Economica

Economica (2021) **88**, 364–398
doi:10.1111/ecca.12358

Electives Shopping, Grading Policies and Grading Competition

By Martin Gregor

Charles University

Final version received 3 November 2020.

This paper analyses grading competition between instructors of elective courses when students shop for high course scores, the instructors maximize class size, and the school imposes a ceiling on mean course scores to limit grade inflation. We demonstrate that curriculum flexibility (more listed courses or fewer required courses) intensifies the competition: in particular, top scores increase. To tame incentives to provide large scores, we suppose that the school additionally introduces a top-score grading policy. We consider three regimes. First, the school caps top scores. Then grading competition segregates students into a concentrated group of achievers and a dispersed group of laggards. This effect extends to constraints on scores at lower quantiles. Second, the school normalizes the range of scores by adjusting the mean-score ceiling. On normalization, scores of a less flexible curriculum first-order stochastically dominate scores of a more flexible curriculum. Hence all students prefer rigid curricula. Third, the school requires that the mean score is evaluated for enrolled students instead of a representative sample of students. Then the instructors stop competing for students, which introduces assortative inefficiencies. Overall, we show that addressing grade inflation through grading policies may generate inequalities, rigidities and inefficiencies.

Introduction

While course instructors and students primarily care about content and quality of courses, it is also the case that students choose courses to shop for the best grades, and some instructors attract students through lenient grading. Indeed, there is extensive evidence of both electives shopping on the side of students (Sabot and Wakeman-Linn 1991; Bar *et al*. 2009; Babcock 2010; Hernández-Julián and Looney 2016) as well as grading competition between instructors. The instructors are motivated by concerns about teaching evaluations (Correa 2001; Butcher *et al*. 2014) and pressure to enrol many students in a class when attendance is a performance metric (Angling and Meng 2000). Grading competition is particularly pronounced in the context of electives that are close substitutes within a single programme, but it may also exist on the level of departments and schools when a programme choice or even a field choice is involved.

For the shape of grading competition, the key is the set of grading policies that bind the instructors. In this paper, we focus on grading competition regulated by a *mean-score ceiling* (also known as a mean-preserving grading policy, a grade ceiling or a mean-grade quota). That is, a school requires a mean course score for a representative sample of students to not exceed a given ceiling, but otherwise leaves grading in the hands of the instructors. The ceiling fully controls the incentives to increase scores across the board but otherwise leaves the shape of the scores distribution unregulated. An example of the policy in practice is the well-studied decision of the small, elite Wellesley College by which average grades in courses with at least 10 students should not exceed a B+ average (Butcher *et al*. 2014). Gorry (2017) documents a similar example in a business school of a large US public university, where in 2014 an average recommended grade was set to 2.8 in introductory courses and 3.2 in intermediate courses.

We analyse grading competition in the presence of the mean-score ceiling, especially when this policy is combined with additional grading policies that regulate top scores. In particular, we analyse how curriculum flexibility and grading policies jointly shape the grading competition. That curriculum flexibility may have an effect on the moments of the distributions can be observed in a variety of higher education environments, but most visibly when comparing grades across levels. For instance, Achen and Courant (2009) report that most departments treat students much better in upper-level courses, and also that upper-level grades are higher than those in the lower levels.

To give an illustrative example, Table 1 lists the grade distributions in the Juris Doctor (JD) programme at Pitt Law School in the year 2019.[1] At Pitt Law, JD students enjoy a large degree of latitude in designing courses of study that meet their individual goals and interests, with only a handful of graduation requirements beyond the first year. In other words, curriculum structures differ across classes, with more senior students having greater flexibility. Table 1 shows that the distribution of inferred grade point averages (GPAs)[2] for the first-year students who face the most rigid curriculum (Class of 2020) exhibits the lowest mean, lowest median (B), and largest variance. In contrast, the median grade of more senior classes is B+, and variance decreases relative to the first-year class.

Using additional data from 2016 and 2018, Table 2 illustrates that grades of more senior students are in all years on average higher (by 0.14–0.22) and exhibit lower variance. Additionally, between-variation of the Class of 2018 reveals that two extra years of seniority have added 0.20 to the average grade.

In the case of a fixed student body, curriculum flexibility affects grades through two simultaneous channels: (i) a change in the instructors' grading schemes, and (ii) an increased self-selection into courses that are more likely to award better grades. To our best knowledge, no existing model in the literature fully characterizes and separates the two channels and explains how they change in the presence of the mean-score ceiling and other grading policies.

To fill the gap, this paper builds a symmetric model of grading competition based on Continuous Lotto games (Hart 2008, 2016). In the model, students exclusively shop for good grades, and instructors maximize enrolment. Each student has a unique personal fit to each elective course, and the fit is *ex ante* unknown both to the student and instructor. The instructor constructs the grading rule as an *arbitrary* mapping from the fit to the course score such that (i) it is non-decreasing, and (ii) the average score for all potential students (or, for a representative sample) is fixed (a mean-score ceiling). In the shopping period, a student costlessly attends all electives and acquires free signals of his or her fit through *interim scores*. When the shopping period closes, each student enrols into courses with the most promising interim scores; since interim scores are assumed to be unbiased signals of *ex post* course scores, the students simply pick up courses with the highest interim scores. Our model of electives shopping is motivated by the existence of pre-registration systems (e.g. a 'shopping week' used at Harvard University), in which the students earn experiences with the instructor, the content of the course, and individual fit to the course. Nonetheless, even in the absence of a formal pre-registration system, students can learn about grading and the expected fit through their access to aggregate quantitative results and student-to-student comments for courses, and through online rating platforms such as ratemyprofessors.com.

Our approach is novel as it introduces a student's type only through a *relative* position among other students in the course. This general characterization of a type admits various microfoundations for students' choices of courses, including various benefit and effort cost functions, and informational and behavioural frictions. In what follows, we restrict fit to be only independent across the courses and invariant to grading;

TABLE 1
GRADES OF JD CANDIDATES AT PITT LAW SCHOOL, 2019

	Grade												
Cohort	A	A-	B+	B	B-	C+	C	C-	D	F	Mean	Median	Variance
Mid GPA	3.94	3.69	3.31	3.00	2.69	2.31	2.05	2.18	1.01	0.38			
Class of 2020	1	24	39	33	28	12	2	1	0	0	3.11	B	0.16
Class of 2019	3	26	55	36	14	5	0	0	0	0	3.18	B+	0.16
Class of 2018	6	29	55	30	15	1	0	0	0	0	3.33	B+	0.10

that is, a student's relative score on a course k is independent of his or her fit in any other course $j \neq k$ and also of the grading schemes in all courses.

In this environment, we characterize both supply side (offers) and demand side (enrolment). We demonstrate that the enrolment rates are *linear* in interim scores and with more flexibility fall proportionally; irrespective of fit, all students become more selective if grading competition intensifies. In addition, we document that increasing curriculum flexibility (i.e. more courses offered or fewer courses required) motivates course instructors to offer larger maximal scores to the best students. In the context of grading that involves an upper bound on scores (e.g. a maximum of 100), this translates into *grades compression* at the top. We can also describe the *ex post* scores that result from the interaction of the supply (interim scores) and demand (enrolments).

The gaps between excessively large scores given to the best students and very low scores given to average and below-average students are among the most unwelcome consequences of flexible curricula. We analyse how the mean-score ceiling can be combined with an extra grading policy in order to eliminate the excessive top scores. For this purpose, we consider adopting one of four top-score grading policies.

- *Top-score cap*: First, we suppose that the school directly caps the exam scores. Then, in enrolment to the policy, the instructors divide the students into two groups, namely attractive students (receiving the maximal score) and unattractive students (receiving low scores); this segregation occurs before the students self-select into the electives. In addition, adding this regulation risks assortative inefficiency associated with grades compression. Namely, compressed top scores are coarser and hence less informative about fit than non-compressed scores.
- *Quantile-score cap*: Second, the school is supposed to cap the number of grades that exceed a predetermined threshold grade level. This corresponds to the famous Princeton University grading policy from 2004 that no department gives more than 35% A-range grades. Stated equivalently, a given quantile of the grades is to be located below a given grade threshold (i.e. the 65-quantile is at B+ or worse). The mechanic effect of the quantile cap is to shift some grades below the threshold. We demonstrate that this shift additionally affects grades distributions below and above the threshold. Above the threshold, the maximal score decreases and thus the policy serves as an *indirect regulation* of the top scores. Below the threshold, the instructors begin to divide the students into attractive students (with compressed scores at the

TABLE 2
GRADE STATISTICS FOR JD CANDIDATES AT PITT LAW SCHOOL (2016, 2018, 2019)

	Mean 2016	Median 2016	Variance 2016	Mean 2018	Median 2018	Variance 2018	Mean 2019	Median 2019	Variance 2019
Class of 2021	—	—	—	—	—	—	3.11	B	0.16
Class of 2020	—	—	—	3.06	B	0.19	3.18	B+	0.16
Class of 2019	—	—	—	3.22	B+	0.13	3.33	B+	0.10
Class of 2018	3.08	B+	0.22	3.28	B+	0.12	—	—	—
Class of 2017	3.18	B+	0.12	—	—	—	—	—	—
Class of 2016	3.22	B+	0.11	—	—	—	—	—	—

threshold) and unattractive students (with low dispersed scores); as in the case of the top-score cap, intermediate scores cease to be offered.

- *Mean-score normalization*: Third, we let the school penalize the high scores by reparametrizing the mean-score ceiling. More precisely, we suppose that the school targets a constant range of the course scores; hence all curricula after normalization will award identical top scores. Interestingly, we demonstrate that such normalization *excessively* counteracts grading competition associated with flexible curricula. More precisely, the normalized score distributions are ranked by first-order stochastic dominance along the flexibility dimension. Thus after normalization, all types of students (with any percentile of fit) receive lower scores when the curriculum becomes more flexible. If the market keeps interpreting scores constantly, irrespective of a change in flexibility, then all students will unequivocally prefer to install a rigid curriculum.
- Ex post *mean ceiling*: The fourth instrument explicitly targets self-selection. We suppose that the mean-score ceiling is evaluated after the students self-select into the courses; in other words, we disregard that enrolled students are not comparable across courses. We find that this regulation does not motivate instructors to attract students with a better fit. As a result, there is an equilibrium where the instructors are discouraged to compete, and rather offer identical scores. Mismatches become prevalent, and positive sorting effects associated with self-selection tend to diminish.

Overall, the paper demonstrates that counteracting lenient grading while preserving the instructors' grading autonomy is a complex task. A number of trade-offs are associated with grading policies that target both mean scores and top scores. We have analysed four specific combinations of mean-score and top-score grading policies, and derived unintended consequences such as segregation, grades compression generating information losses and mismatches, and endogenous preferences for rigid curricula.

The paper proceeds as follows. Section I surveys the related literature. In Section II, we construct the setup. Section III analyses the baseline equilibrium distribution of scores associated with a given curriculum structure under a mean-score ceiling. Then we analyse four regimes in which a top-score grading policy complements the mean-score grading policy: Section IV analyses a cap on the maximal scores, while Section V analyses a quantile cap at an interior grades threshold. Section VI analyses normalization of scores to a fixed interval. Section VII shows that imposing the mean-score ceiling on *ex post* grades completely suppresses grading competition. In Section VIII we summarize and conclude. In Appendices A and B, we provide proofs and examples. Appendix C extends the baseline setup by the instructor's option to invest into resources.

I. LITERATURE

There is much evidence that grades affect students' choices of courses (Sabot and Wakeman-Linn 1991; Matos-Díaz 2012). In the first place, grades affect a student's subjective feelings of academic accomplishment (Boatright-Horowitz and Arruda 2013). Students select courses with better grades to lower their study time (Babcock 2010). In addition, pursuit of easy grades is a rational response to admissions committees that seem to not properly account for course difficulty when evaluating transcripts (Bailey *et al.* 2016). Also, in response to more restrictive dismissal policies, students are more likely to enrol in leniently-graded courses (Keng 2016). Therefore differences in grading mechanisms and incentives to construct lenient grading rules are of central importance to the economics of education. Understanding grading is especially important in the analysis of origins of widely documented grade inflation—whether the improved grades can be attributed to lenient grading, productivity growth or improved sorting.

How is the construction of grading rules addressed in the literature? Most papers look into centralized grading rules in the hands of school administrators. In the classic tradition, the principals maximize social welfare (or, specifically, human capital) subject to the students' best enrolments (Costrell 1994; Betts 1998). A typical problem is the optimal design of absolute or relative grading rules to elicit effort, accounting for characteristics of the environment such as the class size (Becker and Rosen 1992; Dubey and Geanakoplos 2010; Andreoni and Brownback 2017; Paredes 2017; Brownback 2018). In the construction of relative grading rules, a typical trade-off arises between the level of overall effort and an equitable distribution of effort among students, exactly as in other tournaments with heterogeneous contestants.

In more recent literature, the school uses the informative content of grades as a tool in competition for attractive placements. Typically, the pool of students is fixed and placement is in the form of assortative matching between students and firms (Ostrovsky and Schwarz 2010; Popov and Bernhardt 2013; Boleslavsky and Cotton 2015; Harbaugh and Rasmusen 2018). The idea is that a school observes a true ability, and constructs the grading rule as an optimal signal of the true ability. This literature examines the optimal signal technology employed by a school, and the resulting information content of transcripts.

In this paper, we adopt a decentralized perspective on grading that to the best of our knowledge is original in the economics of education. Instead of having full discretion over the grading rules, we suppose that the school constrains transcripts through a few regulations on the basic properties of score distributions, while the grading rules remain in the hands of individual instructors. The assumption of large discretion on the part of instructors is particularly relevant for universities that put a large value on the autonomy of academics. In such an environment, it is natural to analyse transcript structures resulting from the grading competition of non-cooperative instructors.

Our approach belongs in the family of General Lotto games that have originated from Colonel Blotto games (Hart 2008). In the game, for each 'battlefield' a player chooses a distribution of a non-negative random variable with a given expected value, which corresponds to the average 'allocation per battlefield'. The winner in the battlefield is a player with the highest realization. The Lotto games have been solved in the symmetric case of equal expectations by Myerson (1993), and in the general case of unequal expectations by Sahuguet and Persico (2006). Kovenock and Roberson (2021) generalize to heterogeneous valuations and asymmetric resources. Dziubiński (2012) covers the difference between continuous and discrete General Lotto games.

In this literature, typically only two players compete with each other, therefore only one prize per battlefield is allocated. In our setting, the battlefields are represented by a continuum of students, and we consider a more generalized case with multiple courses and multiple prizes depending on the number of courses into which each student has to enrol. We exploit the early results from Continuous General Lotto games, namely that the equilibrium scores distributions for K courses and M required courses are derived analogously to the equilibrium electoral portfolios of K candidates who compete over M votes of each voter, when an electoral portfolio is a distribution of a fixed cake (Myerson 1993). In addition, we borrow and extend the analysis of the caps that restrain from the above values of the random variables, known as Captain Lotto games (Hart 2016).

While most of the literature on the optimal signal technology (i.e. information design) disregards decentralized grading and grading competition, there is a small literature that discusses both provision of information and grading inflation resulting from grading competition. Bar *et al.* (2012) analyse grade disclosure policies as an alternative to grade rationing policies. Revealing the grades distribution to the instructors may elicit social pressure towards lenient graders and thus curb grading competition. Revealing the distribution to the readers of transcripts also reduces the values of inflated grades. In this paper, we abstract from both effects; informal grading norms and instructors' social pressure are absent, and the values of grades for the students are fixed.

II. Setup

Fundamentals

Curriculum There are $K \in \mathbb{N}$ elective courses, each with a single instructor (he). Each student (she) must enrol on exactly $M \in \mathbb{N}$ elective courses, where $1 \leq M \leq K$. A curriculum (M,K) is called more *flexible* than curriculum (M',K') if $M \leq M'$ and $K \geq K'$. If educational standards are measured by the number of courses needed for graduation (Lillard and DeCicca 2001), then flexible curricula imply low educational standards, and vice versa.

Electives shopping period In the shopping period, each student is allowed to costlessly visit any elective course. Through a visit, she obtains an interim score (an offer), which is a perfect signal of her *ex post* score. Hence if a student enrols in the course, then her *ex post* score is identical to her interim score. Our results will clearly extend to the case of noisy signals as long as the noises are independent and identically distributed variables.

Scores The interim (offered) score from a course $k = 1,\ldots,K$ is a realization x_k from a random variable X_k that will be described below. For analytical convenience, we will work with continuous scores, $x_k \in \mathbb{R}_0^+$.

Enrolment Students are pure score-shoppers and maximize their *ex post* scores. Given that the interim scores are perfect signals of the *ex post* scores, the students simply pick up M courses with the highest signal realizations x_k. We thus abstract from search or information costs; with these frictions to self-selection, instructors' grading competition is less pronounced.

Students Ex ante, students are identical; hence we will work with a representative student.[3] For each student and each course k, there is an individual fit to the subject matter of the course k, denoted φ_k. *Ex ante*, for any student and any course, the fit is uncertain, and is without loss of generality uniformly distributed on the unit interval,[4] $\varphi_k \in [0,1]$.

Instructors Each course instructor k chooses how to map the student's fit φ_k into a score. That is, the instructor selects a random variable X_k with a probability density function denoted $f_k(x)$ and cumulative distribution function denoted $F_k(x)$. When setting X_k, each instructor maximizes the expected number of students who enrol into his course.

Transforming fit into score The instructor's distribution function $F_k(x)$ transforms the individual fit to the course φ_k into the interim score x_k through a monotonic (non-decreasing) quantile function $F_k^{-1}(\varphi_k)$. As a result, since the individual fit is uniformly distributed, x_k is distributed by F_k. Monotonicity only reflects the usual incentive-compatibility condition by which high types can mimic low types but not otherwise.

Ex ante mean-score ceiling Each distribution F_k is required to have a constant mean, which we normalize to 1, $\int_0^\infty x f_k(x)\,\mathrm{d}x = 1$. Intuitively, the school requires that each instructor distributes a *fixed* amount of points to prospective students of the course. By this assumption, we suppose that the school compares the course on a level playing field, that is, for representative samples of students. (Later we provide a broader interpretation of the fixed resources and also analyse endogenous resources.) The instructors cannot compete for students by increasing the scores overall, but can change their distribution to encourage or discourage self-selection of particular types of students.

Efficiency We will call an equilibrium *efficient* if each student enrols into those courses that give her the highest values of φ_k with probability 1. For symmetric equilibria, $F_1(x, K, M) = \cdots = F_K(x, K, M)$, a sufficient condition for efficiency is that F is continuous in x (i.e. no mass points exist). In the absence of mass points, the inverse function F^{-1} is defined and thus there exists a one-to-one mapping between x and φ. Then each student who picks M courses with the highest values of x_k effectively picks up M courses with the highest values of φ_k. In contrast, if F is not continuous in x (hence mass points exist), then each mass point $\hat{x}$ is associated with an interval of φ. Hence a student who receives an excessive number of identical maximal offers has to eliminate some of the offers randomly without being able to rank them. By eliminating the offers randomly, efficiency is violated.

Enrolment rates Let $A_k \in \{0,1\}$ be the (random) indicator variable denoting the decision of a student to take the course k. We will calculate the probability of k being taken by a representative student conditional on her signal x_k. This conditional probability is a representative student's *enrolment rate*, and hence also the expected payoff of the instructor k. We will write the enrolment rate as

$$\pi_k(x_k) := \Pr(A_k = 1 | X_k = x_k).$$

A simple productivity perspective

To conveniently describe an instructor's optimization problem, we may use nomenclature from the classic production problem of a firm that maximizes output for given expenditures. In the case of grading competition, each offered score represents an input, and the students' enrolments are outputs. Precisely speaking, there is a continuum of heterogeneous inputs, indexed $x \geq 0$, where the cost of an x input is exactly x. Each instructor has total resources (budget) R, where in the case of exogenous resources (the mean ceiling grading policy), we normalize $R = 1$. A cumulative distribution function $F(x)$ denotes the cumulative number of inputs of index x or lower. If the c.d.f. is continuous, then $f(x)$ denotes the number of employed inputs of index x; if it is discontinuous, then we use Dirac measure at the relevant atom. The budget constraint is

$$\int x \, \mathrm{d}F(x) \leq R = 1.$$

The production function is additively separable and linear in inputs; in other words, the marginal product of each x input is independent on the amount of the inputs, and is denoted (x). The total output is

$$\Pi := \int \pi(x) \, \mathrm{d}F(x).$$

In a symmetric equilibrium, the total instructor's output is $\Pi = M/K$ as each student enrols into M classes offered by K instructors.

Marginal productivity of an input per unit of the cost is the ratio of the marginal product of the input to its cost, $r(x) := (x)/x$; in short, we will call it *productivity* of an input. In the absence of additional constraints, the optimal solution to the output maximization problem is that all inputs have equal marginal productivity per unit of the cost (in short, are equally productive), $r(x)=r$, $x \in [0,1]$. This *common productivity* level is given by the average productivity of resources,

$$r := \frac{\Pi}{R} = \frac{M}{K}\frac{1}{R}.$$

This expression illustrates that endowing all players with more resources (increasing R) decreases the average productivity of inputs. Also, for resources normalized at $R = 1$, we simply obtain $r = M/K$.

III. Equilibrium

This section derives the score distributions in a symmetric non-cooperative equilibrium. Recall that we effectively work with two distributions, interim (offered) scores and *ex post* (realized) score distributions, where the latter reflects the students' enrolments.

Interim scores

In this subsection, we will look into a symmetric equilibrium, denoted as a collection of identical and independent distributions of interim scores, $F(x,K,M) := F_1(x,K,M) = \cdots = F_K(x,K,M)$. The number of offered courses K and the number of required courses $M \leq K$ are two parameters of the distribution.

To understand the structure of the equilibrium, take a course k and a representative student. Recall that the student's *enrolment rate* conditional on an offer x_k, and hence also the expected payoff of the instructor k conditional on x_k, is $\pi_k(x_k) := \Pr(A_k = 1|X_k = x_k)$. The event $A_k = 1|X_k = x_k$ is characterized such that out of $K-1$ independent draws, at most $M-1$ courses are more attractive than the course k. The probability that a single course is more attractive than x_k is $1 - F(x_k)$. Given than each X_j is identically and independently distributed, the enrolment rate is then the value of the cumulative distribution function of a binomial distribution with parameters $M - 1$, $K - 1$ and $1 - F(x)$:

$$\pi_k(x_k) = \Pr(A_k = 1|X_k = x_k) = B(M-1, K-1, 1-F(x)).$$

We may characterize the event equivalently as at least $K - M$ realizations in the series $X_1,\ldots,X_{k-1},X_{k+1},\ldots,X_n$ to be below x_k. Hence the enrolment rate is the $(K - M)$*th-order statistic* of the series of size $K - 1$. Having the conditional probability of course k being taken (i.e. the enrolment rate), we may now proceed to the unconditional probability of course k being taken by a representative student, $\Pr(A_k = 1)$. This constitutes the expected payoff of the instructor k, denoted in line with our productivity perspective $\Pi_k(X_1,\ldots,X_K) := \Pr(A_k = 1)$. In a symmetric equilibrium, students spread equally across courses, and the instructor's expected payoff equals

$$\Pi_k(X_1,\ldots,X_k) = \Pr(A_k = 1) = \int_0^\infty B(M-1, K-1, 1-F(x))\,\mathrm{d}F_k(x) = \frac{M}{K} = r. \tag{1}$$

From the productivity perspective, the symmetric equilibrium simply means that all inputs are equally productive:

$$\frac{\pi(x)}{x} = r(x) = r = \frac{M}{K}.$$

Proposition 1 formally derives the equilibrium. (All proofs are relegated to Appendix A.) Notice that equal productivity is equivalent to the student's enrolment rate being *linear* in the random variable on its support, $(x) = rx$, exactly as observed in other contexts in which General Continuous Lotto games were applied (Myerson 1993; Hart 2008).

1. (Interim scores) Under an *ex ante* mean-score ceiling, the symmetric equilibrium interim scores distribution $F(x,K,M)$ is characterized as a unique solution of

$$B(M-1, K-1, 1-F(x,K,M)) = x\frac{M}{K}. \tag{2}$$

The equilibrium interim scores are derived only in an implicit form. For a low K or an extremely low or high M, a closed-form solution is readily available, which is illustrated for $K = 3$ in panel (a) of Figure 1. For a larger K and an intermediate M, the probability distribution is given by a Gauss hypergeometric function (see the first subsection of Appendix B).[5] From the continuity of the binomial function in the last argument, the distribution does not contain any mass point, hence the equilibrium is efficient.

We now examine the support of the interim scores. As the range of a binomial function is a unit interval, we have $xM/K = xr \in [0,1]$. In other words, the highest given score is inverse to the average productivity, $x \in [0,R/r] = [0,K/M]$. Intuitively, if more instructors compete for students, then instructors offer larger premia to the students with the best fit for the subject. Exactly the same effect occurs when the students have to collect fewer credits and therefore reject more offers. Corollary 1 to Proposition 1 summarizes this. In Appendix C, we also demonstrate that the result remains unchanged if the resources are not fixed and the advisors may increase their resources at a cost.

1. (Top scores) Curriculum flexibility increases top scores.

How do we interpret an increase in top scores in cases when the course scores are naturally bounded from above (e.g. by the value 100)? One way is to interpret the offered scores more broadly than just plain test scores; we may think of extra bonuses in the form of special services that the instructors provide to the best fitting students. These extra services may serve as the margin at which the most promising students are attracted to a class. Examples are promises to serve as thesis advisors, extra attention in classes and office hours, or explicit promises to write recommendation letters for graduate schools. As long as the instructors have an (*ex ante*) fixed supply of these extra services, the mean-score requirement is satisfied. Alternatively, interim scores exceeding natural bounds may be provided in the form of signals that effort needed for a good test score can be low, and thus effort can be saved. Finally, if the course scores indeed represent bounded tests scores, then we have an environment with a top-score cap that will be analysed in detail in Section IV.

Ex post scores

In this subsection, we derive the *ex post* scores distribution $G(x,K,M)$ and examine how the *ex post* scores depend on curriculum flexibility. The key for obtaining the *ex post* scores is that the equilibrium enrolment rate is linear in the offer, hence the *ex post* density is essentially the interim density multiplied by a linear function. Proposition 2 derives the implicitly characterized *ex post* scores.

2. (Ex post scores) Under an *ex ante* mean-score ceiling, the density of the *ex post* scores, denoted $g(x,K,M)$, is in a symmetric equilibrium

$$g(x,K,M) = xf(x,K,M). \tag{3}$$

Flexibility affects *ex post* scores through effects on the offers (instructors' side) and enrolments (students' side). The former channel is captured in the change of the interim scores distribution $F(x,K,M)$, where we have documented an increase in the top score (see

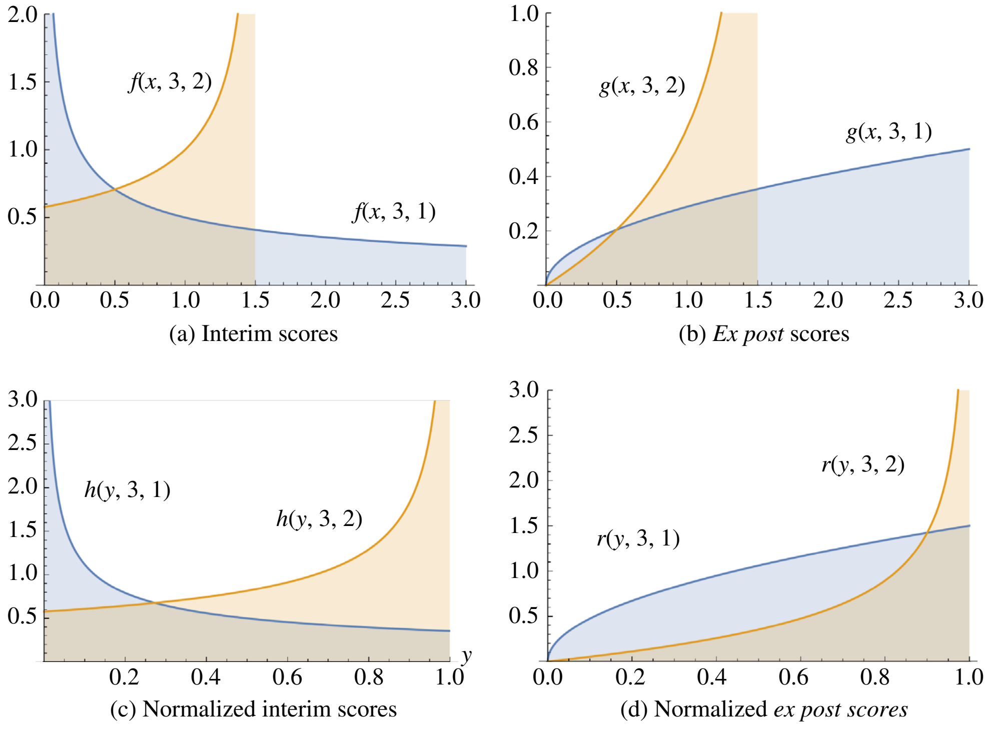

FIGURE 1. Scores densities for $M = 1{,}2$ and $K = 3$. [Colour figure can be viewed at wileyonlinelibrary.com]

Corollary 1). The latter is captured in the change of the enrolment rate (x). The slope of the enrolment rate is given by the average productivity of offers, $r = M/K$, which is decreasing with flexibility (a larger K and/or lower M). Hence with more flexibility, the enrolment rates tend to drop; moreover, in relative terms, the drop is identical for each score.

To give at least a limited perspective on the joint effect of changes in offers and enrolments, we may look at the effect on *the* ex post *mean*. Interestingly, the *ex post* mean exceeds the *ex ante* mean (equal to 1) exactly by the variance of offers,

$$\int_0^\infty x\,\mathrm{d}G(x,K,M) = \int_0^\infty x^2\,\mathrm{d}F(x,K,M) = 1 + \int_0^\infty (x-1)^2\,\mathrm{d}F(x,K,M).$$

In other words, the variance of the interim scores serves as a measure of self-selection into more attractive courses. Since this *ex ante* variance is increasing in flexibility (Myerson 1993, p. 861), and the interim mean is fixed by the mean-score constraint, the *ex post* mean is increasing if the curriculum becomes more flexible. We can interpret an increase in the average realized grades as *grades inflation*. That *ex post* scores increase with flexibility helps us to derive the students' *ex ante* preferences over flexibility in a case when the students are risk-neutral over the scores and market values of scores are exogenous to the curriculum structure. Then the students prefer flexible curricula as the joint effect on the interim distribution and the enrolment rates is to increase the expected *ex post* score.

To illustrate the differences between interim and *ex post* scores, panel (b) of Figure 1 plots the *ex post* distributions for $K = 3$; these *ex post* distributions are comparable with the corresponding interim distributions in panel (a). In addition, the example in the third subsection of Appendix B demonstrates that self-selection can be strong enough such that a positively skewed interim distribution changes into a negatively skewed *ex post* distribution.

IV. Top-score Caps

One of the key lessons of the previous section is that more competitive curricula motivate instructors to provide larger maximal scores. Excessive top scores can be regulated directly, namely by imposing a cap on scores. The top-score cap can be found in two environments. First, if we think of the test scores only, then the top-score cap is in the form of the maximal test score. Yet if the benefits to the students are broad and thus scores have no natural bound (e.g. if the instructor distributes not only test scores but also his own fixed resources such as a fixed amount of time for extra support), then the top-score cap is a limit on maximal total resources provided to a single student, for example, by norms that dictate how office time and attention is allocated to the students.

A Captain Lotto game

To analyse grading competition in the presence of caps that bind the values of the random variables from above, we borrow lessons from a bilateral Captain Lotto game in Hart (2016). For a special case of a bilateral competition, Hart (2016, Proofs of Prop. 1 and Th. 2) shows that the symmetric equilibrium strategy is a lottery over the (uniform) distribution from the unrestricted game truncated to an interval $[0,t]$ and a Dirac measure on the point of the cap c, where $0 < t < c$. This corresponds to our curriculum structure $(M,K)=(1,2)$.

To construct this strategy (i.e. an independent and identically distributed random variable X_k) intuitively, suppose that an instructor decides to use the distribution $F(x)$ from the unrestricted case. Then the probability mass $1 - F(c)$ remains unused. This mass is optimally shifted to the atom of $x = c$ where the enrolment rate is the highest; yet the resource constraint is now relaxed. To make it binding, additional probability mass from the interval $[0,c)$ can be shifted into $x = c$. In the equilibrium, the instructors shift the mass from the upper part of the interval, hence intermediate values on the interval (t, c) are no longer offered.

In our paper, we will observe that an analogical structure arises not only in the special case of bilateral competition, but also in any feasible curriculum structure (M,K). The main difference is that the underlying distribution from the unrestricted game, in our case denoted $F(x,K,M)$, is not necessarily uniform. Proposition 3 derives a unique threshold value for the empty interval and proves that the truncation described above indeed survives in the equilibrium.[6] For the analysis of grading systems, the key properties of grades under this policy are (i) grades compression at the top, and (ii) missing intermediate scores.

3. (Top cap) In the presence of an *ex ante* mean-score ceiling and a binding cap on the scores, $1 \leq c < K/M$, a symmetric equilibrium cumulative distribution of the offers is as follows:

$$\hat{F}(x,K,M) = \begin{cases} F(x,K,M) & \text{if } x \in [0,t], \\ F(t,K,M) & \text{if } x \in (t,c), \\ 1 & \text{if } x \geq c, \end{cases}$$

where the threshold $t \in [0,c)$ is given as a unique solution of

$$\frac{1-G(t)}{1-F(t)} = c. \tag{4}$$

In the equilibrium, notice that the productivity of each given offer is equal to the average productivity, $r(x) = r$; offers are now given from the restricted support of X, that is, $x \in [0,t] \cup \{c\}$. Thus enrolment rates at the given offers are exactly like in the absence of the top-score cap. In the equilibrium, the cap changes *the behaviour of instructors but not of students*.

For slightly higher offers, see that the productivity is now higher, since (x) stepwise increases at $x = c$ from $(c) = cr < 1$ to $(x) = 1$. Intuitively, a student would always accept such an offer since the offer would be the highest offer with probability 1. But given the existence of the cap, these offers are not available. In contrast, productivity at intermediate offers is lower, $r(x) < r$, since the enrolment rate is constant here, $(x) = (t)$ if $x \in (t,c)$, and therefore the productivity is decreasing, $r(x) = (t)/x < (t)/t = r$ if $x \in (t, c)$. Intuitively, an instructor who sets his offer in an empty intermediate interval knows that all other offers are situated below or above the interval. Hence increasing his offer within the interval does not change any student's decision and the enrolment rate is constant. This motivates the instructor to save on resources by moving the offer away from the empty interval.

Finally, see that imposing the top-score policy affects the shape of the distribution of realized scores, $\hat{G}(x,K,M)$, similarly to the distribution of offers. As the enrolment rates are invariant to the top-score policy, all effects of a top cap on the realized scores are due to the change in the structure of offers:

$$\hat{G}(x,K,M) = \begin{cases} G(x,K,M) & \text{if } x \in [0,t], \\ G(t,K,M) & \text{if } x \in (t,c), \\ 1 & \text{if } x \geq c. \end{cases}$$

Segregation and sorting effects

The cap segregates the students into two disjoint groups, *achievers* and *laggards*. With a more restrictive cap, what is the effect on the compositions of the two groups? First, with a more restrictive cap (a lower c), the empty interval begins with a lower threshold t and hence extra intermediate scores are no longer offered.[7] In other words, by capping the scores of the achievers, we indirectly also cap the maximal scores of the laggards.

The expected share of the laggards is $G(t)$. A more restrictive cap thus on average promotes some of the laggards among the achievers. With more achievers, however, we observe a more pronounced grades compression that results in a drop of the informative value of the top scores. Namely, as the most fitting courses are indistinguishable to the students, a sorting inefficiency occurs in expectation whenever a student receives at least $M+1$ top offers and consequently has to randomly eliminate one or more redundant top offers. Equivalently, the event exists when a representative student receives at most $K-(M-1)$ low offers. Thus a sorting inefficiency occurs in expectation with probability $B(K-M-1,K,F(t))$, which is the value of the cumulative distribution function of a binomial distribution with parameters $K-M-1$, K and $F(t)$. With a more restrictive cap, $F(t)$ drops as low offers are less likely, and the probability is increasing.

Higher educational standards

In the educational literature, educational standards are often measured by how restrictive a school is on average (Betts and Grogger 2003; Figlio and Lucas 2004). The question is: do raising educational standards have unequal effects at the top and bottom of the distribution of students? In our model, the overall restriction is captured by the level of resources for each instructor, that is, by the level of the mean ceiling. In the absence of the cap on the scores, higher standards (a more stringent mean-score restriction) only decrease received scores proportionally, hence inequality is not affected. However, if the scores are capped, then lowering the mean-score ceiling involves not only reducing resources for the instructors, but also increasing the 'effective' top-score cap since the environment has fewer resources but the nominal top-score cap has not changed.

The latter effect implies that different groups of students in a class are affected differently:

- top students are *not affected* at all as their score remains at the nominal top-score cap;
- intermediate students experience a *large drop* in scores as they move from the group of achievers into the group of laggards;
- low students' scores decrease *proportionally* to the decrease in the mean ceiling.

This result reveals an inequality-promoting channel that corresponds to observations that with higher educational standards, exogenously measured achievement rises more for students near the top of the achievement distribution rather than for the students near the bottom (Betts and Grogger 2003).

V. Quantile-score Cap

The quantile-score cap is a grading policy that does not restrict the level of the highest offer, but rather restricts a quantile of scores to not exceed a grade threshold. Equivalently, the policy requires a minimal share of students to be assessed below the grade threshold. Formally, a distribution function of offers $F_\varphi(x)$ satisfies a *quantile cap* of quantile φ at the threshold c if and only if[8]

$$F_\varphi(c) \geq \varphi.$$

Clearly, if the unrestricted offer distribution satisfies the quantile cap, $F(c) \geq \varphi$, then the quantile cap is not binding as equalization of productivity across inputs maximizes total output. The quantile cap is binding only if the number of the high offers (those strictly above c), $1-\varphi$, is constrained from above:

$$1-\varphi<1-F(c).$$

Bilateral competition

To illustrate how the quantile cap operates, we will first provide a solution to the simple case of bilateral competition, that is, $(M,K) = (1,2)$. We will illustrate that the quantile cap operates similarly to the top cap as it compresses scores at the cap and eliminates intermediate offers. This is not surprising since the top-score cap from Section IV is a special case of a quantile cap where $\varphi = 1$. In addition, the quantile cap equalizes resources (scores) across low (below threshold, indexed L) and high (above threshold, indexed H) groups of students.

For algebraic convenience, we consider a quantile cap at an average offer, c=1. For the cap to be binding, we let $\varphi > \frac{1}{2}$. The equilibrium distribution function of offers is

$$F^{\varphi}(x)=\begin{cases} x\cdot r_L & \text{if } 0\leq x\leq \left(\frac{1-\varphi}{\varphi}\right)^2, \\ \dfrac{\varphi(1-\varphi)^2}{\varphi^2+2\varphi-1} & \text{if } \left(\frac{1-\varphi}{\varphi}\right)^2\leq x<1, \\ x\cdot r_H & \text{if } 1\leq x\leq \dfrac{1}{\varphi}, \\ 1 & \text{if } x\geq \dfrac{1}{\varphi}, \end{cases}$$

where the input productivities are

$$r_L=\frac{\varphi^3}{\varphi^2+2\varphi-1}<\frac{1}{2}<\varphi=r_H.$$

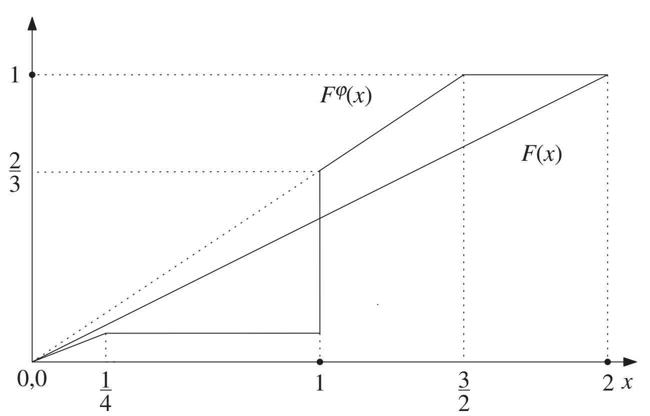

FIGURE 2. Offers in the unrestricted and restricted cases, $F(x)$ and $F^{\varphi}(x)$, for $\varphi = \frac{2}{3}$.

Figure 2 plots the restricted distribution $F^{\varphi}(x)$ along with the unrestricted distribution $F(x)$. A key observation is that the quantile cap equalizes resources. Formally, we may define redistribution in the dimensions of students and redistribution in the dimension of offers.

- To see redistribution in the students' dimension (from high-fit students to low-fit students), see that the initial division of total scores (resources) across the low-fit students of size φ and the high-fit students of size $1-\varphi$ was $(\varphi^2, 1-\varphi^2)$, whereas the new division is

$$\left(\frac{\varphi^2+2\varphi-1}{2\varphi}, \frac{1-\varphi^2}{2\varphi}\right).$$

- To see redistribution in the offers' dimension (from above-threshold offers towards below-average offers), the initial division of scores among below-average and above-average offers was $\left(\frac{1}{4}, \frac{3}{4}\right)$, whereas the new division of scores is identical to the division of scores across low-fit and high-fit students,

$$\left(\frac{\varphi^2+2\varphi-1}{2\varphi}, \frac{1-\varphi^2}{2\varphi}\right).$$

In order to see equalization of resources in both specifications, notice that

$$\frac{3}{4} > 1-\varphi^2 > \frac{1-\varphi^2}{2\varphi}.$$

Multilateral competition

Proposition 4 characterizes the symmetric equilibrium distribution of offers in a general case. In this distribution, both mean-score and quantile-score caps are binding. As in the special case of bilateral competition, high offers are now more productive than low offers, which redistributes resources from high offers to low offers and indirectly lowers the maximal scores.

4. (Quantile cap) For a quantile cap such that $\varphi > F(c)$, the equilibrium offers follow a cumulative distribution $F^{\varphi}(x)$ where

$$F^{\varphi}(x) = \begin{cases} F(\mu_L x) & \text{if } 0 \leq x \leq t, \\ F(\mu_L t) & \text{if } t \leq x < c, \\ F(\mu_H x) & \text{if } c \leq x \leq 1/r_H, \\ 1 & \text{if } x \geq 1/r_H, \end{cases}$$

where the input productivities are $(r_L, r_H) = (\mu_L r, \mu_H r)$, the productivity multipliers are

$$(\mu_L, \mu_H) = \left(\frac{1 - \mu_H R_H^{\varphi}}{1 - R_H^{\varphi}}, \frac{B(M-1, K-1, 1-\varphi)}{B(M-1, K-1, 1-F(c))} \right),$$

and the threshold $t \in [0, p)$ is given as a unique solution of

(5) $$\int_0^t x \, \mathrm{d}F(\mu_L x) + [\varphi - F(\mu_L t)]c = 1 - R_H^{\varphi}.$$

To sum up, a binding quantile cap affects low offers and high offers differently, as follows.

- With a binding quantile cap, the instructor is constrained in reallocating offers from low to high offers. This implies that the instructor prefers high to low offers, that is, high offers are more productive, $r_H > r_L$.
- As the number of available low offers is now excessively high for the disposable interval of scores, $[0, c]$, low offers are distributed on the interval as if an *artificial cap* binds the domain of the offers. This is equivalent to the existence of a top cap. As a consequence, intermediate offers disappear, and a gap between very low offers and offers at the quantile cap emerges.
- The overall number of high offers is now reduced. With higher productivity of high offers, the combined effect is that the top (highest) offer drops dramatically.

We observe two effects on equalization. On one hand, offers are *equalized* through (i) elimination of the highest offers and (ii) reduction in the frequency of the lowest offers. On the other hand, the quantile cap generates an artificial gap between very low offers and high offers, hence *segregating* the students into two disjoint groups.

VI. Normalized Mean-score Ceiling

An indirect way to regulate the excessive top scores is to normalize the top scores to the identical level. This is achieved by tightening the mean-score grading policy, that is, by lowering the ceiling for the mean. In our productivity perspective, this is equivalent to reducing resources available to the instructors on an increase in curriculum flexibility. Formally, in this section we suppose that the school maintains the mean-score policy but reduces the mean-score ceiling such that all scores, including the excessive scores, exactly fall into a predetermined range. How such *normalization* (or reparametrization) affects the interim and *ex post* scores is illustrated in panels (c) and (d) of Figure 1.

If reparametrization occurs *ex post* based on the observed distributions, then there is natural concern about the manipulability of the scheme by individual instructors. However, manipulability is negligible if a single instructor has a negligible effect on the inputs used for reparametrization. For instance, the reparametrization procedure may use the median values of the top scores of instructors instead of the maximum of top scores; then a single instructor's deviation from a symmetric equilibrium profile has no effect on reparametrization.

Normalized interim scores

In the first step, we focus on the normalized interim scores. The score distributions $H(y, K,M)$ are obtained trivially by normalizing the distribution $F(x,K,M)$ with a support $x \in [0,K/M]$ into a distribution $H(y,K,M)$ with a normalized support $y \in [0,1]$ such that $y = Mx/K$:

$$H(y,K,M) := F\left(y\frac{K}{M},K,M\right).$$

Our key result is that the normalized H-distributions can be ranked by first-order stochastic dominance with respect to K and M. That is, Proposition 5 shows that the score distribution for a less flexible curriculum first-order stochastically dominates the distribution with a more flexible curriculum.

5. *(Rigidity increases offers)* Let (M',K') be a less flexible curriculum than (M,K). Then for any $y \in (0,1)$,

$$H(y,K',M') < H(y,K,M).$$

This dominance property can be stated alternatively as the property of the quantile (inverse distribution) function to H-distribution function. That is, for any percentile $p \in [0,1]$, let $Y(p,K,M)$ be the quantile in the normalized distribution satisfying $H(Y(p,K,M),K,M) = p$. When inserted into the implicit solution in equation (2), the quantile is

$$Y(p,K,M) = B(M-1,K-1,1-p).$$

Exploiting well-known properties of the binomial cumulative distribution function, we observe that for any percentile p, an increase in flexibility decreases the respective quantile $Y(p,K,M)$. We demonstrate for an increase from K to $K + 1$ and for a decrease from M to $M - 1$:

$$\begin{aligned} Y(p,K,M) - Y(p,K+1,M) &= B(M,K,p) - B(M,K-1,p) < 0, \\ Y(p,K,M-1) - Y(p,K,M) &= B(M-1,K,p) - B(M,K,p) < 0. \end{aligned}$$

In other words, the H-distribution function moves to the left with an increase in flexibility. As the H-distribution function is non-decreasing in y, this implies that it moves up with flexibility.

Normalized ex post scores

Importantly, the first-order dominance relation in Proposition 5 extends to the *ex post* (realized) scores. We know that the distribution of *ex post* scores $G(x)$ is such that $g(x) = xf(x)$; recall that $G(x,K,M)$ denotes the distribution parametrized by K and M. As for the interim scores, this distribution characterizes a cumulative distribution of normalized scores, denoted $R(y,K,M)$, and its probability density function, $r(y,K,M)$, again such that $y = xM/K$:

$$R(y,K,M) := G\left(y\frac{K}{M},K,M\right).$$

6. *(Rigidity increases scores)* Let (M',K') be a less flexible curriculum than (M,K). Then for any $y \in (0,1)$,

$$R(y,K',M') < R(y,K,M).$$

By Proposition 6, *any percentile* of the distribution $R(y,K,M)$ is decreasing in flexibility. For example, if $\tilde{y}(K,M)$ denotes the median percentile, $R(\tilde{y}(K,M),K,M) = \frac{1}{2}$, then $\tilde{y}(K,M)$ is decreasing in K and increasing in M. In contrast, the second subsection of Appendix B shows that the median of the initial (non-normalized) distribution $G(x,K,M)$, denoted $\tilde{x}(K,M)$, may decrease but also increase in flexibility.

Students' preferences over normalized curricula

The dominance property has clear implications for students' preferences over curricula. If the students' valuations of realized scores are exogenous to the curriculum structure, then having a more flexible normalized curriculum implies lower scores. Given the dominance property, this effect holds *ex post for all types* of students, irrespective of the levels of fit. This reverts the observation from the baseline case in which students with linear valuations always benefit from more flexible curricula; here, the result generalizes beyond the shape of the valuation as the preference holds both *ex ante* and *ex post*.

The assumption that the valuations of scores are exogenous is obviously simplifying as firms in the long term account for changes in score distribution, hence the market valuation at least partly offsets the normalization. Still, grading rules are often imperfectly observable (Zubrickas 2015). Also, there is evidence that students perceive as costly reparametrizations such as decreases in the grade ceiling. For example, Gorry (2017) observes a drop in the students' satisfaction, manifested in worse teaching evaluations, after a grade ceiling dropped.

So far, we have assumed that a student who assesses various curricula maximizes the received scores (or the average of the received scores if the curricula differ in the number of course requirements). From that perspective, a high 'fit' to a course is just instrumental to receiving a high score. In addition to the market valuation of the high scores, we may consider broader benefits of a high fit; for instance, a student consumes higher intrinsic benefits, or exerts less effort if the fit is high. These extra benefits imply that the average realized fit is part of the student's objective. As the average realized fit increases with a larger K (a richer menu) and a lower M (higher selection), the existence of these extra benefits sways the preferences back towards curriculum flexibility.

VII. Ex Post Mean-score Ceiling

Among potentially unwelcome consequences of curriculum flexibility is the pronounced self-selection of the students; we know that with more flexibility, all students decrease their enrolment rates, and the drop (measured proportionally) is identical at each offer. Combined with the effect on the shape of the offers distribution, we have observed grades inflation, defined as an increased wedge between the mean *ex post* score and the (fixed) mean interim score. In this section we analyse a policy through which a school, concerned about the score-inflating effects of self-selection, fixes the mean *ex post*. In other words, the mean score is now evaluated *ex post*.

When the mean-score ceiling is evaluated *ex post*, the instructors need to attract students with a low fit to be able to award high scores to students with a high fit. We will find that this constraint effectively prohibits competition over the enrolments. As a result, scores are compressed at the mean, and all benefits associated with information on the individual fit (benefits of sorting) are wiped out. Specifically, Proposition 7 demonstrates that a symmetric equilibrium in pure strategies (i.e. an identical exam score offered to all electives shoppers) is an equilibrium with the *ex post* mean constraint. For the sake of comparison, we also show that this degenerated scores distribution is not an equilibrium with the *ex ante* mean constraint. We will obtain the result clearly in a discrete setting, for any support of scores that is sufficiently fine-grained. A continuous setting that we have applied to date represents a limit case of the analysis.

Equilibrium

Formally, an exam score is a non-negative integer $x \in \mathbb{N}_0$. The *ex ante* constraint in a discrete setting writes $\sum_{x \in \mathbb{N}_0} x f(x) \leq z$; the *ex post* constraint writes $\sum_{x \in \mathbb{N}_0} x g(x) \leq z$. We let the required *ex post* mean score $z \in \mathbb{N}$ be very large, $z \gg 0$, so that the support grid is fine enough. Next, let F^z be a degenerate cumulative distribution function that offers an identical mean score z to all students, that is, $F^z(x) = 0$ if $x < z$, and $F^z(x) = 1$ otherwise. Proposition 7 checks whether this profile is an equilibrium profile under the *ex post* constraint and the *ex ante* constraint.

7. *(Lack of competition)*Awarding identical scores, $(F_1,\ldots,F_K) = (F^z,\ldots,F^z)$, is an equilibrium strategy profile under the *ex post* mean-score ceiling but not under the *ex ante* mean-score ceiling.

Intuitively, when instructors offer identical scores and thus avoid competing with each other, students spread equally. The enrolment rate is a stepwise function, where $(x) = 0$ if $x < z$, $(z) = rz \in (0,1)$, and $(x) = 1$ if $x > z$. Hence a minor deviation from an above-average offer greatly boosts class size (enrolment). The instructors consider exploiting the stepwise increase in the enrolment rate. To meet the *ex ante* mean constraint, the instructor would just offer to the low-fit students a very low offer. For the *ex post* constraint, however, this compensation is not feasible as low offers are never accepted, and therefore cannot enter the *ex post* scores distribution; only offers $x \geq z$ are in the support of the realized scores. Therefore the instructor is locked to the average offer z; he cannot propose larger scores without violating the *ex post* mean score cap.

A summary of grading policies

Finally, to provide a summary perspective on the grading policies, Table 3 lists the effects of the four grading policies on the offered scores. The effects are presented as changes in the inverse function $F^{-1}(\varphi)$ evaluated at various levels of the individual fit. In addition to

TABLE 3
EFFECTS OF GRADING POLICIES ON THE OFFERED SCORES

Policy	Fit			
	Very low	Low	High	Very high
Top-score cap	0	0	+, compressed	−, compressed
Quantile-score cap	+	+, compressed	−	−, compressed
Normalized mean score[a]	−	−	−	−
Ex post mean score	+, compressed	+, compressed	−, compressed	−, compressed

Notes
[a]By definition, reductions in the mean score apply to flexible curricula, and increases apply to rigid curricula. As we are interested in addressing the excessive top scores, we present only the case of flexible curricula.

giving the signs of the effects, we also show whether the scores are compressed, hence whether the information becomes coarser.

VIII. CONCLUSIONS

High educational standards can be maintained and enforced only through appropriate grading practices. However, grading in higher education institutions is to a great extent the responsibility of individual instructors, and their objectives may differ from the maintenance of standards. For elective courses in particular, instructors are motivated to engage in grading competition whenever class size and students' satisfaction serve as the instructors' performance indicators.

To curb lenient grading associated with grading competition, schools tend to adopt various grading policies, among which the most widely analysed is a constraint on the *average grades*. In this paper, we theoretically analyse how this grading policy, combined with other grading policies, may influence decentralized grading. Our framework focuses on sharply characterizing the effects of grading policies on the *distribution of grades* in a symmetric setting. Like Chan *et al*. (2007), we abstract from the effects on effort (cf. Becker and Rosen 1992; Dubey and Geanakoplos 2010; Andreoni and Brownback 2017; Paredes 2017; Brownback 2018; Olszewski and Siegel 2018), but our setting is consistent with any microfoundation of effort that preserves the outcome rank of a student in a class (i.e. the student's relative position) being independent of the grading scheme.

Technically, we apply Continuous General Lotto games into a setting with *ex ante* homogeneous students, assuming that the school regulates the mean course score. In addition, we suppose that the school compares the courses on a level playing field by calculating the mean for a representative sample of students and not for a self-selected group of students. Our first task is to investigate the effects of having a more flexible curriculum (more listed courses, fewer required courses). With a more flexible curriculum, the gap in scores widens; the best students will receive higher scores, whereas worse students will receive lower scores. In other words, grading competition intensifies and inequality increases.

The very large scores awarded to the best students and the score inequality are two major concerns associated with curriculum flexibility and grading competition. Our second task is to analyse four scenarios of how the incentive to provide excessive scores is counteracted while the mean-score grading policy is preserved. All scenarios illustrate that the combined regulation of both mean scores and top scores comes at a cost. By imposing caps, scores get compressed at the caps, and the capped scores become less informative for both students and employers. Moreover, the students are separated into two disjoint groups; a new kind of inequality emerges in the class. The observation on the segregation of the students and grades compression at the cap is generalized to any binding quantile cap. Next, if the mean grade ceiling is lowered for more flexible curricula, then all types of students receive lower scores and thus will demand rigid curricula even at the cost of lower variety. Finally, by imposing the mean-score policy on the self-selected students (i.e. by disregarding selection effects), favourable self-selection into more fitting courses is entirely eliminated.

In terms of policy consequences, we illustrate that addressing grade inflation through grading policies generates nuanced effects on scores and grade distribution. In particular, we observe unintended consequences of grading policies in terms of *segregation* (division of the class into separate groups of achievers and laggards), *rigidity* (induced support of rigid curricula that more often give relatively high scores), *grades compressions* at caps (both at quantile-score caps and a top-score cap as its special case), and *assortative inefficiency* (inefficient self-selection of students into courses). In particular, our analysis finds a rationale for an inequality-promoting channel of more stringent grading standards. In the literature, grading standards are mostly measured by how restrictive a school is on average (Betts and Grogger 2003; Figlio and Lucas 2004). Here, this restriction is captured by the level of the mean ceiling (instructors' resources). In the absence of the cap on the scores, a more stringent mean-score restriction decreases scores proportionally, hence inequality is not affected. Yet if the scores are capped from above, then drops at the top are less pronounced than drops at other quantiles, and inequality in scores grows. This result corresponds to observations that exogenously measured achievement rises more for students near the top of the achievement distribution than for the students near the bottom (Betts and Grogger 2003).

Our policy effects can be summarized in three broad lessons. (i) Grades get compressed when caps on scores or quantiles are imposed, and when the instructors lack incentives to compete. As a consequence, grades compression reduces information provided to the students and generates assortative inefficiencies. (ii) Score or quantile caps segregate the students into disjoint groups of achievers and laggards. (iii) Providing excessively high grades can be punished by reducing 'rewards' that the instructor allocates to the students (e.g. in the form of a more stringent mean-score requirement or in the form of fewer credits). Nevertheless, to set the penalty correctly is difficult; an excessive penalty, as in the case of the mean-score normalization, biases the students towards rigid curricula.

Our results have been derived for *ex ante* homogeneous students with independent course valuations, but can be extended to correlated course valuations. Negative correlation channels grading competition into clusters of courses that are close substitutes, and hence is equivalent to having a more rigid curriculum. Positive correlation allows the instructors to assign lower grades to below-average students and offer higher grades to above-average students; grading competition intensifies as if the curriculum became more flexible. Additionally, our approach suggests that admissible signal technologies of the schools may be limited by the existence of decentralized grading practices; signal of ability has a specific structure, where the structure depends on the grading policies as well as curriculum structure. Thus the literature on the optimal signal technology in the context of educational markets (Ostrovsky and Schwarz 2010; Popov and Bernhardt 2013; Boleslavsky and Cotton 2015; Harbaugh and Rasmusen 2018) may benefit from incorporating lessons from decentralized grading explicitly into the analysis.

APPENDIX A: PROOFS

PROOF OF PROPOSITION 1

Any admissible $F_k(x)$ is restrained by the mean-ceiling policy that normalizes the average offer to be exactly 1. We multiply this constraint by M/K, and apply the equality with the instructor's expected payoff in equation (1):

$$\int_0^\infty x\frac{M}{K}\,\mathrm{d}F_k(x)=\frac{M}{K}=\int_0^\infty B(M-1,K-1,1-F(x))\,\mathrm{d}F_k(x). \tag{A1}$$

By comparing the outer expressions of equations (A1), the only solution that yields $B(M-1,K-1,1-F(x))$ linear in x (as in Myerson (1993) or Hart (2008)) is $B(M-1,K-1,1-F(x,K,M))=xM/K$. This corresponds to the distribution in equation (2).

It is immediate to see that this distribution is indeed the equilibrium solution. From the linearity of the conditional probability, notice that $B(M-1,K-1,1-F(x,K,M))=xM/K$ if $x\in[0,K/M]$, and $B(M-1,K-1,1-F(x,K,M))=1<xM/K$ if $x>K/M$. Therefore if the instructor deviates to another $F_k(x)$ that complies with the mean-ceiling restriction, $\int_0^\infty x\,\mathrm{d}F_k(x)=1$, then his expected payoff cannot increase:

$$\begin{aligned}\Pi_k(X_1,\ldots,X_K)&=\int_0^\infty B(M-1,K-1,1-F(x))\,\mathrm{d}F_k(x)\\&\le\int_0^\infty x\frac{M}{K}\,\mathrm{d}F_k(x)=\frac{M}{K}.\end{aligned} \tag{A2}$$

PROOF OF PROPOSITION 2

To obtain the distribution of *ex post* scores, denoted $G(x,K,M)$, recall that $A_k\in\{0,1\}$ is the random indicator variable denoting the decision to take the course k. The probability of taking the course conditional on $X_k=x$ is

$$\Pr(A_k=1|X_k=x)=B(M-1,K-1,1-F(x)).$$

The joint probability is

$$\Pr((A_k=1)\cap(X_k=x))=B(M-1,K-1,1-F(x))f(x).$$

We use the fact that the marginal density of A_k=1 is the expected payoff,

$$\Pr(A_k=1)=\Pi_k(X_1,\ldots,X_K)=\frac{M}{K}.$$

Then the conditional probability of $X_k{=}x$ on $A_k{=}1$ is

$$\begin{aligned} g(x) = \Pr(X_k = x | A_k = 1) &= \frac{\Pr((A_k = 1) \cap (X_k = x))}{\Pr(A_k = 1)} \\ &= B(M-1, K-1, 1-F(x)) f(x) \frac{K}{M} \\ &= x f(x). \end{aligned}$$

PROOF OF PROPOSITION 3

First, to see that t is a unique solution, we rewrite the condition in equation (4) as

$$\text{(A3)} \qquad \int_0^t x \,\mathrm{d}F(x) + (1 - F(t))c = 1.$$

The left-hand side of this equation is decreasing in t for $t < c$ (i.e. its derivative is $f(t)(t - c) < 0$), while the right-hand side is constant in t. Notice also that the distribution is defined only for $c \geq 1$; if $c < 1$, then we do not have a single crossing as the left-hand side evaluated at $t{=}0$ is less than the right-hand side. For $c < 1$, the equilibrium distribution is degenerate: $\hat{F}(x, K, M) = 0$ if $x < c$, and $\hat{F}(x, K, M) = 1$ if $x \geq c$.

Second, the resource constraint is satisfied, because the $\hat{F}$ distribution is a lottery over (μ',c) with probabilities $(F(t), 1 - F(t))$, where μ' is a mean of the truncated F-distribution,

$$\mu' = \frac{1}{F(t)} \int_0^t x f(x) \,\mathrm{d}x.$$

Then the mean of the $\hat{F}$ distribution is exactly equivalent to the left-hand side of equation (A3), which yields that the resource constraint is just met:

$$F(t)\mu' + (1 - F(t))c = \int_0^t x f(x) \,\mathrm{d}x + (1 - F(t))c = 1.$$

Third, consider any deviation of an instructor k. Notice immediately that if a deviation imposes positive probability mass on any $x_k \in (t, c)$, then there exists a more profitable deviation that (i) shifts the probability to $x'_k \in (t, x_k)$, and (ii) exploits the relaxed constraint such that it shifts the probability mass from some low x to a higher x. Simply, all shifts in case (i) will be exploited as such shifts relax the constraint without affecting the expected payoff Π_k, given that $_k$ is flat for the interval $x \in (t,c)$. Therefore we will consider only deviations to distributions that impose zero probability on (t,c).

We insert the resource constraint of F_k,

$$\int_0^t x f_k(x) \,\mathrm{d}x + (1 - F_k(t))c \leq 1,$$

into the expected payoff on a deviation to F_k:

$$\begin{aligned}\Pi_k &= \int_0^t B(M-1,K-1,1-F(x))f_k(x)\,\mathrm{d}x + [1-F_k(t)]\,B(M-1,K-1,1-F(c))\\ &= \int_0^t x\frac{M}{K}f_k(x)\,\mathrm{d}x + [1-F_k(t)]\,c\frac{M}{K}\\ &\leq [1-(1-F_k(t))c]\frac{M}{K} + [1-F_k(t)]\,c\frac{M}{K} = \frac{M}{K}.\end{aligned}$$

Hence the deviation does not pay off, $\Pi_k \leq M/K$.

PROOF OF PROPOSITION 4

We will construct the proof in the following steps. In the restricted case, productivity at x is denoted $r^{\varphi}(x)$.

- High offers. When considering the optimal distribution of the offers above the threshold, $x > c$, the instructor is restricted not by the quantile cap but only by the mean constraint. Hence each high offer has an identical productivity $r^{\varphi}(x) = r_H$.
- Given the absence of a gap in scores at high offers, and given equalization of productivities, we can obtain the value of r_H by evaluation of the limit of the input productivity as $x \to c^+$:

$$\begin{aligned} r_H = \lim_{x\to c^+} r^{\varphi}(x) = \lim_{x\to c^+}\frac{\pi(x)}{x} &= \lim_{x\to c^+}\frac{B\big(M-1,K-1,1-F_{\varphi}(x)\big)}{x}\\ &= \frac{B(M-1,K-1,1-\varphi)}{c},\end{aligned}$$

where we have used that for a binding quantile cap, $\lim_{x\to c^+}F_{\varphi}(x) = \varphi$.

- Now, we will express only the high productivity relative to the common productivity of all inputs in the unrestricted case. The common productivity r is for convenience evaluated at $x = c$:

$$r := r(c) = \frac{B(M-1,K-1,1-F(c))}{c} = \frac{M}{K}.$$

We will introduce the productivity *multiplier* as

$$\mu_H := \frac{B(M-1,K-1,1-\varphi)}{B(M-1,K-1,1-F(c))} > 1.$$

Hence $r_H = \mu_{H} r > r$. Notice that a stepwise increase in productivity at $x = c$, namely $r_H = \lim_{x\to c^-} r^{\varphi}(x) > r^{\varphi}(c) = r_L$, is given by the fact that ties cease to exist beyond the quantile cap $x = c$.

- For high offers where mass points do not exist, the distribution of offers is given implicitly as

$$\pi_H^{\varphi}(x) := B(M-1,K-1,1-F^{\varphi}(x)) = r_H x = \mu_H r x. \tag{A4}$$

We can now express $F^{\varphi}(x)$ by reference to the initial distribution $F(x)$. To avoid confusion, let the argument of the initial distribution (i.e. the score) be z:

$$B(M-1,K-1,1-F(z))=rz.$$

If we substitute $z=\mu_{Hx}$, then the implicit characterization is now $B(M-1,K-1,1-F(\mu_H x))=\mu_H rx$. By directly comparing with equation (A4), we receive that at high offers, $x \geq c$,

$$F^{\varphi}(x)=F(\mu_H x).$$

- The top offer $\bar{x}$ is at $F^{\varphi}(\bar{x})=1$, where the enrolment rate is 1, $\pi(\bar{x})=B(M-1,K-1,0)=1$, hence $\bar{x}=1/r_H$. Clearly, the top offer drops relative to the unrestricted case in which the highest offer was $1/r$:

$$\bar{x}=\frac{1}{r_H}=\frac{1}{\mu_H r}<\frac{1}{r}.$$

In other words, introduction of the binding quantile cap at the offer $c<1/\mu_{Hr}$ led to the elimination of offers in the interval $(1/\mu_{Hr},1/r]$.

- What are resources? We can prove that resources are redistributed to low offers. Let $(R_L,R_H)=(1-R_H,R_H)$ be the initial division of resources, where

$$R_H=\int_c^{1/r} zf(z)\,\mathrm{d}z.$$

We use that in the new division of resources,

$$R_H^{\varphi}=\int_c^{1/r_H} xf(x)\,\mathrm{d}x=\int_c^{1/r_H} zf(z)\,\mathrm{d}x=\frac{1}{\mu_H}\int_{\mu_H c}^{1/r} zf(z)\,\mathrm{d}z<R_H.$$

We can interpret the drop in resources by reference to the original division of resources. (i) First, resources corresponding to the original interval $[c,\mu_H c]$ are not included in R_H^{φ}. (ii) Second, the domain of high offers is compressed μ_H times, which, *ceteris paribus*, reduces the resources μ_H times.

- Now we proceed with the low offers. We divide the total output into output from low and high offers:

$$\Pi=\Pi_L+\Pi_H=r_L R_L^{\varphi}+r_H R_H^{\varphi}.$$

For exogenous resources, we use $\Pi=rR=r$ and also $R_L^{\varphi}+R_H^{\varphi}=R=1$ to obtain $r_L=\mu_{Lr}$, where the productivity multiplier for low offers is

$$\mu_L:=\frac{1-\mu_H R_H^{\varphi}}{1-R_H^{\varphi}}<1.$$

- Low offers. Having the input productivity r_L, we can again express $F^{\varphi}(x)$ by reference to the initial distribution $F(x)$. We proceed as for high offers, realizing that for low offers, the relevant productivity multiplier is μ_L. Hence at low offers,

$$F^{\varphi}(x) = F(\mu_L x).$$

- Empty interval. The threshold for low offers is again determined such that all resources for low offers, $1 - R_H^{\varphi}$, are spent. Specifically, the offers are split into two groups: $F(\mu_{Lt})$ of very low offers at $x \in [0,t]$, and $\varphi - F(\mu_{Lt})$ of top-cap offers at $x = c$. Since the left-hand side of equation (5) is decreasing in t and the right-hand side is constant in t, there is a unique value of the threshold.

PROOF OF PROPOSITION 5

To start with, recall that $F(x,K,M)$ is the distribution parametrized by K and M. This distribution characterizes a normalized cumulative distribution $H(y,K,M)$ (and its probability density function $h(y,K,M)$) such that $y = xM/K$, where

$$H(y,K,M) = F\left(y\frac{K}{M}, K, M\right).$$

Notice that $F(x,K,M)$ has a support $[0,K/M]$, whereas any normalized distribution $H(y,K,M)$ has a support $[0,1]$.

In the first step, we exploit the implicit form of the equilibrium cumulative distribution. Throughout the paper, we will slightly abuse the problem by switching from discrete K to continuous K; clearly, the switch is irrelevant as long as the first derivatives are strictly monotonic. Namely, by differentiating $B(m,n,p) := B(M-1, K-1, 1-H(y,K,M)) = y$ with respect to K (i.e. holding M as well as the normalized score y constant), we obtain

$$B_n(m,n,p) - B_p(m,n,p)\,H_K(y,K,M) = 0.$$

Notice that we use subscripts for partial derivatives. We translate the monotonic property of the discrete binomial distribution $B(m,n+1,p) - B(m,n,p) < 0$ into its continuous form, $B_n(m,n,p) < 0$. Using also that $B_p(m,n,p) < 0$ holds, we obtain that the value of the H distribution is increasing in K:

$$H_K(y,K,M) > 0.$$

In exactly the same way, we obtain $H_M(y,K,M) < 0$. These properties are sufficient and necessary for a comparison of distributions in terms of the first-order stochastic dominance.

PROOF OF PROPOSITION 6

We concentrate only on properties of the distribution of normalized scores $R(y)$. To prove that $R(y,K,M)$ is monotonic in K and M, we proceed in three steps.

First, we express the density of the *ex ante* normalized scores $h(y)$:

$$h(y,K,M) = \frac{\mathrm{d}F(yK/M,K,M)}{\mathrm{d}y} = \frac{\mathrm{d}F(yK/M,K,M)}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}y} = f(yK/M,K,M)\,\frac{K}{M}.$$

We enter this into the density $r(y,K,M)$:

$$r(y,K,M)=\frac{\mathrm{d}G(yK/M,K,M)}{\mathrm{d}y}=\frac{\mathrm{d}G(yK/M,K,M)}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}y}$$
$$=g(yK/M)\frac{K}{M}=f(yK/M,K,M)\,y\frac{K^2}{M^2}=h(y,K,M)\,y\frac{K}{M}.$$

Second, we express how the density changes in K:

$$r_K(y,K,M)=\frac{y}{M}h(y)+\frac{yK}{M}h_K(y,K,M).$$

The first component is positive. We prove that the second component is also positive. That is, we apply Leibniz's rule on the *ex ante* scores and our observation on the marginal effect of K on the density $h(y,K,M)$:

$$\int_0^{\hat{y}}h_K(y,\hat{K},M)\,\mathrm{d}y=\frac{\partial\int_0^{\hat{y}}h(y,\hat{K},M)\,\mathrm{d}y}{\partial K}=H_K(\hat{y},\hat{K},M)>0.$$

Since $H_K(\hat{y},\hat{K},M)>0$ for any $(\hat{y},\hat{K})$, we have $h_K(y,K,M)>0$ for all $(\hat{y},\hat{K},M)$. As a consequence, also $r_K(y,K,M)>0$ for any K. Finally, we impose Leibniz's rule on the *ex post* scores to observe

$$R_K(\hat{y},\hat{K},M)=\frac{\partial\int_0^{\hat{y}}r(y,\hat{K},M)\,\mathrm{d}y}{\partial K}=\int_0^{\hat{y}}r_K(y,\hat{K},M)\mathrm{d}y>0.$$

We obtain $R_M(\hat{y},\hat{K},M)<0$ by analogy.

PROOF OF PROPOSITION 7

Take a course k. If this course offers a realization $X_k=x$ to a representative student and all other courses stick to F^z, then the enrolment rate (the event is denoted $A_k\in\{0,1\}$) is

$$\pi(x):=\Pr(A_k=1|x)=\begin{cases}0 & \text{if } x<z,\\ M/K & \text{if } x=z,\\ 1 & \text{if } x>z.\end{cases}$$

The proof now analyses deviations under each of the two constraints separately. For each deviation, we use that the probabilities sum to 1, and also that the respective constraint must be met.

Ex ante mean constraint We show that the instructor of course k can improve his payoff if the support is fine-grained enough. The idea is close to concavification in Kamenica and Gentzkow (2011); namely, if we construct a concave closure of (x), denoted as $c(x)$, then the instructor can obtain the payoff $c(z)$, where $\lim_{z\to\infty}c(z)=1$. In other words, for a sufficiently fine-grained support, the instructor can deviate such that almost all students attending the class in the shopping period will eventually enrol in the class.

The deviation will have the following form. The instructor will redistribute some probability mass to points $x=0$ (low offer) and $x=z+1$ (high offer), $\mathrm{d}f_k(0)>0$ and $\mathrm{d}f_k(z+1)>0$. We suppose the *ex ante* mean constraint is met,

$$f_k(0)\,0+(1-f_k(0)-f_k(z+1))\,z+f_k(z+1)\,(z+1)=z,$$

which implies

$$f_k(0)=\frac{f_k(z+1)}{z}.$$

The marginal effect on the payoff is

$$\frac{\mathrm{d}\Pi_k}{\mathrm{d}f_k(z+1)}=\frac{\mathrm{d}f_k(0)}{\mathrm{d}f_k(z+1)}\left(-\frac{M}{K}\right)+1-\frac{M}{K}=\frac{1}{z}\left(-\frac{M}{K}\right)+1-\frac{M}{K}.$$

If the support becomes fine-grained enough—that is, for a sufficiently large z—then the payoff increases with the redistribution:

$$\lim_{z\to\infty}\frac{\mathrm{d}\Pi_k}{\mathrm{d}f_k(z+1)}=1-\frac{M}{K}>0.$$

Ex post mean constraint First, we will characterize any $F_k(x)$ by two key probabilities, $f_k(z)$ and $f_k(z^+) := \sum_{x=z+1}^{\infty} f_k(x)$. If all other courses stick to F^z, then the instructor's payoff is

$$\Pi_k := \Pr(A_k=1)=f_k(z)\,\pi(z)+f_k(z^+)\,\pi(z+1)=f_k(z)\,\frac{M}{K}+f_k(z^+).$$

In the symmetric profile, $f_k(z) = 1$ and $f_k(z^+) = 0$. Clearly, a necessary condition for an increase in payoff Π_k is $\mathrm{d}f_k(z^+) > 0$. We will show that this violates the *ex post* constraint, and therefore the instructor's optimal distribution is $F_k(x) = F^z(x)$.

First, see that the distribution of the realized scores on the course k is

$$g_k(x) := \Pr(x|A_k=1)=\frac{\Pr(A_k=1|x)\Pr(x)}{\Pr(A_k=1)}=\frac{\pi(x)\,f_k(x)}{f_k(z)\,M/K+f_k(z^+)}.$$

The *ex post* mean satisfies

$$\begin{aligned}\sum_{x\in\mathbb{N}^0} x\,g_k(x) &= \frac{z\,f_k(z)\,M/K}{f_k(z)\,M/K+f_k(z^+)}+\sum_{x=z+1}^{\infty}\frac{x\,f_k(x)}{f_k(z)\,M/K+f_k(z^+)}\\ &\geq \frac{z\,f_k(z)\,M/K+(z+1)\,f_k(z^+)}{f_k(z)\,M/K+f_k(z^+)}.\end{aligned}$$

Now, if we consider a deviation $\mathrm{d}f_k(z^+) > 0$, then in the new distribution, we observe $f_k(z^+) > 0$. Therefore its *ex post* mean violates the *ex post* constraint,

$$\sum_{x\in\mathbb{N}^0} x\,g_k(x)\geq\frac{z\,f_k(z)\,M/K+(z+1)\,f_k(z^+)}{f_k(z)\,M/K+f_k(z^+)}>\frac{z\,f_k(z)\,M/K+z\,f_k(z^+)}{f_k(z)\,M/K+f_k(z^+)}=z.$$

APPENDIX: EXAMPLES

CLOSED-FORM SOLUTION

The existence of a binomial function affects tractability of the problem. To begin with, once students either pick a single course ($M = 1$) or avoid a single course ($M = K - 1$), closed-form solutions are straightforward:

$$F(x,K,1) = \left(\tfrac{x}{K}\right)^{1/(K-1)},$$
$$F(x,K,K-1) = 1 - \left[1-(K-1)\,\tfrac{x}{K}\right]^{1/(K-1)}.$$

In the language of electoral systems (Myerson 1993), the two cases describe campaign promises for plurality and antiplurality voting systems, where a campaign promise is a distribution of a fixed cake in a population of voters. Any other case becomes rather complicated, even for a small K. For instance, the lowest K that generates a case different from plurality and antiplurality is $K = 4$, where $M = 2$. Then

$$B(1,3,p) = (1-p)^3 + 3p(1-p)^2 = (1-p)^2(1+2p) = F^2(x)\,[3-2F(x)] = \frac{x}{2}.$$

Using numerical tools, we obtain an offer function

$$F(x,4,2) = -\frac{1}{4}(1+i\sqrt{3})\left(\sqrt{x^2-2x}-x+1\right)^{1/3} - \frac{1-i\sqrt{3}}{4\left(\sqrt{x^2-2x}-x+1\right)^{1/3}} + \frac{1}{2},$$

on $x \in [0,2]$. Its probability distribution function has a U-shape, with many offers being very small and many offers being very large:

$$f(x,4,2) = \frac{(1-i\sqrt{3})\left(\frac{x-1}{\sqrt{x^2-2x}}-1\right)}{12\left(\sqrt{x^2-2x}-x+1\right)^{4/3}} - \frac{(1+i\sqrt{3})\left(\frac{x-1}{\sqrt{x^2-2x}}-1\right)}{12\left(\sqrt{x^2-2x}-x+1\right)^{2/3}}.$$

The case $K = 4$ serves as a clean illustration of the property found in Proposition 5: for a small M, the distribution is skewed to the right with a higher frequency of small offers (bad scores); for an intermediate M, there are both very small offers and very large offers (separation); and for a large M, the distribution is skewed to the left with a higher frequency of large offers (good scores).

To obtain a closed-form solution in general, we may represent the binomial function as a regularized incomplete beta function:

$$B(M-1,K-1,1-F(x,K,M)) = (M-1)\binom{K-1}{M-1}\int_0^{F(x,K,M)} t^{M-2}(1-t)^{K-M}\,\mathrm{d}t.$$

Hence its first derivative with respect to x must be constant. However, when evaluating the integral $\int t^{M-2}(1-t)^{K-M}\,\mathrm{d}t$, we obtain a product of an exponential function evaluated at x and the Gauss hypergeometric function evaluated at $F(x,K,M)$.

MEDIAN SCORE IS NOT MONOTONIC IN FLEXIBILITY

Suppose that we increase the requirements from $M = 1$ to $M = 2$; the number of electives is fixed and satisfies $K \geq 3$. Denote the median value $\tilde{x}(K, M)$, hence in this specific case $\tilde{x}(K, 1)$ and $\tilde{x}(K, 2)$. Given that $F(x) = \frac{1}{2}$ characterizes the median, we can easily characterize the median values, and find that increasing rigidity from $M = 1$ to $M = 2$ (i.e. limiting flexibility from its maximum) in fact *increases* the median value of noise:

$$\frac{\tilde{x}(K,2)}{\tilde{x}(K,1)} = \frac{B\left(1, K-1, \frac{1}{2}\right)}{2B\left(0, K-1, \frac{1}{2}\right)} = \frac{K}{2} > 1.$$

In contrast, suppose that we increase requirements from $M = K - 2$ to $M = K - 1$; again, the number of electives is fixed and satisfies $K \geq 3$. Analyse the ratio of the corresponding median values:

$$\begin{aligned}\frac{\tilde{x}(K,K-1)}{\tilde{x}(K,K-2)} &= \frac{K-2}{K-1} \cdot \frac{B\left(K-2, K-1, \frac{1}{2}\right)}{B\left(K-3, K-1, \frac{1}{2}\right)} \\ &= \frac{K-2}{K-1} \cdot \frac{1 - B\left(0, K-1, \frac{1}{2}\right)}{1 - B\left(1, K-1, \frac{1}{2}\right)} \\ &= \frac{K-2}{K-1} \cdot \frac{2^{K-1} - 1}{2^{K-1} - K}.\end{aligned}$$

We observe that $\tilde{x}(K, K-1) > \tilde{x}(K, K-2)$ for $3 \leq K \leq 5$, and $\tilde{x}(K, K-2) > \tilde{x}(K, K-1)$ for $K \geq 6$. In other words, increasing rigidity to its maximum *decreases* the median offered score if the number of the electives is sufficiently large.

FOR $M = 1$, SELF-SELECTION TURNS POSITIVE SKEW INTO NEGATIVE SKEW

We briefly show that self-selection greatly impacts scores distribution if the students pick just a single course. Consider $M = 1$. The probability density function of the *ex ante* grades is

$$f(x, K, 1) = \frac{1}{K-1} \left(\frac{1}{K}\right)^{1/(K-1)} x^{(2-K)/(K-1)}.$$

The probability density of the *ex post* grades is

$$g(x, K, 1) = x f(x, K, 1) = \frac{1}{K-1} \left(\frac{1}{K}\right)^{1/(K-1)} x^{1/(K-1)}.$$

Clearly, $f(x,K,1)$ is decreasing in x (median below mean; here also positive skew), whereas $g(x, K,1)$ is increasing in x (median above mean; here also negative skew).

APPENDIX: ENDOGENOUS RESOURCES

Recall that the resource constraint (mean-score cap) of an instructor is

$$\int_0^\infty x\,\mathrm{d}f(x) \le R.$$

Suppose that the school allocates a single unit of resources for free (i.e. total test points), while any extra resources $R-1$ (such as the instructor's attention) are costly to the instructor. The cost function $C(R)$ is continuous, increasing, convex and twice-differentiable, and to comply with our interpretation, it is normalized at $C(1)=0$.

The instructor k sets (R_k,F_k) simultaneously; in other words, he is flexible enough to adjust the amount of resources in the electives-shopping period if necessary. We will again look for the symmetric equilibrium characterized by (R,F). Clearly, by inserting the symmetric equilibrium value of R into equation (2), the implicitly characterized distribution F is now simply

$$B(M-1,K-1,1-F(x,K,M)) = x\frac{M}{KR}. \tag{A5}$$

To derive the equilibrium value of R, consider an instructor k's payoff on a deviation to any r_k. By inserting equation (A5) into the payoff and combining with the resource constraint, we obtain that extra resources increase the market share M/K proportionally:

$$\Pi_k = \int_0^\infty B(M-1,K-1,1-F(x,K,M))f(x)\,\mathrm{d}x - C(R_k) = \frac{M}{K}\frac{R_k}{R} - C(R_k).$$

The equilibrium value of R is characterized by the first-order condition, which yields the implicit function

$$I\left(\frac{K}{M};R\right) := C'(R) - \frac{M}{KR} = 0.$$

What is the effect of curriculum flexibility on resources? As more flexible curricula are more competitive, there is a lower 'market share' for each instructor. Since extra resources increase the market share proportionally, the marginal benefit is also lower. Hence flexibility somewhat unexpectedly *decreases* extra attention paid to the students. Formally, we use that

$$\frac{\partial I}{\partial K/M} = \frac{M^2}{K^2R} > 0 \text{ and } \frac{\partial I}{\partial R} = C''(R) + \frac{M}{KR^2} > 0$$

to obtain

$$\frac{\mathrm{d}R}{\mathrm{d}K/M} = -\frac{\partial I}{\partial K/M}\Big/\frac{\partial I}{\partial R} < 0.$$

In spite of this effect, we will show that the result about top scores observed in Corollary 1 is robust to the existence of the option to provide extra resources. The upper bound for the scores is $\bar{x} := rK/M$, hence the first-order condition is simply

$$\bar{x} = \frac{1}{C'(R)}.$$

By derivation, the top score is indeed increasing in curriculum flexibility:

$$\frac{\mathrm{d}\bar{x}}{\mathrm{d}K/M} = -\frac{C''(R)}{C'^2(R)} \frac{\mathrm{d}R}{\mathrm{d}K/M} > 0.$$

ACKNOWLEDGMENTS

I would like to thank participants at the 2018 European Meeting on Game Theory in Bayreuth for useful comments, and specifically Larry Samuelson for suggesting an extension towards an endogenous resource constraint. Support by institutional research funding at Charles University (PROGRES Q24) is acknowledged.

NOTES

1. Source:https://www.law.pitt.edu/student-resources/grades-and-transcripts (accessed 11 November 2020).
2. As individual GPAs are not observable, we assign students the mid-value of the interval for their received grade (mid GPA); e.g. for A+, the individual GPAs lie in [3.876,4], hence we consider the inferred GPA to take the mid-interval value of 3.938.
3. Recall that the setup features identical (symmetric) instructors as well as students. *Ex ante* homogeneity rules out *ex ante* heterogeneity in the grading standards associated with separating types of different abilities. Thus our paper does not address the ability channel documented by evidence on heterogeneous grading standards as in Bailey *et al.* (2016). When both courses and students are heterogeneous, sorting primarily involves signalling on the part of instructors (Herron and Markovich 2017). Our analysis of symmetric grading competition constitutes an initial step for a more complex analysis of asymmetric competition.
4. With continuous fit, notice that continuity in scores allows us to disregard issues associated with mapping from continuous fit to discrete grades. Among others, with flat mapping from fit to grades we would observe an incentive for the bunching of test scores (Diamond and Persson 2016).
5. The first moment of the interim scores distribution is fixed at 1, but what is known about the other properties of the distribution? Above all, Myerson (1993) numerically computes standard deviations for $K \leq 8$ and shows that the standard deviation is increasing in curriculum flexibility. Not all statistics are monotonic in curriculum flexibility, however. While the maximum (100th percentile) is decreasing in the number of required courses, the median (50th percentile) may decrease but also increase in the number of required courses, as the second subsection of Appendix B illustrates.
6. If the cap is binding extremely, $c < 1$, then the resource constraint is loose for any admissible distribution. In such a case, the equilibrium distribution obviously degenerates to the atom at $x = c$, and the equilibrium threshold is in a corner, $t = 0$.
7. That the threshold t is non-decreasing in c is given by the implicit function theorem imposed on equation (4).
8. Notice that low offers are $x \in [0,c]$, hence include the boundary offer $x = c$. High offers are all offers such that $x > c$.

REFERENCES

Achen, A. C. and Courant, P. N. (2009). What are grades made of? *Journal of Economic Perspectives,* **23**(3), 77–92.

Andreoni, J. and Brownback, A. (2017). Grading on a curve, and other effects of group size on all-pay auctions. *Journal of Economic Behavior & Organization*, **137**, 361–73.

Angling, P. M. and Meng, R. (2000). Evidence on grades and grade inflation at Ontario's universities. *Canadian Public Policy*, **26**(3), 361–8.

Babcock, P. (2010). Real costs of nominal grade inflation? New evidence from student course evaluations. *Economic Inquiry*, **48**(4), 983–96.

BAILEY, M. A., ROSENTHAL, J. S. and YOON, A. H. (2016). Grades and incentives: assessing competing grade point average measures and postgraduate outcomes. *Studies in Higher Education*, **41**(9), 1548–62.

BAR, T., KADIYALI, V. and ZUSSMAN, A. (2009). Grade information and grade inflation: the Cornell experiment. *Journal of Economic Perspectives*, **23**(3), 93–108.

BAR, T., KADIYALI, V. and ZUSSMAN, A. (2012). Putting grades in context. *Journal of Labor Economics*, **30**(2), 445–78.

BECKER, W. E. and ROSEN, S. (1992). The learning effect of assessment and evaluation in high school. *Economics of Education Review*, **11**(2), 107–18.

BETTS, J. R. (1998). The impact of educational standards on the level and distribution of earnings. *American Economic Review*, **88**(1), 266–75.

BETTS, J. R. and GROGGER, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, **22**(4), 343–52.

BOATRIGHT-HOROWITZ, S. L. and ARRUDA, C. (2013). College Students' categorical perceptions of grades: it's simply 'good' vs. *'bad'*. *Assessment & Evaluation in Higher Education*, **38**(3), 253–9.

BOLESLAVSKY, R. and COTTON, C. (2015). Grading standards and education quality. *American Economic Journal: Microeconomics*, **7**, 248–79.

BROWNBACK, A. (2018). A classroom experiment on effort allocation under relative grading. *Economics of Education Review*, **62**, 113–28.

BUTCHER, K. F., MCEWAN, P. J. and WEERAPANA, A. (2014). The effects of an anti-grade-inflation policy at Wellesley College. *Journal of Economic Perspectives*, **28**(3), 189–204.

CHAN, W., HAO, L. and SUEN, W. (2007). A signaling theory of grade inflation. *International Economic Review*, **48**(3), 1065–90.

CORREA, H. (2001). A game theoretic analysis of faculty competition and academic standards. *Higher Education Policy*, **14**(2), 175–82.

COSTRELL, R. M. (1994). A simple model of educational standards. *American Economic Review*, **84**(4), 956–71.

DIAMOND, R. and PERSSON, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. NBER Working Paper no. 22207.

DUBEY, P. and GEANAKOPLOS, J. (2010). Grading exams: 100, 99, 98,... or A, B, C? *Games and Economic Behavior*, **69**(1), 72–94.

DZIUBIńSKI, M. (2012). Non-symmetric discrete General Lotto games. *International Journal of Game Theory*, **42**, 801–33.

FIGLIO, D. N. and LUCAS, M. E. (2004). Do high grading standards affect student performance? *Journal of Public Economics,* **88**(9), 1815–34.

GORRY, D. (2017). The impact of grade ceilings on student grades and course evaluations: evidence from a policy change. *Economics of Education Review*, **56**, 133–40.

HARBAUGH, R. and RASMUSEN, E. (2018). Coarse grades: informing the public by withholding information. *American Economic Journal: Microeconomics*, **10**(1), 210–35.

HART, S. (2008). Discrete Colonel Blotto and General Lotto games. *International Journal of Game Theory*, **36**, 441–60.

HART, S. (2016). Allocation games with caps: from Captain Lotto to all-pay auctions. *International Journal of Game Theory*, **45**, 37–61.

HERNáNDEZ-JULIáN, R. and LOONEY, A. (2016). Measuring inflation in grades: an application of price indexing to undergraduate grades. *Economics of Education Review*, **55**, 220–32.

HERRON, M. C. and MARKOVICH, Z. D. (2017). Student sorting and implications for grade inflation. *Rationality and Society*, **29**(3), 355–86.

KAMENICA, E. and GENTZKOW, M. (2011). Bayesian persuasion. *American Economic Review*, **101**(6), 2590–615.

KENG, S. H. (2016). The effect of a stricter academic dismissal policy on course selection, student effort, and grading leniency. *Education Finance and Policy*, **11**(2), 203–24.

KOVENOCK, D. and ROBERSON, B. (2021). Generalizations of the General Lotto and Colonel Blotto games. *Economic Theory*, forthcoming.

LILLARD, D. R. and DECICCA, P. P. (2001). Higher standards, more dropouts? Evidence within and across time. *Economics of Education Review*, **20**(5), 459–73.

MATOS-DíAZ, H. (2012). Student evaluation of teaching, formulation of grade expectations, and instructor choice: explorations with random-effects ordered probability models. *Eastern Economic Journal*, **38**(3), 296–309.

MYERSON, R. B. (1993). Incentives to cultivate favored minorities under alternative electoral systems. *American Political Science Review*, **87**(4), 856–69.

OLSZEWSKI, W. and SIEGEL, R. (2019). Pareto improvements in the contest for college admissions. Working Paper, Northwestern University. https://cpb-us-e1.wpmucdn.com/sites.northwestern.edu/dist/2/1249/files/2019/06/paretoimprovements-6-5-2019.pdf

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

PAREDES, V. (2017). Grading system and student effort. *Education Finance and Policy*, **12**(1), 1070–128.

POPOV, S. V. and BERNHARDT, D. (2013). University competition, grading standards, and grade inflation. *Economic Inquiry*, **51**(3), 1764–78.

SABOT, R. and WAKEMAN-LINN, J. (1991). Grade inflation and course choice. *Journal of Economic Perspectives*, **5**(1), 159–70.

SAHUGUET, N. and PERSICO, N. (2006). Campaign spending regulation in a model of redistributive politics. *Economic Theory*, **28**, 95–124.

ZUBRICKAS, R. (2015). Optimal grading. *International Economic Review*, **56**(3), 751–76.