_{c,k} - h_c) - a_i + (d_c + h_c) = \tau_{c,k} - a_i + d_c.$$

Thus, every cumulative grade probability, and therefore every grade probability, is unchanged. More strongly, for any realization of the shock vector, the grade event associated with each edge is the same under the original and transformed parameters; hence, every joint grade-vector probability is unchanged under the maintained shock law. The transformation preserves the ordering of the finite cutpoints because it shifts all cutpoints in course c by the same amount. Hence, ordinal transcript data alone identify only the composite course-specific thresholds $\tau_{c,k} + d_c$, not d_c and the grading standard separately. □

Theorem 6.14 is the formal reason the common-cutpoint or external-anchor assumption matters. A model with freely shifting course-specific cutpoints may still support some comparisons under additional restrictions, but it does not by itself deliver the scalar additive decomposition used by the exact eigengrade results.

Anchoring and estimation. A practical ordinal implementation must supply the grading law that Theorem 6.12 treats as known. With common cutpoints, known link F, and conditionally independent edge shocks, the natural finite-sample estimator is the constrained ordered-response likelihood

$$(\hat{a}, \hat{d}) \in \underset{\sum_i a_i = 0}{\arg\max} \sum_{(i,c)\in E} \sum_{k=1}^{K} \mathbf{1}\{G_{ic} = k\} \log\big[F(\tau_k - (a_i - d_c)) - F(\tau_{k-1} - (a_i - d_c))\big].$$

The normalization fixes the common additive indeterminacy. In applications, the common grading law can be anchored by standardized assessments, common rubrics, audit grading, externally calibrated assignments, or historical bridge samples that put course grades on a shared latent scale. Without such anchors, the exact-recovery claims should be read as applying to latent cardinal performance or to externally calibrated ordinal reports. If cutpoints are allowed to vary freely by course, Theorem 6.14 shows that course-specific grading standards and course effects are no longer separately identified by transcript data alone.

Noisy cardinal data admit an equally transparent precision calculation for recovery of the same target.

Proposition 6.15 (Noisy-cardinal precision bound)**.** *Suppose the observed transcript graph $\mathcal{G}$ is connected and the observed edge outcomes satisfy*

$$p = Bx^\star + \varepsilon, \qquad x^\star = (a, d),$$

where $\mathbb{E}[\varepsilon] = 0$ *and* $\mathrm{Cov}(\varepsilon) = \sigma^2 I_{|E|}$. *Let*

$$\hat{x} = L^+ B^\top p$$

and write $\hat{x} = (\hat{\alpha}, \hat{\delta})$. *Then*

$$\mathbb{E}[\hat{x}] = L^+ L x^\star = P_{\ker(L)^\perp} x^\star, \qquad \mathrm{Cov}(\hat{x}) = \sigma^2 L^+.$$

If $\tilde{\alpha} = M_n \hat{\alpha}$ *is the zero-sum normalized student block, then*

$$\mathbb{E}[\tilde{\alpha}] = a - \bar{a}\, \mathbf{1}_n$$

and

$$\mathbb{E}\|\tilde{\alpha} - (a - \bar{a}\, \mathbf{1}_n)\|^2 \le \sigma^2 \operatorname{tr}(L^+) \le \sigma^2 \frac{n + m - 1}{\lambda_2(L)},$$

where $\lambda_2(L) > 0$ *is the second-smallest eigenvalue of* L, *equivalently the algebraic connectivity of the connected transcript graph.*

Proof. Because $p = Bx^\star + \varepsilon$,

$$\hat{x} = L^+ B^\top p = L^+ L x^\star + L^+ B^\top \varepsilon.$$

Taking expectations yields

$$\mathbb{E}[\hat{x}] = L^+ L x^\star = P_{\ker(L)^\perp} x^\star.$$

For the covariance,

$$\mathrm{Cov}(\hat{x}) = L^+ B^\top \mathrm{Cov}(\varepsilon) B L^+ = \sigma^2 L^+ B^\top B L^+ = \sigma^2 L^+ L L^+ = \sigma^2 L^+.$$

Therefore

$$\mathbb{E}\|\hat{x} - \mathbb{E}[\hat{x}]\|^2 = \operatorname{tr}(\mathrm{Cov}(\hat{x})) = \sigma^2 \operatorname{tr}(L^+).$$

Let $e_\alpha = \hat{\alpha} - \mathbb{E}[\hat{\alpha}]$. Because $\mathcal{G}$ is connected, $\ker(L) = \operatorname{span}\{(\mathbf{1}_n, \mathbf{1}_m)\}$. Hence $P_{\ker(L)^\perp} x^\star$ differs from $x^\star$ by a common shift, so the student block of $\mathbb{E}[\hat{x}]$ equals $a - \kappa \mathbf{1}_n$ for some scalar κ. Thus

$$\mathbb{E}[\tilde{\alpha}] = M_n \mathbb{E}[\hat{\alpha}] = M_n(a - \kappa \mathbf{1}_n) = a - \bar{a}\, \mathbf{1}_n.$$

Moreover,

$$\tilde{\alpha} - (a - \bar{a}\, \mathbf{1}_n) = M_n e_\alpha.$$

Since M_n is an orthogonal projection,

$$\|M_n e_\alpha\|^2 \le \|e_\alpha\|^2 \le \|\hat{x} - \mathbb{E}[\hat{x}]\|^2.$$

Taking expectations gives

$$\mathbb{E}\|\tilde{\alpha} - (a - \bar{a}\,\mathbf{1}_n)\|^2 \le \sigma^2 \operatorname{tr}(L^+).$$

If the eigenvalues of the connected-graph Laplacian are $0 = \lambda_1(L) < \lambda_2(L) \le \cdots \le \lambda_{n+m}(L)$, then the nonzero eigenvalues of L^+ are their reciprocals, so

$$\operatorname{tr}(L^+) = \sum_{j=2}^{n+m} \frac{1}{\lambda_j(L)} \le \frac{n+m-1}{\lambda_2(L)}. \qquad \square$$

Theorem 6.15 formalizes a simple point: Connectivity is enough for exact identification in the noiseless benchmark. In noisy data, however, the overlap structure matters quantitatively through the spectrum of the graph Laplacian. Thin overlap or near-disconnected transcript graphs can make estimation unstable even when identification technically holds. If the graph is disconnected, then the covariance identity for $\hat{x}$ remains valid with the appropriate pseudoinverse, but the expectation is only componentwise centered; applying a single global student-centering operator need not recover $a - \bar{a}\mathbf{1}_n$.

6.6 Multi-Dimensional Abilities and Skill Matches

The preceding results complete the scalar additive measurement problem, showing exact recovery in the noiseless benchmark, an ordinal identification analogue under a common grading law, and a precision bound for noisy cardinal data. The next question is conceptual rather than statistical: What changes if the student component itself is not one-dimensional, so that transcripts reflect comparative advantage as well as a common baseline?

The additive model treats the student component as a scalar. That is the right benchmark for studying the basic cross-course comparability problem, but it is not the only empirically relevant structure. Students may have broad academic preparation as well as comparative advantage across fields, and courses may differ not only in average difficulty but also in the skills they reward.

A parsimonious extension is the Multi-Dimensional Additive Interaction (MDAI) model

$$P_{ic} = a_i - d_c + u_i^\top v_c + \varepsilon_{ic}, \tag{9}$$

where a_i remains a scalar baseline student effect, d_c remains a scalar baseline course effect, $u_i \in \mathbb{R}^K$ is student i's comparative-advantage profile, and $v_c \in \mathbb{R}^K$ is course c's skill-demand profile. The term $u_i^\top v_c$ is not a second scalar ability measure; it is a match term. A student can perform especially well in one family of courses and less well in another without being globally above or below another student.

The algebra for MDAI is closely related to the scalar case. Under the mean-zero normalizations $\sum_i u_i = 0$ and $\sum_c v_c = 0$, let U and V collect the student and course profiles. In complete noiseless data,

$$M_n P M_m = UV^\top, \qquad M_m = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^\top;$$

thus, double-centering removes the additive baselines a_i and d_c and leaves the low-rank interaction matrix. The student-side singular vectors of the residual matrix recover the latent skill-match subspace. These are the natural multi-dimensional analogue of eigengrades, but their interpretation is geometric rather than ordinal.

The main implication is that multidimensional transcript adjustment changes the object being reported. In the scalar additive model, the institution can recover $a_i - \bar{a}$, a one-dimensional comparison signal. In the MDAI model, transcript data identify a comparative-advantage subspace. Turning that profile into a single number requires an explicit benchmark. For example, given course weights $\omega_c \geq 0$ with $\sum_c \omega_c = 1$, one may define

$$s_i(\omega) = a_i + \sum_{c=1}^{m} \omega_c u_i^\top v_c, \tag{10}$$

which is expected baseline-adjusted performance on the course bundle ω. If ω is concentrated on a shared core curriculum, $s_i(\omega)$ is a meaningful core-performance score. A different bundle would generally induce a different ordering, which is a feature of multidimensional comparative advantage rather than a defect of the method.

The data requirements also become stronger. Connectivity of the enrollment graph is enough to identify scalar additive effects in the noiseless benchmark; it is not enough to identify a K-dimensional interaction structure. Each student must be observed in sufficiently many informative courses, each course must enroll sufficiently diverse students, and the overlap graph must span the relevant dimensions. Appendix A gives the formal MDAI model, complete-data subspace recovery result, benchmark-score invariance result, and incomplete-transcript rank and overlap conditions.

7 Student Course Choice and Incentive Neutrality

We now ask whether transcript adjustment changes strategic course choice. In the benchmark additive environment, the answer is yes in a precise sense: exact eigengrade-style adjustment eliminates the direct labor-market return to choosing an easier course. Throughout this section and the next, an "eigengrade-style" mechanism means a transcript mechanism that uses enrollment overlap to recover the centered student effect, including the complete-data affinity score and the connected-graph Laplacian score developed in Section 6. The common feature is exact difficulty adjustment: the public report depends on the student effect, not on the course effect through which the raw grade was generated.

The intuition is straightforward. Under raw GPA, replacing a harder course with an easier one can raise the observed grade because performance contains the term $-d_c$. Under exact adjustment, that same course effect is netted out. A higher raw grade earned only because the course is easier is offset by the estimated course effect, so the adjusted report does not increase. Thus, eigengrades do not merely give the market a better ex post estimate; in the benchmark, they remove the signaling gain from course-shopping itself. This is the mechanism-level counterpart of empirical and theoretical work showing that students respond to grading environments when choosing courses or departments (Sabot and Wakeman-Linn, 1991; Bar et al., 2009; Hernández-Julián and Looney, 2016; Herron and Markovich, 2017).

A student values both pedagogical match and labor-market signaling. To isolate the direct grade-arbitrage channel, this section uses a *report-only market benchmark*: the downstream observer's payoff-relevant comparison signal for student i is $Y_i^{\mathcal{R}}$ together with the rule $\mathcal{R}$. If course labels or full course portfolios are also disclosed and used by the market, they are additional signals; they must either be included in $Y_i^{\mathcal{R}}$ and satisfy the exactness condition below, or be assumed conditionally uninformative for the posterior object being valued. When comparing courses, we hold fixed the latent student-effect vector a.

We represent the downstream labor-market benefits of a course in terms of a reduced-form valuation statistic—for a transcript mechanism $\mathcal{R}$, define the *expected market valuation from choosing course c* by

$$m_i(c;\mathcal{R}) = \mathbb{E}\big[\hat{a}_i^{\mathcal{R}}(Y_i^{\mathcal{R}}) \mid a,\ i \text{ chooses } c,\ \mathcal{R}\big]\,. \tag{11}$$

The object $m_i(c;\mathcal{R})$ is the expected labor-market value of choosing course c under reporting rule $\mathcal{R}$, holding fixed the latent student-effect vector; it summarizes how the market is expected to interpret the report consequences of that course choice.

Define the *course-choice bias*

$$b_i(c;\mathcal{R}) = m_i(c;\mathcal{R}) - a_i$$

and the *strategic susceptibility*

$$\Omega(\mathcal{R}) = \sup_{i\in S}\left(\max_{c\in C}\{b_i(c;\mathcal{R})\} - \min_{c\in C}\{b_i(c;\mathcal{R})\}\right). \tag{12}$$

The bias term $b_i(c;\mathcal{R})$ measures the transcript advantage or disadvantage created by course choice alone. It is the gap between the market's expected assessment after taking course c and the student's true effect. The strategic susceptibility $\Omega(\mathcal{R})$ is the spread of the market valuation determined by course choice, that is, it measures the strength of the course-shopping incentive built into the transcript system. The larger $\Omega(\mathcal{R})$ is, the more a student's expected market evaluation can be improved or worsened purely by choosing one course rather than another; $\Omega(\mathcal{R})$ is 0 when the direct labor-market return to course shopping disappears.

The course choice below is a marginal enrollment decision, not a complete transcript-generation model. A student is choosing one course slot against a background of required courses, prior enrollments, or other institutional structure that creates the transcript graph. Formally, one may start from a background edge set E^0 and let a feasible course-choice profile add edges (i, c_i), yielding $E(c) = E^0 \cup \{(i, c_i) : i \in S\}$. The exact-adjustment conclusions below are conditioned on every feasible $E(c)$ belonging to the transcript class under consideration, for example, a connected graph. If each student instead chose exactly one course and nothing else, the graph would generally fail this condition, and the graph-Laplacian corollary would not apply.

Suppose student i chooses among courses according to

$$U_{ic} = v_{ic} + \gamma m_i(c;\mathcal{R}) + \nu_{ic}, \tag{13}$$

where v_{ic} is pedagogical value, $\gamma > 0$, and the shocks ν_{ic} are iid Type-I extreme-value. Let

$$\overset{*}{\Pr_i}(c) = \frac{\exp\{v_{ic}\}}{\sum_{k=1}^{m}\exp\{v_{ik}\}}$$

denote the purely pedagogical benchmark.

Under (13), let

$$\Pr_i(c;\mathcal{R}) = \Pr(U_{ic} \geq U_{ic'} \text{for all } c' \in C)\,; \tag{14}$$

with iid Type-I extreme-value shocks, this probability takes the multinomial-logit form

$$\Pr_i(c;\mathcal{R}) = \frac{\exp\{v_{ic} + \gamma m_i(c;\mathcal{R})\}}{\sum_{c'=1}^{m} \exp\{v_{ic'} + \gamma m_i(c';\mathcal{R})\}}. \tag{15}$$

For comparative statics, it is convenient to allow a sequence of reporting rules. A *sequence of transcript mechanisms* is a family $(\mathcal{R}_t)_{t\geq 1}$ such that each $\mathcal{R}_t$ is a transcript mechanism in the sense of Definition 3.5, with induced objects $Y_i^{\mathcal{R}_t}$, $\hat{a}_i^{\mathcal{R}_t}$, $m_i(c;\mathcal{R}_t)$, and $\Omega(\mathcal{R}_t)$.

Proposition 7.1 (Incentive neutrality)**.** *If $\Omega(\mathcal{R}) = 0$, then $\Pr_i(c;\mathcal{R}) = \Pr_i^*(c)$ for every student i and course c. More generally, if $\Omega(\mathcal{R}_t) \to 0$ along a sequence of mechanisms, then $\Pr_i(c;\mathcal{R}_t) \to \Pr_i^*(c)$ for every i and c.*

Proof. Write $m_i(c;\mathcal{R}) = a_i + b_i(c;\mathcal{R})$. In the logit formula generated by (13), the common term $\exp\{\gamma a_i\}$ cancels. If $\Omega(\mathcal{R}) = 0$, then $b_i(c;\mathcal{R})$ is constant across courses for each i, so that term also cancels and the remaining choice probabilities equal $\Pr_i^*(c)$. The sequence claim follows from continuity of the logit map as the spread in $b_i(c;\mathcal{R}_t)$ converges to zero. □

Theorem 7.1 is purely behavioral: if the reporting rule makes expected market value independent of the chosen course, then course choice is governed by pedagogical value rather than transcript arbitrage. The remaining question is whether eigengrades have this property. The answer is yes under the exact-adjustment conditions established in Section 6 and under the report-only market benchmark just stated.

To connect Theorem 7.1 to transcript adjustment, it is useful to define exactness directly.

Definition 7.2 (Exact difficulty adjustment)**.** Fix a class $\mathcal{T}$ of feasible transcript environments generated by the additive benchmark. A transcript mechanism $\mathcal{R}$ is an *exact difficulty adjustment on $\mathcal{T}$* if, for every transcript in $\mathcal{T}$ generated by parameters (a, d), the market-facing report for student i is a deterministic function of the student-effect vector alone:

$$Y_i^{\mathcal{R}} = y_i(a) \qquad \text{for all } i.$$

In the zero-sum normalization used in Section 6, $y_i(a) = a_i - \bar{a}$.

Exact difficulty adjustment means that, within the maintained transcript environment, the market-facing report for each student depends only on the student-effect vector and not on the course effects through which performance was measured. In that sense, the mechanism strips out course difficulty exactly rather than only approximately. The definition applies to the comparison report used in the report-only benchmark; separate disclosure of course portfolios is a different information channel unless it is made part of $Y_i^{\mathcal{R}}$.

Proposition 7.3 (Exact adjustment eliminates direct benchmark course shopping)**.** *Under the report-only market benchmark, suppose every feasible course-choice profile generates a transcript in a class $\mathcal{T}$ on which $\mathcal{R}$ is an exact difficulty adjustment. Then $\Omega(\mathcal{R}) = 0$.*

Proof. Fix a student i and a latent student-effect vector a. Under exact difficulty adjustment, for every feasible course choice c, the reported object $Y_i^{\mathcal{R}}$ is the same deterministic function $y_i(a)$ of the student-effect vector and does not depend on the realized course effect. Under the report-only market benchmark, the observer's valuation is computed from $Y_i^{\mathcal{R}}$ and $\mathcal{R}$, with no additional course-portfolio signal. Therefore $\hat{a}_i^{\mathcal{R}}(Y_i^{\mathcal{R}})$ is the same across all feasible c, and so is its conditional expectation $m_i(c; \mathcal{R})$. Hence $b_i(c; \mathcal{R})$ is constant across courses for every i, which implies $\Omega(\mathcal{R}) = 0$. □

Theorem 7.3 establishes the incentive logic at the level of an abstract mechanism class. To connect that statement back to the measurement results, we now verify that the transcript adjustments constructed in Section 6 are exact in precisely the sense required. The corollary therefore translates the abstract neutrality claim into a concrete property of eigengrades.

Corollary 7.4 (Eigengrades eliminate direct benchmark course shopping)**.** *Suppose every feasible course-choice profile generates either complete noiseless transcript data or a connected incomplete noiseless transcript graph. Then each of the following mechanisms is an exact difficulty adjustment on the corresponding feasible class:*

1. *centered row means in complete data;*
2. *the scaled affinity score of Definition 6.5 in complete data;*
3. *the graph-Laplacian adjusted score of Definition 6.8 on connected incomplete graphs.*

Consequently, under the report-only market benchmark, each mechanism satisfies $\Omega(\mathcal{R}) = 0$ on its feasible class, and student course-choice probabilities coincide with the purely pedagogical benchmark: $\Pr_i(c; \mathcal{R}) = \Pr_i^*(c)$ for every i and c.

Proof. Theorem 6.4 shows that centered row means equal $a_i - \bar{a}$ in complete noiseless data. Theorem 6.6 shows the same for the scaled affinity score. Theorem 6.9 shows the same for the graph-Laplacian adjusted score on connected incomplete graphs. Thus, each mechanism is exact on the stated class, and Theorem 7.3 applies. □

Theorem 7.4 shows a sense in which eigengrades eliminate the direct benchmark course-shopping problem. If the market evaluates students using a report that has removed d_c, then a student cannot raise expected market valuation by choosing a lower-d_c course. The student may still choose a course because it is more enjoyable, more useful, less time-consuming, or

better matched to their skills, but the transcript-arbitrage motive generated by raw grades is gone.

Running example, third pass: course-shopping incentives. The same transcript also shows the incentive problem. Suppose Ava considers replacing Organic Chemistry with Econ 101 while holding the rest of her transcript fixed. Under raw grades, the swap is attractive: Ava's Organic Chemistry score is 2.86, a C, while her counterfactual Econ 101 score would be 4.88, an A. Her letter GPA would rise from

$$(B + B + C + C)/4 = 2.50$$

to

$$(B + B + C + A)/4 = 3.00.$$

That is a direct transcript reward for moving from a harder course to an easier one.

Table 6: Ava's raw and difficulty-adjusted scores under a course swap

Course	Difficulty d_c	Raw score R_{ic}	Letter grade	$R_{ic} + d_c$
Organic Chemistry	0.90	2.86	C	3.76
Econ 101	-1.20	4.88	A	3.68

The last column of Table 6 adds back the course difficulty. The large raw-grade gain essentially disappears. In the exact scalar additive case without match terms, both entries in the last column would equal $2.5 + a_{\text{Ava}} = 3.60$ exactly. In the displayed example, the small remaining difference comes from real subject-specific fit, not from the fact that Econ 101 is easier. Thus, the eigengrade-style report removes the direct signaling return to choosing the easier raw-grade course, leaving only genuine pedagogical or match reasons to choose it.

The statement is deliberately a benchmark statement. It removes the *direct* signaling return to easier courses. It does not say that course portfolios carry no information in richer environments, nor that students have no comparative advantage across subjects. In the MDAI environment, for example, a portfolio may reveal specialization or match, and a benchmark-specific score may still reward courses that are informative for the chosen benchmark. Nor does the result claim exact finite-sample neutrality when grades are ordinal, transcripts are noisy, or adjustment is only approximate. In those cases, the relevant conclusion is approximate: as the adjustment error shrinks, the direct course-shopping return shrinks with it.

8 Instructor Leniency and Competitive Grade Inflation

Students are not the only strategic actors in the university grading context. Instructors may choose grading leniency, assessment difficulty, or curving policy. Here too the benchmark answer is yes, but again in a specific sense: exact difficulty adjustment in the cardinal or externally anchored additive benchmark eliminates the *competitive enrollment* component of grade inflation. It does so because a unilateral course-level grade shift is treated as a course effect rather than as evidence that the students in that course are better. The instructor can no longer attract students by making raw grades more generous if the market-facing report nets out that generosity.

This does not mean that eigengrades make grade inflation impossible in every sense. Instructors may have direct preferences for giving high grades, may face pressure from students or departments, or may compress within-course variation in ways that damage measurement without simply raising the mean. Nor does our reduced-form instructor game incorporate the investment-quality tradeoff emphasized by Boleslavsky and Cotton (2015), in which strategic grading can reduce transcript informativeness while increasing incentives to invest in education quality. Those are governance problems for the raw input process and broader welfare questions. The point of the model below is narrower and sharper: once difficulty adjustment is exact, the race to attract enrollment by being lenient loses its signaling payoff.

We model this margin through course mean grades. Let $\bar{G}_c \in [\underline{G}, \overline{G}]$ denote the mean raw grade in course c. The share equation below is a reduced-form representation of the enrollment effect of course-level mean grades after the reporting rule is fixed. We interpret $\eta_{\mathcal{R}}$ as the report-mediated semi-elasticity of enrollment demand generated by market valuation of raw mean-grade shifts. Let

$$s_c(\bar{G}; \mathcal{R}) = \frac{\exp\{\eta_{\mathcal{R}} \bar{G}_c\}}{\sum_{k=1}^{m} \exp\{\eta_{\mathcal{R}} \bar{G}_k\}}, \tag{16}$$

where $\eta_{\mathcal{R}} \geq 0$ is the reduced-form mean-grade demand semi-elasticity under mechanism $\mathcal{R}$.

The link to the student-choice primitives is the pass-through from a course-level mean-grade shift into the market valuation $m_i(c; \mathcal{R})$. Suppose, locally around a symmetric course environment, that a shift in course c's mean raw grade changes the market value of choosing that course by

$$m_i(c; \mathcal{R}; \bar{G}_c + h) - m_i(c; \mathcal{R}; \bar{G}_c) = \rho_{\mathcal{R}} h + o(h),$$

while leaving the report-mediated value of other courses unchanged. Then the logit model in (13) puts the local coefficient $\gamma \rho_{\mathcal{R}}$ on course c's mean grade in the demand index. The

softmax form (16) is the symmetric reduced form with

$$\eta_{\mathcal{R}} = \gamma \rho_{\mathcal{R}}. \tag{17}$$

Thus $\eta_{\mathcal{R}}$ is not an additional structural primitive when this mapping is imposed; it is the grade-to-market-value pass-through from Section 7 multiplied by the signaling weight in student utility. If enrollment also responds to workload, reputation, convenience, or other non-report channels, those forces are outside $\eta_{\mathcal{R}}$ and are not eliminated by transcript adjustment.

A transcript class must also be stable under the counterfactual course shifts being considered. A class $\mathcal{T}$ is *course-shift closed* if, whenever an environment in $\mathcal{T}$ is generated by (a, d) on a given enrollment graph, the environment generated on the same graph by $(a, d-h)$ also belongs to $\mathcal{T}$ for every course-shift vector h under consideration.

Definition 8.1 (Course-shift invariance)**.** Fix a class $\mathcal{T}$ of transcript environments generated by the additive benchmark. A transcript mechanism $\mathcal{R}$ is *course-shift invariant on* $\mathcal{T}$ if, for every transcript in $\mathcal{T}$ generated by (a, d) and every course-shift vector h such that the shifted environment generated by $(a, d-h)$ is also in $\mathcal{T}$, replacing (a, d) by $(a, d-h)$ leaves every student's market-facing report unchanged.

Requiring course-shift invariance is stronger than assuming $\Omega(\mathcal{R}) = 0$. The former concerns unilateral course-level shifts in raw grades; the latter concerns course choice at a fixed grade policy. In the reduced-form demand system (16), setting $\eta_{\mathcal{R}} = 0$ is a modeling implication only when enrollment demand responds to raw mean-grade shifts through the market-facing report and not through some separate preference for easy workloads, generous grading, or course reputation.

Proposition 8.2 (Exact adjustment implies course-shift invariance)**.** *If $\mathcal{T}$ is course-shift closed and $\mathcal{R}$ is an exact difficulty adjustment on $\mathcal{T}$, then $\mathcal{R}$ is course-shift invariant on $\mathcal{T}$. If, in addition, enrollment demand in* (16) *responds to course mean grades only through the market-facing report, then the reduced-form demand system satisfies $\eta_{\mathcal{R}} = 0$ on that class.*

Proof. Take any transcript environment in $\mathcal{T}$ generated by (a, d) and any course-shift vector h such that the shifted environment generated by $(a, d-h)$ is also in $\mathcal{T}$. Under exact difficulty adjustment, the market-facing report $Y_i^{\mathcal{R}}$ depends only on the ability vector a and not on the course-effect vector d. Both the original and shifted environments have the same a, so every $Y_i^{\mathcal{R}}$ is unchanged. This is exactly course-shift invariance. Under the stated report-mediated demand interpretation of (16), a unilateral course-level grade shift that leaves every market-facing report unchanged has no enrollment-demand effect through

expected market valuation. The corresponding reduced-form grade semi-elasticity is therefore zero. □

Theorem 8.2 shows how the same exactness condition that eliminates student course-shopping also disciplines instructor incentives: if eigengrade-style adjustment strips course effects out of the market-facing report, then an instructor cannot raise demand by mechanically shifting grades upward. This is the bridge to the next result, which embeds that invariance in the instructor payoff function and studies the resulting comparative-statics of leniency.

Instructor c has payoff

$$U_c(\bar{G}_c, \bar{G}_{-c}) = \alpha \bar{G}_c + \beta \log s_c(\bar{G}; \mathcal{R}) - \frac{\xi}{2}(\bar{G}_c - \bar{G}_0)^2, \tag{18}$$

where $\alpha \geq 0$ is a direct benefit from higher grades, $\beta \geq 0$ is the value of enrollment share, and $\xi > 0$ is the marginal cost of departing from the professional norm $\bar{G}_0$.

Proposition 8.3 (Instructor equilibrium)**.** *Fix $\eta_{\mathcal{R}} \geq 0$. The game with payoffs* (18) *and shares* (16) *has a unique Nash equilibrium, and it is symmetric. Let*

$$\Pi_{[\underline{G},\overline{G}]}(x) = \min\{\overline{G}, \max\{\underline{G}, x\}\}.$$

The equilibrium mean grade is

$$\bar{G}^{\mathrm{NE}} = \Pi_{[\underline{G},\overline{G}]}\left(\bar{G}_0 + \frac{\alpha + \beta\eta_{\mathcal{R}}(1 - 1/m)}{\xi}\right). \tag{19}$$

If the interior solution lies in $[\underline{G}, \overline{G}]$, the competitive enrollment component of grade inflation is

$$\frac{\beta\eta_{\mathcal{R}}(1 - 1/m)}{\xi}.$$

Proof. Define the potential

$$\Phi(\bar{G}) = \sum_{c=1}^{m}\left[(\alpha + \beta\eta_{\mathcal{R}})\bar{G}_c - \frac{\xi}{2}(\bar{G}_c - \bar{G}_0)^2\right] - \beta \log\left(\sum_{k=1}^{m} e^{\eta_{\mathcal{R}}\bar{G}_k}\right).$$

Because

$$\log s_c(\bar{G}; \mathcal{R}) = \eta_{\mathcal{R}}\bar{G}_c - \log\left(\sum_{k=1}^{m} e^{\eta_{\mathcal{R}}\bar{G}_k}\right),$$

we have $\partial\Phi/\partial\bar{G}_c = \partial U_c/\partial\bar{G}_c$ for every c. Thus the game is an exact potential game. The Hessian of Φ is

$$\nabla^2\Phi = -\xi I_m - \beta\eta_{\mathcal{R}}^2(\mathrm{diag}(s) - ss^\top),$$

where $s = (s_1, \ldots, s_m)$. The matrix $\operatorname{diag}(s) - ss^\top$ is positive semidefinite, so $\nabla^2\Phi$ is negative definite because $\xi > 0$. Hence Φ is strictly concave and has a unique maximizer. Since each player's payoff is concave in that player's own mean grade and the game is an exact potential game, the Nash equilibria are exactly the coordinatewise maximizers of Φ on the product interval. Strict concavity makes that coordinatewise maximizer the unique global maximizer, so the game has a unique Nash equilibrium.

By symmetry, the unique equilibrium is symmetric. At an interior symmetric profile, $s_c = 1/m$ for every course and the first-order condition is

$$0 = \alpha + \beta\eta_{\mathcal{R}}\left(1 - \frac{1}{m}\right) - \xi(\bar{G}_c - \bar{G}_0).$$

Solving gives

$$\bar{G}_c = \bar{G}_0 + \frac{\alpha + \beta\eta_{\mathcal{R}}(1 - 1/m)}{\xi}.$$

Projection onto $[\underline{G}, \overline{G}]$ handles boundary cases and yields (19). □

Theorem 8.3 shows that equilibrium grading has two separable components: a direct taste for higher grades, captured by α/ξ, and a competitive enrollment term, captured by $\beta\eta_{\mathcal{R}}(1 - 1/m)/\xi$. The next corollary combines that decomposition with Theorem 8.2: under exact eigengrade-style adjustment, the report-mediated demand term disappears because $\eta_{\mathcal{R}} = 0$.

Corollary 8.4 (Eigengrades eliminate the competitive enrollment component)**.** *If the relevant transcript class is course-shift closed, $\mathcal{R}$ is an exact difficulty adjustment on that class, and enrollment demand is report-mediated as in Theorem 8.2, then $\eta_{\mathcal{R}} = 0$. The unique Nash equilibrium in Theorem 8.3 therefore reduces to*

$$\bar{G}^{\mathrm{NE}} = \Pi_{[\underline{G}, \overline{G}]}\left(\bar{G}_0 + \frac{\alpha}{\xi}\right).$$

The competitive enrollment component of grade inflation vanishes.

Theorem 8.4 disentangles two motives for grading leniency. One is strategic: instructors may inflate grades to attract students when easier grading improves the transcript signal attached to their course. The other is direct: instructors or institutions may simply prefer higher grades for independent reasons. In the model, that residual motive is summarized by α. This distinction is useful when comparing our mechanism to strategic-grading models at the school level: Boleslavsky and Cotton (2015) show that noisy grading may improve welfare

by strengthening school investment incentives, whereas here, exact transcript adjustment targets the competitive raw-grade race within a fixed measurement environment.

This is the sense in which eigengrades neutralize competitive grade inflation in the model. Under raw transcripts, an instructor who raises mean grades makes the course more attractive because students expect higher market valuations. Under exact eigengrades, the market-facing report is invariant to a course-level shift, so the enrollment demand term no longer rewards unilateral leniency. What remains is not the competitive externality but a residual preference for high grades, plus any unmodeled attempts to degrade the measurement system itself. Thus, eigengrades eliminate the grade-inflation *race* generated by raw-grade comparability, but they do not remove the need to govern the raw grades from which the adjustment is computed.

9 Governance and Policy Architecture

The formal results suggest that grading policy should separate measurement from governance. The external measurement problem is to construct a comparison signal that adjusts for course effects as far as the transcript structure and maintained model allow. The internal governance problem is to preserve the quality of the raw inputs from which that comparison signal is built.

The preceding two sections have a sharper implication than simply that eigengrades may be useful. In the exact cardinal or externally anchored additive benchmark, eigengrade-style adjustments eliminate two specific incentive distortions: direct course-shopping by students under a report-only market benchmark and report-mediated competitive enrollment-driven leniency by instructors. The reason is the same in both cases. A raw grade advantage created by a favorable course environment is reclassified as a course effect rather than as student ability. Adjustment, therefore, changes behavior because it changes what the transcript signal rewards.

For external reporting, the institution can release an adjusted score, a course-effect estimate, contextual transcript information, or a coarser category derived from an adjusted signal. The exact format is a disclosure decision. The key measurement point is that course identity and enrollment overlap should not be discarded. In the additive benchmark, these objects are precisely what allow the observer to separate student and course effects.

For internal governance, the institution should monitor the raw grade process. Exact adjustment removes the competitive demand for easy grades in the model, but it does not remove direct instructor incentives to raise grades or compress distributions. A governance system should therefore track course means, within-course dispersion, assessment design, and

the stability of grading standards over time. Empirical price-indexing evidence that rising grades reflect a mixture of student characteristics, course choices, and residual grade growth reinforces this separation of measurement, composition, and governance (Hernández-Julián and Looney, 2016). A Taylor-rule-style feedback on course mean grades is one natural first-moment instrument (Taylor, 1993); such a rule targets average inflation without dictating the entire within-course distribution. In this architecture, the Taylor rule is complementary to eigengrades rather than a substitute for them: eigengrades neutralize the external signaling reward from leniency, while the Taylor rule disciplines residual raw-grade drift (see our earlier working paper version (Gans ℗ Kominers, 2026) for further discussion).

Note also that our proposed architecture differs sharply from a distributional grade cap. When such a cap binds, the effective threshold for a given grade is no longer a common institutional standard; it becomes endogenous to the course and the realized cohort. A cap may therefore reduce mean grades while making grade labels less comparable across classes and over time. It can also compress or destroy within-course variation, which is precisely the variation transcript-based adjustment exploits. In the language of the model, a cap alters the measurement system directly while addressing the incentive problem only indirectly. A grade cap, thus, may improve a headline statistic such as the share of A grades, yet worsen the institution's ability to measure and compare student performance; the same is true of any other form of forced curve.

Curricular design also matters. Theorem 6.7 gives breadth requirements and shared core courses an informational interpretation: they create paths in the student–course graph. If the course enrollment graph is disconnected, transcript data alone cannot identify relative course effects across components; this yields the natural implication that grades cannot provide informative comparisons of students in different divisions of a highly siloed academic program. In noisy data, Theorem 6.15 strengthens the point by showing that not all connected graphs are equally informative.

The policy conclusion is therefore disciplined rather than merely modest. Cross-course comparability requires course-context information, and attempts to manage grade inflation should preserve rather than destroy the variation needed for adjustment. Exact eigengrade-style adjustments eliminate the direct benchmark course-shopping channel and the report-mediated competitive grade-inflation channel by making the market-facing signal invariant to course effects. Governance must then focus on the remaining margins: the quality, stability, and dispersion of the raw grade inputs, and the possibility of residual noncompetitive pressure for high grades.

10 Scope and Limitations

Our analysis has several limitations.

One-dimensional baseline. The constructive results in the main additive benchmark rely on a scalar student effect and a scalar course effect. The MDAI extension in Section 6.6 and Appendix A allows multidimensional skills and demands, but the conclusion changes: the interaction component identifies a skill-match subspace rather than a canonical one-dimensional ranking. A scalar report is therefore a benchmark choice, such as performance on a shared core curriculum or on a specified bundle of courses.

Interpretation of the course effect. The additive course effect should not be read as pure difficulty. In applications, it can absorb grading standards, instructor strictness, assessment design, department norms, and other additive shifters. The theory, therefore, identifies a composite course effect. Normative claims about "difficulty" require additional structure.

Ordinal grades. Letter grades are ordinal. The spectral exact-recovery results therefore apply either to latent cardinal performance or to an ordered-response model with a specified or externally anchored common link and cutpoints. Theorem 6.12 is an identification result at the level of the joint grade distribution, not a claim that one noisy letter grade per edge is enough for exact finite-sample recovery. Unknown course-specific thresholds require additional anchors; if they are freely adjustable, the ordinal transcript alone identifies only composite course-specific thresholds and cannot separate grading standards from course effects.

Overlap and missingness. Connectedness is necessary for exact identification in the noiseless graph model, but it is not enough for precise estimation in noisy data. Theorem 6.15 shows that precision depends on the spectrum of the Laplacian. Sparse transcripts, low-degree courses, and near-disconnected graphs can therefore make adjustment unstable. If missingness depends on unobserved match shocks, further structure is needed.

Strategic portfolios. The course-choice neutrality result isolates the direct signaling return to easier courses under a report-only market benchmark. In richer environments, employers or graduate schools may also observe course portfolios and infer ambition, specialization, or comparative advantage from them. Those informational channels can interact with transcript adjustment and are not eliminated by exact adjustment of the scalar comparison report alone.

Partial reforms. While the main text emphasizes exact adjustment, we must note that *partial* adjustment has ambiguous incentive effects. Indeed, a partially adjusted signal may become more trusted before it fully removes the gaming margin, so the remaining arbitrage can become more valuable (see Kominers ⓡ Gans (2026)). That is a caution about implementation, not an argument against exact adjustment itself.

11 Conclusion

Cross-course comparability is a fundamental challenge in grading because the same grade label is typically used to summarize performance generated in very different course environments. A single raw grade cannot generally distinguish high ability in an unfavorable course from lower ability in a favorable one. Thus, a grading policy focused only on grade levels or frequencies misses the deeper issue.

Transcript structure offers a constructive response. In the additive benchmark, overlapping course enrollments identify student and course effects. In complete data, centered row means and a scaled affinity-eigenvector representation recover the same centered student effect. On connected incomplete transcript graphs, the corresponding graph-Laplacian adjustment recovers the same target. With ordinal grades, the same benchmark survives at the level of the joint grade distribution under an explicit ordered-response model. With noisy data, the reliability of adjustment depends on overlap strength, not merely on technical connectivity. With multidimensional skill matches, the same transcript logic recovers a comparative-advantage subspace, while any scalar ranking must be tied to an explicit benchmark.

Strategic behavior strengthens rather than weakens the case for exact adjustment, subject to the maintained measurement and demand assumptions. Raw grades create a direct signaling return to easy courses and a competitive enrollment return to lenient grading. Exact difficulty adjustment removes the first channel under the report-only market benchmark and, in the instructor game, removes the second competitive component when demand is mediated by the market-facing report through the pass-through parameter in (17). In that precise benchmark sense, eigengrades neutralize the direct course-shopping and competitive-inflation channels rather than merely diagnosing them. They do not remove every grading incentive, and they do not substitute for governance of raw inputs.

The appropriate reform is, therefore, not to try and control the grade distribution, but rather to introduce a better measurement and inference architecture: use transcript overlap to restore cross-course comparability, preserve informative within-course variation, and govern the remaining margins on which grades can lose information.

A Multi-Dimensional Abilities and Skill Matches

This appendix develops the Multi-Dimensional Additive Interaction (MDAI) extension discussed in Section 6.6. The purpose is to make precise what transcript data recover once student performance has both a scalar baseline component and a multidimensional comparative-advantage component. The main message is deliberately different from the scalar additive model: the interaction term identifies a low-rank skill-match subspace, not a canonical universal ranking of students.

A.1 Model and Normalizations

Fix an integer $K \geq 1$. Student i has a scalar baseline effect $a_i \in \mathbb{R}$ and a skill profile $u_i \in \mathbb{R}^K$. Course c has a scalar baseline effect $d_c \in \mathbb{R}$ and a skill-demand profile $v_c \in \mathbb{R}^K$. Latent performance is

$$P_{ic} = a_i - d_c + u_i^\top v_c + \varepsilon_{ic}, \tag{20}$$

where $\mathbb{E}[\varepsilon_{ic}] = 0$. Let $U \in \mathbb{R}^{n \times K}$ have rows $u_i^\top$ and let $V \in \mathbb{R}^{m \times K}$ have rows $v_c^\top$. In the noiseless case,

$$P^0 = a\mathbf{1}_m^\top - \mathbf{1}_n d^\top + UV^\top. \tag{21}$$

As in the scalar additive model, (a, d) are identified only up to a common additive constant: replacing (a, d) by $(a + \kappa\mathbf{1}_n, d + \kappa\mathbf{1}_m)$ leaves every fitted difference $a_i - d_c$ unchanged. For the interaction term, impose the centering normalizations

$$\sum_{i=1}^{n} u_i = 0, \qquad \sum_{c=1}^{m} v_c = 0, \tag{22}$$

or equivalently $M_n U = U$ and $M_m V = V$, where

$$M_n = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top, \qquad M_m = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^\top.$$

These normalizations separate the interaction term from the additive row and column effects.

The factorization $UV^\top$ has the usual change-of-basis indeterminacy. For any invertible matrix $R \in \mathrm{GL}(K, \mathbb{R})$,

$$(UR)(VR^{-\top})^\top = UV^\top.$$

Thus, the data can identify the interaction matrix and its row and column spaces, but not a uniquely labeled coordinate system for skills without additional restrictions such as anchor courses or externally specified skill labels.

A.2 Complete-Data Subspace Recovery

Definition A.1 (Multi-dimensional eigengrades)**.** Given a latent performance matrix P, define its double-centered residual matrix

$$P^{dc} = M_n P M_m.$$

The *multi-dimensional eigengrade* (MDE) representation of dimension K is the matrix $\widehat{U}_K \in \mathbb{R}^{n\times K}$ whose columns are the top K left singular vectors of P^{dc}.

Theorem A.2 (Complete-data MDAI subspace recovery)**.** *In the noiseless MDAI model with complete data, suppose U and V each have full column rank K. Then*

$$P^{dc} = M_n P^0 M_m = UV^\top,$$

so P^{dc} has rank K and the columns of $\widehat{U}_K$ span $\mathrm{col}(U)$*. Symmetrically, the top K right singular vectors span* $\mathrm{col}(V)$*.*

Proof. Using (21),

$$M_n P^0 M_m = M_n(a\mathbf{1}_m^\top)M_m - M_n(\mathbf{1}_n d^\top)M_m + M_n UV^\top M_m.$$

The first term is zero because $\mathbf{1}_m^\top M_m = 0^\top$, and the second is zero because $M_n\mathbf{1}_n = 0$. By (22), $M_nU = U$ and $V^\top M_m = V^\top$, so the remaining term is $UV^\top$. If both U and V have full column rank, then $UV^\top$ has rank K. Its left singular space is $\mathrm{col}(U)$ and its right singular space is $\mathrm{col}(V)$. □

Theorem A.2 is the multidimensional analogue of the complete-data scalar eigengrade result. The important difference is interpretive. In the scalar additive case, the student-side object is the centered vector $M_n a$, which directly orders students along the maintained student-effect dimension. In the MDAI case, the student-side object is a K-dimensional subspace. That subspace describes comparative advantage across course families; it does not by itself say that one student is unconditionally better than another.

A tempting basis-free dominance relation would require student i's interaction performance to exceed student j's in every course:

$$u_i^\top v_c \geq u_j^\top v_c \quad \text{for every course } c.$$

Under the normalization $\sum_c v_c = 0$, however, this relation is degenerate. Summing the

nonnegative differences across courses gives

$$\sum_c (u_i - u_j)^\top v_c = (u_i - u_j)^\top \sum_c v_c = 0,$$

so every inequality must hold with equality. If the course-demand vectors span $\mathbb{R}^K$, this further implies $u_i = u_j$. Hence, the MDAI interaction term does not provide a useful basis-free dominance order. Meaningful scalar comparisons require an explicit benchmark, such as the reference-bundle scores below.

A.3 Benchmark-Specific Scalar Scores

If an institution wants a one-dimensional summary in an MDAI environment, it must choose the benchmark explicitly. Let $\omega = (\omega_1, \ldots, \omega_m)$ be course weights with $\omega_c \geq 0$ and $\sum_c \omega_c = 1$.

Definition A.3 (Reference-bundle score)**.** The *reference-bundle score* associated with ω is

$$s_i(\omega) = a_i + \sum_{c=1}^{m} \omega_c u_i^\top v_c. \tag{23}$$

If ω is supported on a shared core curriculum, we call $s_i(\omega)$ a *core score*.

The course-effect term d_c does not appear in (23) because the score is intended to measure expected student performance on the reference bundle net of common course difficulty. In the noiseless model,

$$s_i(\omega) = \sum_{c=1}^{m} \omega_c (P_{ic} + d_c).$$

Proposition A.4 (Basis invariance of benchmark scores)**.** *For any ω, the reference-bundle score $s_i(\omega)$ is invariant to the change of basis $(U, V) \mapsto (UR, VR^{-\top})$ for any $R \in \mathrm{GL}(K, \mathbb{R})$. Once a and $UV^\top$ are identified up to the common additive normalization of a, rankings by $s_i(\omega)$ are therefore well defined for the chosen benchmark.*

Proof. Let $V^\top \omega = \sum_c \omega_c v_c$. Under the change of basis,

$$(UR)_i^\top (VR^{-\top})^\top \omega = u_i^\top R R^{-1} V^\top \omega = u_i^\top V^\top \omega = \sum_{c=1}^{m} \omega_c u_i^\top v_c.$$

Thus, the interaction component of $s_i(\omega)$ is basis-invariant. The remaining common shift $a \mapsto a + \kappa \mathbf{1}_n$ adds the same constant κ to every score, so it does not affect rankings. □

The proposition is why shared core courses are especially useful in a multidimensional environment. A core curriculum supplies a natural benchmark bundle ω. The resulting scalar score is not a universal ranking of all possible talents; it is expected performance on the institution's chosen reference bundle.

A.4 Incomplete Transcripts

With incomplete transcripts, recovering the MDAI interaction is a low-rank matrix-completion problem after the additive main effects have been removed or jointly estimated, closely paralleling the broader matrix-completion literature on recovering latent row–column structure from partially observed panels (Athey et al., 2021). Let $E \subseteq S \times C$ be the observed enrollment edge set, and suppose, for this subsection, that the noiseless residuals

$$r_{ic} = P_{ic} - a_i + d_c = u_i^\top v_c$$

are observed on E.

Proposition A.5 (Local degree and spanning requirement)**.** *Fix the course profiles $(v_c)_{c \in C}$. If student i is observed in the course set C_i and*

$$\operatorname{rank}\{v_c : c \in C_i\} < K,$$

then u_i is not uniquely determined by the observed residuals, absent additional restrictions linking u_i to other student profiles. In particular, $|C_i| < K$ implies non-uniqueness. Symmetrically, fixing the student profiles, if course c is observed for the student set S_c and

$$\operatorname{rank}\{u_i : i \in S_c\} < K,$$

then v_c is not uniquely determined, absent additional restrictions linking v_c to other course profiles.

Proof. If student i is observed in the course set C_i, the residual observations solve

$$u_i^\top v_c = r_{ic} \qquad (c \in C_i).$$

Let V_i be the matrix whose rows are the vectors $v_c^\top$ for $c \in C_i$. The equations are $V_i u_i = r_i$. Because the true profile is feasible, the solution set is nonempty. If $\operatorname{rank}(V_i) < K$, the null space of V_i contains a nonzero vector z, so $u_i + z$ generates the same residuals on every

observed course in C_i. Thus u_i is not uniquely determined. The course-side argument is identical with students and courses interchanged. □

Theorem A.2 gives the exact analogue of scalar eigengrade recovery in the multidimensional setting, but it also shows why the interpretation must change. What is recovered on the student side is now a subspace of interaction directions, not a single ordered latent trait.

In particular, the local degree-and-spanning requirement is not sufficient. The observed courses must also span the relevant directions in a way that is globally compatible with a single factorization. Nor is the proposition a claim that a degree of at least K is necessary under every possible set of external restrictions or cross-row normalizations; it is the local requirement for unrestricted profile recovery conditional on the opposite-side profiles. For example, observing a student in K courses with linearly dependent demand vectors still leaves some skill directions unidentified.

Theorem A.6 (Sequential sufficient condition)**.** *Suppose the noiseless residuals $r_{ic} = u_i^\top v_c$ are observed on E. Assume there are seed sets $S_0 \subseteq S$ and $C_0 \subseteq C$ with $|S_0| = |C_0| = K$ such that every pair in $S_0 \times C_0$ is observed and the corresponding residual submatrix has rank K.*

Suppose further that the remaining vertices can be ordered so that each new student is observed in at least K previously recovered courses whose v_c vectors span $\mathbb{R}^K$, and each new course is observed for at least K previously recovered students whose u_i vectors span $\mathbb{R}^K$. Then the interaction profiles (U, V) are uniquely determined up to a single global change of basis.

Proof. The seed submatrix has rank K, so it admits a rank-K factorization with full-rank student and course factors. Any two such factorizations differ by a single invertible change of basis. This fixes a coordinate system up to that global transformation.

Proceed along the stated ordering. When a new student i is observed in previously recovered courses whose profiles span $\mathbb{R}^K$, the equations $u_i^\top v_c = r_{ic}$ uniquely determine u_i in the current coordinate system. When a new course c is observed for previously recovered students whose profiles span $\mathbb{R}^K$, the equations $u_i^\top v_c = r_{ic}$ uniquely determine v_c. Induction recovers all profiles in the same coordinate system, leaving only the initial global change of basis. □

Theorem A.6 makes explicit why simple connectedness is no longer enough. Scalar additive effects can be chained through single overlap paths. A K-dimensional interaction requires overlap that spans K directions.

A.5 Noisy Latent Scores under MCAR Missingness

The next result records the direct perturbation analogue of Theorem A.2. It is an oracle latent-score statement, not a complete theory of sparse ordinal transcript estimation.

Theorem A.7 (MDE consistency under dense MCAR latent-score observation)**.** *Assume the MDAI model* (20)*. Let* $\Omega_{ic} \sim \text{Bernoulli}(p)$ *be independent observation indicators, independent of the shocks, with fixed* $p \in (0, 1]$*. Observe* $Y_{ic} = \Omega_{ic} P_{ic}$ *and form the inverse-probability-weighted matrix* $\widehat{P}$ *with entries* $\widehat{P}_{ic} = Y_{ic}/p$*. Let*

$$\widehat{P}^{dc} = M_n \widehat{P} M_m,$$

and let $\widehat{U}_K$ *be its top* K *left singular vectors. Let* $S = UV^\top$ *and let* U_K *be any orthonormal basis for* $\text{col}(U)$*. Suppose* $\sigma_K(S) > 0$ *and that, for constants* $A, D, B_0 < \infty$*,*

$$|a_i| \le A, \qquad |d_c| \le D, \qquad |u_i^\top v_c| \le B_0$$

for all i, c*. Define*

$$\Delta_{n,m,p} = \frac{A + D + B_0 + \sigma}{p}(\sqrt{n} + \sqrt{m}).$$

If the shocks are independent mean-zero random variables with sub-Gaussian norm $\|\varepsilon_{ic}\|_{\psi_2} \le \sigma$*, then there exist universal constants* $C, c > 0$ *such that, with probability at least* $1 - 2\exp\{-c(n+m)\}$*,*

$$\left\|\sin\Theta(\widehat{U}_K, U_K)\right\|_{\text{op}} \le \min\left\{1,\ C\frac{\Delta_{n,m,p}}{\sigma_K(S)}\right\}.$$

Consequently, along any sequence of problems with fixed p *and* $\Delta_{n,m,p}/\sigma_K(S) \to 0$*, the recovered student interaction subspace is consistent in operator-norm sine distance.*

Proof. Write

$$P^0 = a\mathbf{1}_m^\top - \mathbf{1}_n d^\top + UV^\top, \qquad M_0 = A + D + B_0.$$

Then $|P^0_{ic}| \le M_0$ for every i, c, and

$$\widehat{P} = P^0 + W, \qquad W_{ic} = \left(\frac{\Omega_{ic}}{p} - 1\right) P^0_{ic} + \frac{\Omega_{ic}}{p}\varepsilon_{ic}.$$

Conditional on the fixed arrays (a, d, U, V), the entries W_{ic} are independent and mean zero. The first term is bounded in absolute value by M_0/p, hence has sub-Gaussian norm at most a universal constant times M_0/p. The second term has sub-Gaussian norm at most a universal constant times σ/p, because multiplication by the Bernoulli indicator and scaling by $1/p$

preserve sub-Gaussianity up to that factor. Therefore

$$\|W_{ic}\|_{\psi_2} \le C_1 \frac{M_0 + \sigma}{p}$$

for a universal constant C_1.

Double-centering yields

$$\widehat{P}^{dc} = M_n P^0 M_m + M_n W M_m = S + M_n W M_m,$$

where $M_n P^0 M_m = S$ by the proof of Theorem A.2. A standard rectangular matrix concentration bound for independent, centered, uniformly sub-Gaussian entries gives universal constants $C_2, c > 0$ such that, with probability at least $1 - 2\exp\{-c(n+m)\}$,

$$\|M_n W M_m\|_{\mathrm{op}} \le \|W\|_{\mathrm{op}} \le C_2 \frac{M_0 + \sigma}{p}(\sqrt{n} + \sqrt{m}) = C_2 \Delta_{n,m,p}.$$

On this event, if $C_2\Delta_{n,m,p} \le \sigma_K(S)/2$, Wedin's singular-subspace perturbation bound for the rank-K signal S gives

$$\left\|\sin\Theta(\widehat{U}_K, U_K)\right\|_{\mathrm{op}} \le \frac{2\|M_n W M_m\|_{\mathrm{op}}}{\sigma_K(S)} \le 2C_2 \frac{\Delta_{n,m,p}}{\sigma_K(S)}.$$

If instead $C_2\Delta_{n,m,p} > \sigma_K(S)/2$, the trivial bound $\|\sin\Theta(\widehat{U}_K, U_K)\|_{\mathrm{op}} \le 1$ is bounded by $C\Delta_{n,m,p}/\sigma_K(S)$ for a sufficiently large universal constant C. Combining the two cases yields the displayed minimum bound. □

The theorem deliberately assumes dense MCAR observation and latent cardinal scores. Ordinal grades require an ordered-response first stage, and sparse or strategically selected transcripts require additional graph-dependent and selection-model assumptions. Those qualifications are not nuisances; they are the multidimensional version of the paper's central point that transcript information is valuable only when the observation structure contains enough overlap to separate student effects, course effects, and match effects.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Athey, S., M. Bayati, N. Doudchenko, G. W. Imbens, and K. Khosravi (2021): "Matrix Completion Methods for Causal Panel Data Models," *Journal of the American Statistical Association*, 116, 1716–1730.

Bar, T., V. Kadiyali, and A. Zussman (2009): "Grade Information and Grade Inflation: The Cornell Experiment," *Journal of Economic Perspectives*, 23, 93–108.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *The Annals of Mathematical Statistics*, 24, 265–272.

Boleslavsky, R. and C. Cotton (2015): "Grading Standards and Education Quality," *American Economic Journal: Microeconomics*, 7, 248–279.

Bradley, R. A. and M. E. Terry (1952): "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," *Biometrika*, 39, 324–345.

Chan, W., L. Hao, and W. Suen (2007): "A signaling theory of grade inflation," *International Economic Review*, 48, 1065–1090.

Gans, J. S. ⓡ S. D. Kominers (2026): "Informative Grading Requires Cross-Course Comparability," Working paper, Harvard University.

Hernández-Julián, R. and A. Looney (2016): "Measuring Inflation in Grades: An Application of Price Indexing to Undergraduate Grades," *Economics of Education Review*, 55, 220–232.

Herron, M. C. and Z. D. Markovich (2017): "Student Sorting and Implications for Grade Inflation," *Rationality and Society*, 29, 355–386.

Johnson, V. E. (2003): "Grade Inflation: A Crisis in College Education," *Springer*.

Kominers, S. D. (2026): "Grade Caps Fail the Game Theory Exam," *The Harvard Crimson*.

Kominers, S. D. ⓡ J. S. Gans (2026): "Manipulable Signals in Matching Markets," Working paper, University of Toronto.

Luce, R. D. (1959): *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley.

Negahban, S., S. Oh, and D. Shah (2017): "Rank Centrality: Ranking from Pairwise Comparisons," *Operations Research*, 65, 266–287.

Ostrovsky, M. and M. Schwarz (2010): "Information disclosure and unraveling in matching markets," *American Economic Journal: Microeconomics*, 2, 34–63.

Page, L., S. Brin, R. Motwani, and T. Winograd (1999): "The PageRank Citation Ranking: Bringing Order to the Web," *Stanford InfoLab Technical Report*.

Popov, S. V. and D. Bernhardt (2013): "University Competition, Grading Standards, and Grade Inflation," *Economic Inquiry*, 51, 1764–1778.

Rojstaczer, S. and C. Healy (2012): "Where A is Ordinary: The Evolution of American College and University Grading, 1940–2009," *Teachers College Record*, 114, 1–23.

Sabot, R. and J. Wakeman-Linn (1991): "Grade Inflation and Course Choice," *Journal of Economic Perspectives*, 5, 159–170.

Spence, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374.

Taylor, J. B. (1993): "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.