

Wagner, Lilo

Working Paper

Pass/Fail, A-F, or 0-100? Optimal grading of eager students

DIW Discussion Papers, No. 1445

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Wagner, Lilo (2015) : Pass/Fail, A-F, or 0-100? Optimal grading of eager students, DIW Discussion Papers, No. 1445, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/106989>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

1445

Discussion
Papers

Deutsches Institut für Wirtschaftsforschung

2015

Pass/Fail, A-F, or 0-100? –
Optimal Grading of Eager Students

Lilo Wagner

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2015

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<http://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
<http://ideas.repec.org/s/diw/diwwpp.html>
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

Pass/Fail, A-F, or 0-100? Optimal Grading of Eager Students

Lilo Wagner*

DIW Berlin

January 23, 2015

Abstract

This paper analyzes optimal grading in a world that focuses on top grades. Students choose an effort level, their performance is graded, and their grade correlates with their future income. Ex-ante, the policy maker chooses the optimal coarseness of the grading scale to maximize student welfare. When choosing their effort, students overweight outstanding – or salient – grades. I show that this behavior leads to excessive effort levels when grading is fully informative, and that coarse grading can be used to counterbalance incentives. Thus, salience can help explain why grading ranges from Pass/Fail scales (tenure decisions) via A-F-scales (school) to fully disclosing scores (e.g. SAT).

Keywords: Optimal Grading, Effort Incentives, Salience Theory, Education

JEL Classification Numbers: D83; D81; I21

*I especially thank Yves Breitmoser whose extensive suggestions had an enormous impact on this work. Pio Baake, Alfonso Caiazzo, Dorothea Kübler, Tobias Schmidt, Paul Viefers, Georg Weizsäcker and seminar participants at DIW Berlin provided many helpful comments for which I am grateful. Contact: Mohrenstrasse 58, 10117 Berlin, Germany, Email: lwagner@diw.de, Phone: +49 30 897 89 661.

1 Introduction

Students' examination results are typically recorded on a very accurate scale, like 1, 2..., 99, 100. Yet, educational institutions pool different scores together in broad categories, for instance by adopting a letter grading scale. In fact, perfectly accurate final reporting can rarely be found. This prevalence of coarse grading scales is puzzling as it reduces the informativeness of diplomas to employers.

Even more surprisingly, selectivity seems to be a good indicator for the degree to which information is suppressed. Yale Law school for instance employs a broad scale on the basis of Fail/Low Pass/Pass/Honors. Other top law schools like Harvard and Stanford have adapted similar scales while less selective schools tend to stick to the finer traditional (and usually curved) letter grading scale.

A more subtle method to install a coarse scale is to factually abolish lower marks by just never or very rarely assigning them, a phenomenon recently discussed in the news under the headline of grade inflation. Stuart Rojstaczer, one of the initiators of the debate states in a *Washington Post* article: “I recently handed in my grades for an undergraduate course I teach at Duke University. They were a very limited assortment: A, A-minus, B-plus, B and B-minus. There were no C’s of any flavor and certainly no D’s or F’s. It was a good class, but even when classes aren’t very good, I just drop down slightly, to grades that range from A-minus to B-minus.”¹

At Harvard, the most frequently given grade is now an A, the median an A-.² When the *Business Insider* lists 13 schools where it’s almost impossible to fail³, the list comprises almost exclusively elite private schools. More selective institutions assign higher GPA’s to their students than do their peers for the same education level (Healy and Rojstaczer, 2012), presumably corresponding to a lower support in the scale.⁴

The purpose of the present paper is to provide a justification for the prevalence of such coarse grading. Different explanations come to mind which however do not account for the fact that coarse scales can be found most and for all at top schools.

For instance, coarse grading scales may be the equilibrium outcome of a game in which employers can be misled about educational quality (as for instance investigated in Chan et al., 2007). Alternatively, they may reflect the noisiness of performance or performance

¹“Where All Grades Are Above Average”, *Washington Post*, January 28, 2003

²“Substantiating Fears of Grade Inflation, Dean Says Median Grade at Harvard College Is A-, Most Common Grade Is A”, *The Harvard Crimson*, December 3, 2013

³“13 schools where it’s almost impossible to fail”, *The Business Insider*, May 29, 2013

⁴Albeit not perfectly correlated, the fact that the share of D’s at private colleges has now dropped to less than 3% (from the same report) leads me to conclude that. Also, many reports indicate that this phenomenon is prevalent most and for all at the elite schools.

evaluation. In that sense, scores may offer a spurious precision which in reality does not exist. Existing work demonstrates that coarse grading can induce students to exert higher effort (Dubey and Geanakoplos, 2010). Also, coarse grading may be optimal for profit maximizing schools whose students are distinct not only in their abilities but who also pay different tuition fees (Rayo and Segal, 2010).

I call attention to a different explanation showing that coarse grading may maximize students' ex-ante welfare. In particular, information suppression may be a school's best response to a student body that consists of individuals who place too much weight on top grades. Consider the following conjecture: *"Ivy League educational institutions attract a disproportionate share of grade-obsessed overachievers. [...] Their compulsion to succeed as others define it and their sheepish failure to prioritize higher-order benefits with their time at college perhaps makes a grading scale based on obvious inflation the best option available."*⁵ Similarly, when Yale Law School changed its scale from numerical grading to the broader scale in 1967, it was reported that *"it is believed by those members of the faculty who voted in favor the plan that it will offer some relief from what one educator described as 'the excessive preoccupation with number or letter grades' "*⁶

It is this study's aim to formally test the validity of this classic conjecture. The intuition behind it appears to be that coarse grading might serve to increase student welfare by offering relief, that is by lowering effort levels. This however stands in contrast to existing models where coarse grading raises student effort. In this case, coarseness may be exactly the wrong measure to cope with grade obsession.

In the present study, coarse grading scales are used by a school to maximize student welfare. In a simple model in which grades are based on exam scores I show that on the one hand, fully disclosing results is uniquely optimal when students are expected utility maximizers. On the other hand, it is demonstrated that grade obsession induces students to exert inefficiently high effort and that coarse scales can be used to counterbalance incentives. For this purpose, I model grade obsession by applying salience theory according to Bordalo, Gennaioli and Shleifer (2012) – henceforth BGS.

In their theory, a decision maker places too much weight on the most outstanding – or salient – payoffs when having to choose among lotteries. Thus, grade obsession is translated into a focus on both the very good and the very bad outcomes at the expense of average grades.⁷ Since students can never be certain about the success of their efforts, they are faced with the choice among independent lotteries, with payoffs being determined on the

⁵"In Defense of Grade Inflation at Harvard", *The Atlantic*, December 6, 2013

⁶"Yale College inaugurates pass-fail marking scale", *The Heights*, November 13, 1967

⁷The hypothesis that focusing equates to focusing on large differences is also explored by Bordalo et al. (2013), Kőszegi and Szeidl (2013) and Tversky (1969).

job market. Due to the salience of extreme outcomes, students replace the true outcome probabilities by salience-distorted weights. More specifically, the salience of a payoff is context-dependent, meaning that outcomes are compared to those of other lotteries which realize in the same state of the world.

This model of salience captures the main insights from educational psychology in a stylized form. Those students who do not focus on salient grades may pursue mastery goals, aiming at increasing their competency and deeper knowledge. Others – the grade obsessed students – are either performance approachers who strive to achieve extrinsic success, or failure avoiders (for a review see e.g. Covington, 2000). In a sense, the focus on salient grades – the very best and worst grades – can be understood as both performance approach and failure avoidance orientation.⁸

The following is the setting: an institution which seeks to maximize the expected utility of its students commits to and publicly announces a grading scale. Based on this, students of different cost types⁹ make an unobserved decision between employing costly high effort, and costless low effort. The associated probability mass functions are standardly assumed to satisfy the monotone likelihood ratio property. Scores then realize, grades are assigned, and students go on the job market where risk neutral employers are willing to pay higher salaries for better expected scores.

Clearly, when students maximize expected utility, a grading scale which fully discloses scores is socially efficient, and it is even uniquely optimal. A focus on the most salient states however brings students to exert inefficiently high effort, from the school’s perspective. Roughly speaking, there are two most salient states, but a student who considers employing high effort overweights the good grades relatively more than the bad grades. Similarly, when considering to exert low effort, the student overweights bad grades relatively more than good grades. This effect arises due to the monotone likelihood ratio property.

The share of hard working students is a continuous function of the degree of salience, implying that coarse rules which mimic full disclosure at the extreme have an intersection with the expected utility maximizing cut-off and are therefore optimal for some degree of

⁸Failure avoidance is perhaps better captured by disappointment aversion which – like a focus on the very worst grades – results in lower effort level as compared to expected utility maximizing behaviour. Also, note that here, the focus is externally given while educational psychologists take into account that altering extrinsic incentives possibly affects goal orientation types. Especially curved scales are seen to foster extrinsic motivation (performance goal orientation), intrinsic motivation (mastery goal orientation) and student satisfaction is better supported by absolute scales.

⁹This may be the opportunity costs for what is named higher order benefit in the above citation: it is a well established thesis among psychologists that while extrinsic motivation responds well to incentives, it comes at the cost of a reduced intrinsic motivation (for an overview and critical assessment: Cameron and Eisenberger, 1996). Similarly, institutions may be concerned about the health of their students: at Harvard Law School, the scale change reportedly served to reduce stress and anxiety levels.

salience. I further show that optimal coarse rules exist whenever full disclosure induces too many students to choose the high effort lottery, from the school's perspective. This is typically so for all degrees of salience.

Related literature. To the best of my knowledge, this is the first study to formally investigate the relationship between the focus on grades of students and optimal grading scales. Coarse grading otherwise may raise effort levels. Dubey and Geanakoplos (2010) show that if students care about their rank in class and grading scales are chosen such that students maximize their studying efforts, coarse scales may be the best choice and are always so when students are disparate. The reason for the fact that coarsening increases effort levels in Dubey and Geanakoplos (2010) stems from the contest structure: if students are very different in abilities, then fully revealing results may fix the student ranking even if no student employs high effort. In the present paper by contrast, it is assumed that students care about their salaries rather than their rank in class, and that higher expected scores are associated to higher salaries. This setting however invalidates their ranking-based argument such that here, coarseness reduces effort levels rather than increasing them.

A related topic is treated by Zubrickas (Forthcoming) who theoretically shows that the best students should receive the same best grade when schools maximize student effort (which is equal to the educational achievement here) whereas students seek to signal their abilities and the grading rule is not observable by the job market. Zubrickas (Forthcoming) thus addresses a different question. In contrast to the present paper, different student types are compressed for the best grades – scores are not. Rather, different scores receive different grades by the exogeneity of the grade utilities.

Effort incentives of grading rules are also theoretically investigated by Costrell (1994) and Betts (1998), who however restrict their study to optimal Pass/Fail scales under different welfare conceptions. In that sense, their focus is rather on grading standards than on rules. In empirical research, the effect of grading standards on student motivation has been an issue for a while. For instance, Figlio and Lucas (2004) and Betts and Grogger (2003) investigate the effect of raising a Pass/Fail standard. Both works find that students' educational achievement generally responds positively to increased standards. Others explore whether effort is chosen strategically, i.e. such that scores closely above a threshold become probable (Oettinger, 2002).

Some works theoretically analyze grading scales in adverse selection models. Rayo and Segal (2010) show that coarse grading scales can be optimal if grading is based not solely on the abilities or educational achievement of students but also on their profitability to the sender (the paid tuition fees). Further, Ostrovsky and Schwarz (2010) analyze best disclosure policies when employers rank students according to their exogenously given abilities. Partial

disclosure can be a market equilibrium if student bodies exhibit different exogenously given talent levels. In particular, an average school, if faced with low quality competitors, may take into account non-revelation of highly talented students in order to attain a good placement for less able persons. A similar effect is present in Boleslavsky and Cotton (Forthcoming).

In industrial organization, a stream of literature starting with Lizzeri (1999) shows why profit maximizing certifiers may find it optimal to employ coarse disclosure policies in adverse selection settings. In Lizzeri (1999), a certifier who is bound to offer his service at a fixed non-discriminatory fee to sellers with different quality goods considers disclosure on a Pass/Fail basis optimal. Similarly, intermediaries who seek to maximize not their own profits but the information being provided to the public, partial disclosure may outrule fully disclosing rules if a selection process is at work and certification is costly (Harbaugh and Rasmusen, 2013).

The paper proceeds as follows. Section 2 describes the model. In Section 3, I derive two general insights, which are independent of the focus on salient payoffs but relate only to the game structure. The notation of salience characterized in Section 4, and optimal rules are analyzed in Section 5. The last section discusses extensions and concludes. All proofs are relegated to the appendix.

2 The model

Students are assumed to have a cost for studying of $\theta \in [0, 1]$. θ is uniformly distributed across students. A student's choice set is $\{\mathcal{L}_0, \mathcal{L}_e\}$ where the choice of lottery \mathcal{L}_e costs θ whereas that of \mathcal{L}_0 is free. Students are assumed to be risk neutral but allowing for risk aversion would not affect the qualitative properties of the equilibria.¹⁰ An expected utility maximizing student of type θ then chooses lottery \mathcal{L}_e if and only if the expected value from doing so, minus the costs is at least as high as the expected value of choosing \mathcal{L}_0 .

The lottery choice determines the probability distribution of exam scores. In particular, an exam score is an element of the ordered set $\Omega = \{q^1, q^2, \dots, q^N\}$ with $N \geq 3$. The probability that a score in Ω is q^n is given by p_e^n if the student chooses lottery \mathcal{L}_e and it is p_0^n otherwise. p_e and p_0 are the according probability vectors. All elements are assumed to be strictly positive. In order to ensure interior solutions, it is standardly assumed that probability functions satisfy the monotone likelihood ratio property. Define $\tau_n := p_e^n/p_0^n$.

(A1) (*Monotone likelihood ratio property*) $\tau_n > \tau_m \quad \forall n > m$

¹⁰Full disclosure will always bring students to exert inefficiently high effort because the distortive effect weights heavier on high than on low probability states. This effect arises due to the monotone likelihood ratio property (see below) but is independent of student utility as long as it is monotonically increasing in payoffs.

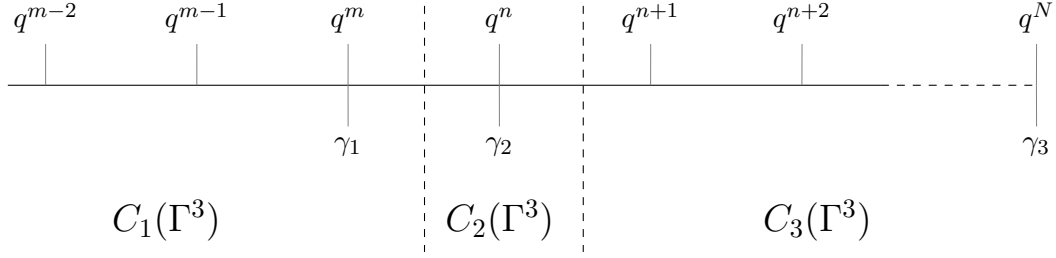


Figure 1: A grading rule $\Gamma^3 \subset \Psi_3$.

The condition guarantees that students who choose the costly lottery \mathcal{L}_e will attain higher expected scores than the ones whose choice is \mathcal{L}_0 .

Scores in Ω are ordered according to the salaries they are paid on the job market. There, high scores are rewarded. By assumption, an employer is willing to pay a salary of n/N for a score q^n . With slight abuse of notation, q^n will be used to denote both the score and the salary the score commands on the market. Thus, $q^n = n/N$ and $q^N = 1$. Hence, when the market exhibits symmetric information, that is, when employers can perfectly observe scores, the expected valuations for both lotteries are the expected scores, i.e. $E[Q|\mathcal{L}_e] = \sum_{n=1}^N q^n p_e^n$ or $E[Q|\mathcal{L}_0] = \sum_{n=1}^N q^n p_0^n$ respectively, where q^n is the realization of random variable Q .

Teachers convert scores into grades using a grading rule that maps every possible score q^n into a grade. A deterministic grading rule¹¹ is defined as a set of thresholds $\Gamma^H = \{\gamma_1, \gamma_2, \dots, \gamma_H\}$ with $\gamma_H = q^N$ and $\gamma_1 < \gamma_2 < \dots < \gamma_N$. The associated grades are denoted by $C_1(\Gamma^H), C_2(\Gamma^H), \dots, C_H(\Gamma^H)$. For instance, γ_1 could be 30 out of 100 points and $C_1(\cdot)$ would be grade D . Further, let Ψ be the set of all possible grading rules. In particular, a grading rule in Ψ induces a partition of Ω into consecutive intervals. It can be expressed as a subset of Ω : $\Gamma^H \subseteq \Omega$.

The mapping between scores and grades according to rule Γ^H and grades $C_1(\Gamma^H), C_2(\Gamma^H), \dots, C_H(\Gamma^H)$ is as follows. All scores q^n that are equal to or worse than γ_1 are granted grade $C_1(\Gamma^H)$, all scores that are strictly better than γ_1 but equal to or worse than γ_2 are assigned grade $C_2(\Gamma^H)$ and so forth. The best grade $C_H(\Gamma^H)$ is awarded to scores greater than γ_{H-1} . An example for a rule Γ^3 is depicted in figure 1: if $\gamma_1 = q^m, \gamma_2 = q^n$, then the lowest grade $C_1(\Gamma^3)$ is awarded to all scores lower or equal to q^m , $C_2(\Gamma^3)$ is awarded to scores q^{m+1} to q^n and the best grade to all scores strictly greater than q^n . From this follows that $\gamma_H := q^N = 1$ and $1 \leq H \leq N$.

If $H = 1$, no information is disclosed, if $H = N$, scores are fully disclosed. I use Γ to describe a general rule and Γ^H to describe some rule with $H - 1$ cuts. The same holds for Ψ

¹¹A deterministic rule is a rule in which each score is deterministically assigned a grade. Simple probabilistic rules, i.e. rules in which a score may be assigned different grades with positive probability are used later to derive an existence condition of optimality for given degrees of salience.

and Ψ_H respectively: $\Psi_H \subset \Psi$ is the set of disclosure rules that are described by some Γ^H . For instance, all possible Pass/Fail rules are contained in the set Ψ_2 , where γ_1 is an element of $\Omega/\{q^N\}$.

Employers are informed about the grading rule. Also, the distribution of costs and the probability distributions are assumed to be common knowledge whereas the students' lottery choices cannot be observed. On the basis of the information available employers form expectations about the value of a grade $C_h(\cdot)$, denoted by $v_h(\cdot)$ ¹² (i.e. the salary). The equilibrium concept I use is the Weak Perfect Bayesian Equilibrium. This requires that students' lottery choices are optimal given employer beliefs and that employer beliefs are consistent with the strategies being played by students.¹³

By definition of the grading rule, the value of a grade $v_h(\cdot)$ is increasing in h and located on the interval $[q^1, q^N]$. Further, the value $v_h(\cdot)$ of a grade $C_h(\cdot)$ which is awarded for scores in Ω $q^n, q^{n+1}, \dots, q^{m-1}, q^m$ is located on the interval $[q^n, q^m]$. Also note that the values $v_h(\cdot)$ are independent of the distribution of grades, i.e. of the relative rank of students. A grade is an absolute measure of achievement. Hence, the relative ranking of students is not required for potential employers to gauge their value for the company if the grading rule is common knowledge. If it is not, the relative ranking might become important. The game proceeds as follows.

- (1) The grading rule is publicly announced,
- (2) students make a choice between \mathcal{L}_e and \mathcal{L}_0 ,
- (3) scores are drawn according to the chosen distribution \mathcal{L}_e or \mathcal{L}_0 , exams are graded according to the grading rule,
- (4) students go on the job market and realize payment $v_h(\cdot)$ for grade C_h .

3 Optimal grading with expected utility

Prior to making a lottery choice, a student evaluates both lotteries \mathcal{L}_e and \mathcal{L}_0 . Let $V_e(\Gamma|\phi)$ and $V_0(\Gamma|\phi)$ be the expected valuations from choosing the respective lottery for a given

¹²The grade value v_h is a function of the grading rule and the belief employers hold about students' lottery choices as will become clear below.

¹³The Weak Perfect Bayesian Equilibrium is for instance formally defined in Mas-Colell et al. (1995), p. 285. Note that off-path beliefs do not play a role here since by definition, a grading rule specifies a grade for each score. Every deviating lottery choice therefore induces an observation (a grade) on the path. As a result, the Weak Perfect Bayesian Equilibrium is identical to all notions of Perfect Bayesian Equilibrium.

grading rule where ϕ denotes the belief that the job market holds about values of grades.¹⁴ Define $\gamma_0 := 0$ and define the random variable capturing respective future income as V . Then, $V_e(\Gamma|\phi)$ and $V_0(\Gamma|\phi)$ are given by

$$\begin{aligned} V_e(\Gamma|\phi) &= E[V|\mathcal{L}_e] = \sum_{h=1}^H v_h(\Gamma|\phi) \sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} p_e^i, \\ V_0(\Gamma|\phi) &= E[V|\mathcal{L}_0] = \sum_{h=1}^H v_h(\Gamma|\phi) \sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} p_0^i. \end{aligned} \quad (1)$$

Note that the choice of Γ determines, in addition to the market values for grades, the sums in equation (1). Note further that the second sum in both expressions of equation (1) is the probability for grade h to occur given the lottery choice and the grading rule: if for instance $\gamma_{h-1} = q^m$ and $\gamma_h = q^n$, then the probability that a student choosing lottery \mathcal{L}_e is assigned grade $C_h(\cdot)$ is given by $p_e^{m+1} + p_e^{m+2} + \dots + p_e^n$.

An expected utility maximizing student is then defined as follows.

(A2) (*Expected utility maximization*) An expected utility maximizing student θ chooses lottery \mathcal{L}_e if and only if $V_e(\Gamma|\phi) - V_0(\Gamma|\phi) \geq \theta$.

The school seeks to maximize its conception of student welfare, which is overall expected utility. An immediate consequence from (A2) is that it is never an optimal strategy (from the viewpoint of the school and the students) that high types choose \mathcal{L}_e when lower types do not. Thus, student strategies are fully captured by some critical type $\hat{\theta}$ and welfare is given by

$$W(\Gamma, \hat{\theta}|\phi) = \int_0^{\hat{\theta}} (V_e(\Gamma|\phi) - \theta) d\theta + \int_{\hat{\theta}}^1 V_0(\Gamma|\phi) d\theta \quad (2)$$

for a rule Γ .

Transforming equation (2) and maximizing with respect to $\hat{\theta}$ gives the first best critical value of θ for a given grading rule such that all students with lower costs should choose \mathcal{L}_e whereas higher types should choose \mathcal{L}_0 . For a given grading rule, this optimal critical type is given by $\hat{\theta}^r(\Gamma|\phi) = V_e(\Gamma|\phi) - V_0(\Gamma|\phi)$. Thus, if the school was able to choose the critical type for a given grading rule, it would make the same choice as expected utility maximizing students do.

The market exhibits symmetric information if scores are fully disclosed to the public.¹⁵

¹⁴More precisely, employers hold a belief ϕ about the critical cost type who just chooses lottery \mathcal{L}_e whenever scores are not fully disclosed. For a discussion, see later in this section.

¹⁵Recall that a grading rule which fully discloses scores is described by Γ^N which equals Ω .

In this case of full disclosure, $\hat{\theta}^r(\Gamma|\phi)$ is also the first best critical type.¹⁶ It is given by the difference in expected scores for a given lottery choice \mathcal{L}_e or \mathcal{L}_0 , i.e.

$$\hat{\theta}^{FB} = E[Q|\mathcal{L}_e] - E[Q|\mathcal{L}_0] = \sum_{n=1}^N q^n (p_e^n - p_0^n).$$

Denote by $W^{FD}(\hat{\theta})$ the $W(\Gamma, \hat{\theta}|\phi)$ when scores are fully disclosed, i.e. if $\Gamma = \Gamma^N = \Omega$.¹⁷ In words, $W^{FD}(\hat{\theta})$ is the expected utility with full disclosure as a function of the critical type $\hat{\theta}$. The highest expected utility for the school is then $W^{FD}(\hat{\theta}^{FB})$, that is, if students are expected utility maximizers and scores are fully disclosed.

When scores are not fully disclosed, employers form, for a given grading rule, a belief ϕ about the critical type $\hat{\theta}$. The values of grades to students are then a function of this belief. To see why, consider some rule such that grade $C_H(\Gamma)$ is coarse: if many students are believed to have chosen \mathcal{L}_e , that is, if ϕ is large, then the value of this grade is high as compared to a lower ϕ because the likelihood of a student with an awarded grade $C_H(\Gamma)$ having attained a top score rather than a still good score is high whereas it is lower for lower ϕ . If scores are fully disclosed however, the salary associated to a grade C_n is simply the score $q^n = n/N$. For a general rule, the value of a grade $C_h(\Gamma)$ to students is given by

$$v_h(\Gamma|\phi) = \frac{\sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} q^i (\phi p_e^i + (1-\phi)p_0^i)}{\sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} (\phi p_e^i + (1-\phi)p_0^i)}. \quad (3)$$

Note that the function argument Γ determines the sums. Hence, because the values of grades are a function of the belief ϕ , so are the expected valuations from choosing the lotteries, $V_e(\Gamma|\phi)$ and $V_0(\Gamma|\phi)$. They can be derived by inserting equation (3) into equations (1).

Following the concept of Weak Perfect Bayesian Equilibrium, students' lottery choices can be foreseen by employers and beliefs are correct. Denote by $\hat{\theta}^r(\Gamma)$ a ϕ that solves $\phi = \hat{\theta}^r(\Gamma|\phi)$. By Brouwer's fixed-point theorem, because $\hat{\theta}^r(\Gamma|\phi)$ is a mapping from the closed interval $[0, 1]$ into itself, $\hat{\theta}^r(\Gamma)$ exists for every grading rule Γ . The school whose student body consists of expected utility maximizers considers the following welfare function:

$$W^r(\Gamma) = \int_0^{\hat{\theta}^r(\Gamma)} (V_e(\Gamma|\hat{\theta}^r(\Gamma)) - \theta) d\theta + \int_{\hat{\theta}^r(\Gamma)}^1 V_0(\Gamma|\hat{\theta}^r(\Gamma)) d\theta. \quad (4)$$

¹⁶Lemma 1 further below demonstrates that first, the critical type solely determines welfare, and second, that full disclosure indeed maximizes welfare with expected utility maximizing students.

¹⁷Recall that when the grading rule is full disclosure, the market values do not depend on beliefs. Thus, nor does student welfare.

A rule is called *expected utility optimal* if it maximizes equation (4), that is, if it induces utility $W^{FD}(\hat{\theta}^{FB})$. Full disclosure is expected utility optimal and implies that $\hat{\theta}^r(\Gamma) = \hat{\theta}^{FB}$. The following lemma proves that this condition suffices to determine expected utility optimality for other rules.

Lemma 1. *A grading rule is considered expected utility optimal if and only if it induces the first best critical type $\hat{\theta}^{FB}$. Suppose $\hat{\theta}^r(\Gamma) = \hat{\theta}^{FB}$ for some $\Gamma \subset \Psi$. Then, it holds that $W^r(\Gamma) = W^{FD}(\hat{\theta}^{FB})$. By contrast, if $\hat{\theta}^r(\Gamma) \neq \hat{\theta}^{FB}$, it holds that $W^r(\Gamma) < W^{FD}(\hat{\theta}^{FB})$.*

Lemma 1 states that whether a grading rule is expected utility optimal is exclusively determined by the critical type which it induces in equilibrium. Put differently, coarseness itself does not affect welfare, only its effect on the lottery choices does. Thus, the value of a grading rule is given by its ability to incentivize student effort. This result is surprising, and the more so since in equilibrium, even when the induced critical type is the same, students' expected valuations for the lotteries are typically different under coarse rules than they are under a full disclosure rule. The insight provided in Lemma 1 considerably simplifies matters since the analysis of critical types suffices to reveal expected utility optimal grading rules.¹⁸

It is this study's main purpose to demonstrate why schools may find it optimal to employ a coarse grading rule rather than fully disclosing examination results. By Lemma 1, a grading rule Γ is expected utility optimal if and only if $\hat{\theta}^r(\Gamma) = \hat{\theta}^{FB}$. This is equivalent to $\hat{\theta}^r(\Gamma|\phi) = \phi = \hat{\theta}^{FB}$. The following can be shown.

Proposition 1. *In any equilibrium with expected utility maximizing students, students employ less than the efficient effort level when disclosure is coarse. That is, for all rules in Ψ/Ψ_N , it holds that for any belief ϕ , $\hat{\theta}^r(\Gamma|\phi) < \hat{\theta}^{FB}$.*

Thus, when students are expected utility maximizers, no other than the full disclosure rule is optimal. More precisely, any rule that does not fully disclose scores induces also too less effort from the school's point of view, a result which does not depend on the belief ϕ . Hence, the intuition that incentives to work are lowered if better scores are not fully rewarded applies.

4 Salience of extreme grades

Following BGS, students evaluate H^2 different payoff states in S where S is determined by the grading rule Γ . In particular, in a state $s_{gh} \in S(\Gamma)$, a student is awarded grade $C_g(\Gamma)$

¹⁸Note that Lemma 1 fails to hold for risk averse students. The according analysis would then require case-to-case evaluations of grading rules. With risk neutral students by contrast, it is possible to focus on critical types induced by different grading rules. This result extends to a focus on salient grades, see below.

when choosing lottery \mathcal{L}_0 and grade $C_h(\Gamma)$ when choosing lottery \mathcal{L}_e . Each state $s_{gh} \in S(\Gamma)$ then occurs with a commonly known probability, given by

$$\pi_{s_{gh}}(\Gamma) = \sum_{i=\gamma_{g-1}N+1}^{\gamma_g N} p_0^i \sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} p_e^i \quad \forall s_{gh} \in S(\Gamma).$$

Accordingly, it holds that $\sum_{s_{gh} \in S(\Gamma)} \pi_{s_{gh}}(\Gamma) = 1$.

The salience of a state $s_{gh} \in S(\Gamma)$ is described by a continuous, non-negative and symmetric (states s_{gh} and s_{hg} are equally salient) salience function $\sigma_{s_{gh}}(\Gamma|\phi) := \sigma(v_g(\Gamma|\phi), v_h(\Gamma|\phi))$. As proposed by BGS, $\sigma_{s_{gh}}(\cdot)$ satisfies the ordering property.¹⁹ Formally:

(A3) (*Ordering*) For all states $s_{g'h'}, s_{g''h''} \in S(\Gamma)$ for which $\min\{v_{g'}(\cdot), v_{h'}(\cdot)\} \leq \min\{v_{g''}(\cdot), v_{h''}(\cdot)\}$ and $\max\{v_{g'}(\cdot), v_{h'}(\cdot)\} > \max\{v_{g''}(\cdot), v_{h''}(\cdot)\}$ or $\min\{v_{g'}(\cdot), v_{h'}(\cdot)\} < \min\{v_{g''}(\cdot), v_{h''}(\cdot)\}$ and $\max\{v_{g'}(\cdot), v_{h'}(\cdot)\} \geq \max\{v_{g''}(\cdot), v_{h''}(\cdot)\}$, it holds that $\sigma_{s_{g'h'}}(\cdot) > \sigma_{s_{g''h''}}(\cdot)$.

The ordering property says that whenever one state's values are nested within the values of some other state, the latter is more salient. For instance, consider three grades $C_1(\Gamma^3), C_2(\Gamma^3)$ and $C_3(\Gamma^3)$ where by definition of a grading rule, $v_1(\cdot) < v_2(\cdot) < v_3(\cdot)$. Then the states $s_{13} = s_{31}$ are most salient and one of the states s_{11}, s_{22} and s_{33} are least salient. By contrast, (A3) does not make any statement on the relative salience of states s_{11} to s_{23} or s_{23} to s_{12} .

Generally speaking, it implies that s_{1H} and s_{H1} are the most salient states, whereas some state s_{gg} , $g \in \{1, 2, \dots, H\}$, exhibits the least salience.

To obtain the decision weights attached to some state s_{gh} , define the weighting function as

$$\omega_{s_{gh}}(\Gamma|\phi, \delta) = \frac{\delta^{-\sigma_{s_{gh}}(\Gamma|\phi)}}{\sum_{s_{st} \in S(\Gamma)} \delta^{-\sigma_{s_{st}}(\Gamma|\phi)} \pi_{s_{st}}(\Gamma)}. \quad (5)$$

$1 - \delta \in (0, 1]$ measures the degree of salience, where $\delta = 1$ means that students are expected utility maximizers in the sense of (A2). The perceived probability for a state s_{gh} is given by

$$\omega_{s_{gh}}(\Gamma|\phi, \delta) \pi_{s_{gh}}(\Gamma)$$

where $\sum_{s_{gh} \in S} \omega_{s_{gh}}(\cdot) \pi_{s_{gh}}(\cdot) = 1$.

The values of grades are a function of the belief ϕ about the critical type, as before. Hence, $v_h(\Gamma|\phi)$ is the same as in equation (3). As in the previous section, a student evaluates lottery

¹⁹BGS also propose that the salience function should exhibit diminishing sensitivity, that is, for any two states $s_{g'h'}, s_{g''h''} \in S(\Gamma)$, where $v_{g''} = v_{g'} + \varepsilon$ and $v_{h''} = v_{h'} + \varepsilon$, for any $\varepsilon > 0$, $\sigma_{s_{g'h'}} \geq \sigma_{s_{g''h''}}$. I do not need this property to derive the general result, I will however, use a function that satisfies it when conducting the numerical analysis.

\mathcal{L}_e against lottery \mathcal{L}_0 . Contrary to an expected utility maximizing student however, these values are distorted by the weighting function. Let denote these perceived values $\tilde{V}_e(\Gamma|\cdot)$ or $\tilde{V}_0(\Gamma|\cdot)$ respectively. They are given by

$$\begin{aligned}\tilde{V}_e(\Gamma|\phi, \delta) &= \sum_{h=1}^H v_h(\Gamma|\phi) \sum_{i=1}^H \omega_{s_{ih}}(\delta) \pi_{s_{ih}}(\Gamma) \\ \tilde{V}_0(\Gamma|\phi, \delta) &= \sum_{h=1}^H v_h(\Gamma|\phi) \sum_{i=1}^H \omega_{s_{hi}}(\delta) \pi_{s_{ih}}(\Gamma).\end{aligned}$$

A student who focusses on salient grades is defined as follows.

(A4) (*Focus on salient grades*) A student θ with a degree of salience of $1 - \delta$ chooses lottery \mathcal{L}_e if and only if $\tilde{V}_e(\Gamma|\phi, \delta) - \tilde{V}_0(\Gamma|\phi, \delta) \geq \theta$.

As with expected utility maximizing students, student strategies are again captured by some critical type such that all students with lower costs choose \mathcal{L}_e while higher types choose \mathcal{L}_0 . The critical type is now a function of the degree of salience, given by

$$\hat{\theta}^s(\Gamma^H|\phi, \delta) = \sum_{g=1}^H \sum_{h=g+1}^H (v_h(\Gamma^H|\phi) - v_g(\Gamma^H|\phi)) \omega_{s_{gh}}(\Gamma^H|\phi, \delta) (\pi_{s_{gh}}(\Gamma^H) - \pi_{s_{hg}}(\Gamma^H)), \quad (6)$$

where $\hat{\theta}^s(\Gamma^H|\phi, 1) = \hat{\theta}^r(\Gamma^H|\phi)$.

The school is assumed to maximize students' expected utility. In analogy to the previous analysis, define $\hat{\theta}^s(\Gamma|\delta)$ as a belief ϕ which solves $\phi = \hat{\theta}^s(\Gamma|\phi, \delta)$. Note that by continuity of $\hat{\theta}^s(\cdot)$ in ϕ , such a $\hat{\theta}^s(\cdot)$ exists for any combination of δ and a grading rule Γ . When choosing a grading rule, the school considers the following welfare function:

$$W^s(\Gamma|\delta) = \int_0^{\hat{\theta}^s(\Gamma|\delta)} (V_e(\Gamma|\hat{\theta}^s(\Gamma|\delta)) - \theta) d\theta + \int_{\hat{\theta}^s(\Gamma|\delta)}^1 V_0(\Gamma|\hat{\theta}(\Gamma|\delta)) d\theta. \quad (7)$$

A rule is called *first best optimal* if and only if welfare is the same as when students are expected utility maximizers and scores are fully disclosed, that is if and only if $W^s(\Gamma|\cdot) = W^{FD}(\hat{\theta}^{FB})$. The following lemma generalizes Lemma 1 to a focus on salient grades.

Lemma 2. *A grading rule is considered first best optimal if only if it induces the first best critical type $\hat{\theta}^{FB}$. Suppose $\hat{\theta}^s(\Gamma|\delta) = \hat{\theta}^{FB}$ for some $\Gamma \subset \Psi$ and a given degree of salience $1 - \delta$. Then, it holds that $W^s(\Gamma|\delta) = W^{FD}(\hat{\theta}^{FB})$. By contrast, if $\hat{\theta}^s(\Gamma|\delta) \neq \hat{\theta}^{FB}$, it holds that $W^s(\Gamma|\delta) < W^{FD}(\hat{\theta}^{FB})$.*

In analogy to Lemma 1, Lemma 2 says that in order to determine whether a grading rule

is first best optimal, it suffices to analyze the critical type induced by it. Here, as before, a rule which induces any other critical type is not first best optimal. Thus, also here, $\hat{\theta}^{FB}$ is called the first best critical type. In the following, I restrict my analysis to first best optimal rules.

5 Optimal grading with salience

My focus lies on first best optimal rules for a given degree of salience $1 - \delta$. Therefore, I restrict my analysis to exploring whether $\hat{\theta}^s(\Gamma|\hat{\theta}^{FB}, \delta) = \hat{\theta}^{FB}$ has a solution. In words: given that employers believe the critical type to be the first best critical type, is there a grading rule which indeed induces the first best critical type? Such a grading rule is first best optimal.

In order to approach this question, I start by considering the case where the students' degree of salience is extremely high. For $\delta \rightarrow 0$, students take into account only the most salient states whereas all payoffs not contained in these states are ignored. To see this, write

$$\omega_{s_{gh}}(\Gamma|\phi, \delta) = \frac{\delta^{\sigma_{s_{1H}}(\Gamma|\phi) - \sigma_{s_{gh}}(\Gamma|\phi)}}{\sum_{s_{st} \in S(\Gamma)} \delta^{\sigma_{s_{1H}}(\Gamma|\phi) - \sigma_{s_{st}}(\Gamma|\phi)} \pi_{st}(\Gamma)} \quad \forall s_{gh} \in S(\Gamma).$$

By the ordering property, the most salient states are s_{1H} and s_{H1} for a given rule. For these states, the weighting function becomes $\omega_{s_{1H}}(\cdot) = \omega_{s_{H1}}(\cdot) = (\pi_{s_{1H}}(\cdot) + \pi_{s_{H1}}(\cdot))^{-1}$. All other states are not taken into account at all, that is, $\omega_{s_{gh}}(\cdot) = 0$. The critical type under a full disclosure rule can then be written as

$$\lim_{\delta \rightarrow 0} \hat{\theta}^s(\Gamma^N|\phi, \delta) = (q^N - q^1) \frac{\tau_N - \tau_1}{\tau_N + \tau_1}.$$

The following can be shown.

Lemma 3. *When results are fully disclosed, student effort is inefficiently high for an extremely high degree of salience. That is, $\lim_{\delta \rightarrow 0} \hat{\theta}^s(\Gamma^N|\delta) > \hat{\theta}^{FB}$.*

Lemma 3 states that under a full disclosure rule, too many students choose \mathcal{L}_e when their degree of salience is very high. To give an intuition for the result, consider the overweighting of the high effort lottery \mathcal{L}_e relative to the low effort lottery \mathcal{L}_0 . The choice of \mathcal{L}_e attaches too much weight to grade $C_1(\Gamma^N)$ in state s_{N1} (recall that $H = N$ for full disclosure), in which case high effort becomes less attractive as compared to expected utility maximization. However, it also attaches too much weight to the best grade $C_N(\Gamma^N)$ in state s_{1N} . From this point of view, high effort becomes more attractive. The analogue holds for low effort:

in state s_{N1} , too much weight is placed on the highest grade, low effort in this case becomes more attractive. On the contrary, it becomes less attractive in state s_{1N} . Students compare both lotteries, and the critical value is determined by the perceived difference in valuations between employing high effort and low effort, see (A4). Thus, if state s_{N1} is relatively more overweighted than state s_{1N} , the critical type decreases as compared to expected utility maximization. Otherwise, the critical type increases. Lemma 3 then basically proves that in sum, state s_{1N} is relatively more overweighted. This is because the initial state probabilities differ by (A1). In particular, the monotone likelihood ratio property requires that $\pi_{1N}(\Gamma^N) > \pi_{N1}(\Gamma^N)$. Thus, although the weighting factor by which state probabilities are overweighted is the same for both states, initial probabilities are not.

The lemma has an important implication: since, for high degrees of salience, students focus only on the most salient states, the only grades that are taken into account are the grade with the lowest value and that with the highest value. In the case of full disclosure, by Lemma 3, this induces students to employ too much effort. Values of the extreme grades $C_1(\cdot)$ and $C_N(\cdot)$ equal the lowest and highest score respectively, that is, $v_1(\Gamma^N|\phi) = q^1$ and $v_H(\Gamma^N|\phi) = q^N$. Thus, any other rule with the same values for the lowest and highest grades yields the same limit result, meaning that these rules also induce students to employ too much effort for very high degrees of salience. This is the case for all grading rules with $\gamma_1 = q^1$ and $\gamma_{H-1} = q^{N-1}$. Obviously, it requires that the rule has at least two cuts, i.e. $H \geq 3$.

Further, from Proposition 1, every rule different from the full disclosure rule induces inefficiently low effort if students maximize expected utility, i.e. if $\delta = 1$. This is the case for any employer belief, thus also for the belief that the critical type is the first best critical type. As a result, since the critical type is continuous in the degree of salience, there exists some $\delta < 1$ for which a rule of this form – $\gamma_1 = q^1$ and $\gamma_{H-1} = q^{N-1}$ – is first best optimal. This insight provides an explanation for the occurrence of coarse grading rules and shall therefore be stated separately.

Proposition 2. *For every $H \geq 3$, Ψ_H contains at least one rule that is first best optimal for some $\delta \in (0, 1)$.*

Proposition 2 demonstrates that the focus on salient grades can explain the occurrence of a wide variety of grading rules across different coarseness categories. Grading rules which exhibit the property that the worst and best score can be detected unambiguously are first best optimal for some degree of salience. The proposition however makes no statement on the existence of first best optimal rules for different degrees of salience. Do first best optimal grading rules exist for all degrees of salience, i.e. for all δ ?

Because a grading rule is a subset of Ω , its entries are rational numbers. This implies that the first best outcome may not exactly be restored by a deterministic rule, even if some first best optimal probabilistic counterpart exists. Thus, in order to be able to analyze global first best optimality, it is necessary to extend the set of rules to probabilistic ones. A probabilistic rule is a grading rule which assigns to certain scores different grades with positive probability. As will be shown, to derive a general optimality result, it suffices to take into account grading rules which are probabilistic only to a small degree in the sense that probabilistic assignment to grades is necessary only for single scores. Moreover, it suffices that these scores are assigned to at most two different grades with positive probability. All other scores continue to be assigned grades deterministically. The next proposition proves this. Before, slightly more general rules are introduced.

Extended set of rules: probabilistic rules. I strive to generate continuity of the critical type both within and across coarseness categories H . For this purpose, define \mathcal{G} as the set of probabilistic rules with the following properties: every score q^n which is not placed on a cut, that is, $q^n \neq \gamma_h$ for any h , is deterministically awarded a grade, like before. All other scores may be assigned to grades stochastically. In particular, denote by r_h a probability vector, with its elements summing up to one. More specifically, an element r_h^g denotes the probability with which score $q^n = \gamma_h$ is assigned to grade $C_g(\cdot)$. A rule in \mathcal{G} is then characterized by the set $G^H = \{\gamma_1, \gamma_2, \dots, \gamma_h, \dots, \gamma_H, r_1, \dots, r_H\}$, with $\gamma_H = q^N = 1$, as before. If r_h is such that $r_h^h = 1$ for all h , the rule equals the respective deterministic rule. For instance, $\tilde{G}^2 = \{\gamma_1, q^N, (1, 0), (0, 1)\}$ is the Pass/Fail rule $\tilde{\Gamma}^2 = \{\gamma_1, q^N\}$.

As turns out, it suffices to consider rules which have that (a) at most one r_h contains elements different from 1 and 0, and (b) at most two elements of r_h are different from 0, and these are neighbouring elements. In words, these rules are probabilistic only to a small degree: at most one score is assigned to at most two grades with positive probability. An example for an according probabilistic rule is $\tilde{G}^3 = \{\gamma_1, \gamma_2, q^N, (r, 1-r, 0), (0, 1, 0), (0, 0, 1)\}$. Here, all scores smaller than γ_1 are assigned grade $C_1(\tilde{G}^3)$, the score which coincides with the cut γ_1 is assigned the lowest grade $C_1(\tilde{G}^3)$ with probability r and $C_2(\tilde{G}^3)$ with probability $1-r$, all scores higher than γ_1 but lower than or equal to γ_2 are assigned grade $C_2(\tilde{G}^3)$ with certainty and all higher scores are assigned the best grade $C_3(\tilde{G}^3)$.

The following Proposition makes a statement about the existence of first best optimal rules for varying degrees of salience.

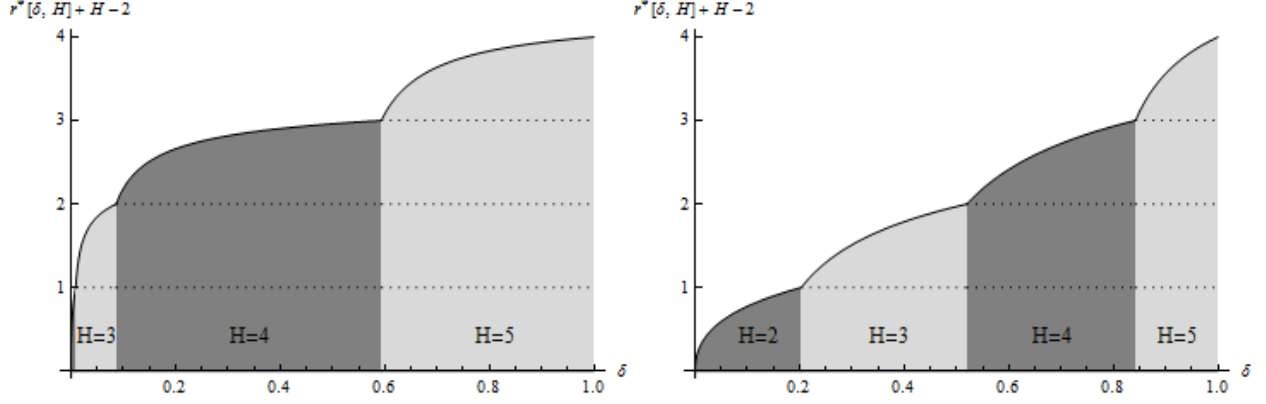
Proposition 3. *For any degree of salience $1-\delta$ for which the focus on salient grades induces inefficiently high effort when results are fully disclosed, i.e. for which $\hat{\theta}^s(\Gamma^N|\delta) \geq \hat{\theta}^{FB}$, a first best optimal rule in \mathcal{G} exists.*

Proposition 3 proves that first best optimal rules exist for each degree of salience for which students employ inefficiently high effort levels under a full disclosure rule. The idea of the proof is as follows: the rule which does not disclose any information to the market induces no student to employ high effort, independent of his degree of salience and employers' beliefs. It is then a strictly dominant strategy to choose low effort, meaning that the critical type is zero. Probabilistic rules generate continuity between the critical types for a no-disclosure policy and a full disclosure rule. To construct this effect, consider the following class of deterministic rules from which probabilistic rules are derived: for any coarseness category H , the grading rule $\tilde{\Gamma}^H$ is such that scores are fully disclosed up to a certain threshold. This threshold is q^{H-1} . All other scores are assigned to a single grade. That is, $\tilde{\Gamma}^2 = \{q^1, q^N\}$, $\tilde{\Gamma}^3 = \{q^1, q^2, q^N\}$, $\tilde{\Gamma}^4 = \{q^1, q^2, q^3, q^N\}$ and so forth. Then, continuity between the grading rules $\tilde{\Gamma}^H$ and $\tilde{\Gamma}^{H+1}$ is generated by implementing a probabilistic rule in which score q^H is awarded grade $C_H(\cdot)$ with probability $r \in (0, 1]$ and $C_{H+1}(\cdot)$ with probability $1 - r$.

For instance, consider $\tilde{\Gamma}^2$ and $\tilde{\Gamma}^3$. q^2 is awarded either grade $C_2(\cdot)$ (with probability r) or $C_3(\cdot)$ (with probability $1 - r$). If it is assigned to $C_2(\cdot)$ with probability one, the critical type is the same as under the grading rule $\tilde{\Gamma}^3$, because the second lowest score q^2 is unambiguously identified by employers. On the other hand, when it is almost surely awarded grade $C_3(\cdot)$, the critical type limits the one under a grading rule $\tilde{\Gamma}^2$. In this case, the second lowest score q^2 is awarded almost certainly the same grade as all higher scores. According probabilistic rules generate continuity of the critical type for all combinations of H and $H + 1$. As a result, whenever the focus on salient grades induces inefficiently high effort, some rule of this kind is first best optimal.²⁰

For a given salience function $\sigma_{s_{gh}}(\cdot) = (|v_g(\cdot) - v_h(\cdot)|) / (|v_g(\cdot) + v_h(\cdot)|)$ and given probability vectors p_e, p_0 , an example is depicted in figure 2 for $N = 5$. In particular, the picture shows the probability r , denoted by $r^*(\delta, H)$, which, for the described class of rules then yields first best optimality for given degrees of salience and coarseness categories H . For instance, in figure 2b, for the approximate interval $(0.2, 0.54]$, a rule with three distinct grades ($H = 3$), where q^1 is awarded grade $C_1(\cdot)$, q^2 is awarded grade $C_2(\cdot)$ with probability $r^*(\delta, H)$ and

²⁰ Also, note that probabilistic rules in \mathcal{G} can be constructed such that the critical types of deterministic rules within a coarseness category are linked. To see how, consider the following example: let there be two rules of the same coarseness category H , and assume they are identical with respect to $H - 1$ elements, and differ in one single element where the difference between these two elements is one score, i.e. $1/N$. For instance, let the coarseness category be $H = 2$ and consider grading rules $\Gamma^{2'} = \{q^n, q^N\}$ and $\Gamma^{2''} = \{q^{n+1}, q^N\}$. Now consider the following probabilistic rule: the score q^{n+1} is assigned grade $C_1(\cdot)$ with probability r and $C_2(\cdot)$ with probability $1 - r$, while all other scores are graded according to the initial deterministic rules, that is, q^1 to q^n are assigned to $C_1(\cdot)$ and q^{n+2} to q^N are assigned to $C_2(\cdot)$. Let this new continuous rule be denoted $G^2(r)$. Then, the critical type can be written as a function of this new grading rule, i.e. $\hat{\theta}^s(G^2(r)|\phi, \delta)$. When the probability r is zero, this critical type equals the critical type for rule $\Gamma^{2'}$. When r is one, it equals the critical type for rule $\Gamma^{2''}$ for a given degree of salience and belief.



(a) Scores are distributed according to binomial distributions $B(4, \alpha)$ with $\alpha = 0.7$ if the lottery choice is \mathcal{L}_e and $\alpha = 0.4$ if it is \mathcal{L}_0 . (b) The pmfs p_e^n and p_0^n are represented by the function $1/5(1 + \beta(2n - 6)/4)$ with $\beta = 1$ if the lottery choice is \mathcal{L}_e and $\beta = 0$ if it is \mathcal{L}_0 .

Figure 2: An example for first best optimal rules for $N = 5$ and $\sigma_{s_{gh}} = (|v_g - v_h|)/(v_g + v_h)$. $r^*(\delta, H)$ denotes the probability r , which, for the described class of rules yields first best optimality for given degrees of salience and coarseness categories H .

$C_3(\cdot)$ with probability $1 - r^*(\delta, H)$ while all other scores are assigned $C_3(\cdot)$, is first best optimal. Here, first best optimal rules exist for all degrees of salience.

Note that the constructed probabilistic rules well capture different degrees of grade inflation: higher scores are assigned equal grades and can therefore not be distinguished by employers. The less cuts (the smaller H), the more severe is grade inflation. By Proposition 3, some degree of grade inflation is first best optimal for any degree of salience $1 - \delta$ for which full disclosure yields inefficiently high effort levels. The example displayed in Figure 2 further suggests that higher degrees of salience – lower δ – account for higher degrees of grade inflation.

6 Discussion

It was assumed that students of the same institutions are equal with respect to (a) the probability distributions and (b) their inherent competitiveness (their δ). Selection in schools may justify these assumptions to a certain degree, albeit not fully. In the following, I am going to scetch the implications of deviations.

6.1 Different ability levels

First, consider students with different ability levels. In particular, let there be M student groups t_1, t_2, \dots, t_M , whose probability distributions have different supports. That is, students of group t_m score between \underline{q}_m and \bar{q}_m , $\underline{q}_m < \bar{q}_m$ and $\underline{q}_m, \bar{q}_m \in \Omega$.

Denote by ϕ_m the belief employers hold about the critical cost type in group m and let $\phi^t = (\phi_1, \phi_2, \dots, \phi_M)$. The market value of a certificate v_h is given by

$$v_h(\Gamma|\phi^t, \delta) = \frac{\sum_{m=1}^M \sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} q^i (\phi_m p_e^{im} + (1 - \phi_m) p_0^{im})}{\sum_{m=1}^M \sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} (\phi_m p_e^{jm} + (1 - \phi_m) p_0^{jm})},$$

where p_e^m and p_0^m are probability vectors for group m with *zero* elements outside the support interval. If students are disparate in the sense that their supports do not overlap, my results from the previous analysis apply accordingly: every group can be analyzed separately, and cuts are chosen such that first best optimality is restored. There is no benefit from further distorting information. As an example, when all groups are optimally incentivized by employing a Pass/Fail rule, the now first best optimal rule is one with M cuts.

Similarly, if supports overlap and the optimal cuts are placed outside the overlapping area, the optimal solution can easily be restored. If not so, the decision maker has to trade off the effect of merging the different cuts against that of moving cuts outside the support intervals of students of different abilities. First best optimality is then typically not restored. I leave the in-depth analysis of these effects for further research.

6.2 Different degrees of salience

A related question is, in how far does the behavior of students with different degrees of salience affect the behavior their classmates? Assume the school has chosen a coarse rule $\tilde{\Gamma}$ which in equilibrium implemented just the right effort levels given the average δ , denoted by $\tilde{\delta}$. More specifically, $\hat{\theta}^s(\tilde{\Gamma}|\hat{\theta}^{FB}, \tilde{\delta}) = \hat{\theta}^{FB}$. Consider a single (atomistic) student whose degree of salience is lower, namely $\hat{\delta}$, therefore $\hat{\theta}^s(\tilde{\Gamma}|\hat{\theta}^{FB}, \hat{\delta}) < \hat{\theta}^{FB}$.

Consider by contrast an institution where the average δ is $\hat{\delta}$ but who has employed the same, then not first best optimal, rule $\tilde{\Gamma}$. It can be verified that market values for each grade are lower, the reason being that employers foresee the reduced willingness to work. As a result, my analysis predicts that for a given grading rule, students with a lower degree of salience employ more effort in the presence of classmates with higher degrees of salience.

7 Concluding comments

I model grade obsession as a focus on salient grades using the framework of BGS. It is analyzed how an institution seeking to maximize its students' expected utility should optimally design a grading rule when students are grade obsessed. It is shown that an extreme grade obsession induces students to employ inefficiently high effort levels and that a coarse grading rule can counterbalance this effect. Thus, the common intuition that coarsening grades reduces incentives applies. Moreover, it is demonstrated how first best optimal behavior can always be restored by the use of a coarse grading rule.

Grade inflation is a different means of coarsening grades. At least some elite schools are known to factually have abolished lower grades. By assuming that grade obsession is present first and foremost in student bodies of very selective colleges, the model provides an explanation for the occurrence of coarse grading and grade inflation in these institutions. The phenomenon of grade inflation is a hotly debated topic world wide. It is mostly viewed as being negative and politicians often request to implement measures to end it. The present article on the other hand points to the positive aspects of grade inflation.

This suggests to further analyze the relationship between the eagerness of students and grading scales of colleges, e.g. by empirically testing the predictions derived above. And of course, research need not stop here, as phenomena such as promotion obsession and reputation obsession are likely to have similar roots. Solutions there may be less obvious to find, but substantial progress seems to be possible.

A Appendix

Proof of Lemma 1. Suppose $\hat{\theta}^{FB} = \hat{\theta}^r(\Gamma)$ for some Γ and define $\bar{\theta} := \hat{\theta}^{FB} = \hat{\theta}^r(\Gamma)$. It is to show that $W^{FD}(\bar{\theta}) = W^r(\Gamma)$. Define $V_e^{FD} := E[Q|\mathcal{L}_e]$ and $V_0^{FD} := E[Q|\mathcal{L}_0]$. Then $W^{FD}(\bar{\theta}) = W^r(\Gamma)$ is given if and only if

$$\bar{\theta}V_e^{FD} + (1 - \bar{\theta})V_0^{FD} = \bar{\theta}V_e(\Gamma|\cdot) + (1 - \bar{\theta})V_0(\Gamma|\cdot). \quad (8)$$

Using $V_e(\cdot)$ and $V_0(\cdot)$ as given in equation (1) and $v_h(\Gamma|\phi)$ as in equation (3) transforms equation (8) to

$$\sum_{i=1}^N q^i(\bar{\theta}p_e^i + (1 - \bar{\theta})p_0^i) = \sum_{h=1}^H \frac{\sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} q^i(\phi p_e^i + (1 - \phi)p_0^i)}{\sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} \phi p_e^i + (1 - \phi)p_0^i} \sum_{i=\gamma_{h-1}N+1}^{\gamma_h N} \bar{\theta}p_e^i + (1 - \bar{\theta})p_0^i.$$

Since by definition of $\hat{\theta}^r(\Gamma)$ it holds that $\bar{\theta} = \phi$, the condition becomes

$$\sum_{i=1}^N q^i(\phi p_e^i + (1 - \phi)p_0^i) = \sum_{h=1}^H \sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} q^i(\phi p_e^i + (1 - \phi)p_0^i),$$

which is clearly given. By contrast, consider some Γ for which $\hat{\theta}^r(\Gamma) \neq \hat{\theta}^{FB}$. Then, $W^{FD}(\hat{\theta}^{FB}) = W^r(\Gamma)$ if and only if

$$\hat{\theta}^{FB}V_e^{FD} + (1 - \hat{\theta}^{FB})V_0^{FD} = \hat{\theta}^r(\Gamma)V_e(\Gamma|\cdot) + (1 - \hat{\theta}^r(\Gamma))V_0(\Gamma|\cdot). \quad (9)$$

Transforming accordingly and using $\phi = \hat{\theta}^r(\Gamma)$ gives that equation (9) is given if and only if

$$\begin{aligned} \sum_{i=1}^N q^i(\hat{\theta}^{FB}p_e^i + (1 - \hat{\theta}^{FB})p_0^i) &= \sum_{h=1}^H \sum_{i=1+\gamma_{h-1}N}^{\gamma_h N} q^i(\hat{\theta}^r(\Gamma)p_e^i + (1 - \hat{\theta}^r(\Gamma))p_0^i) \\ &= \sum_{i=1}^N q^i(\hat{\theta}^r(\Gamma)p_e^i + (1 - \hat{\theta}^r(\Gamma))p_0^i). \end{aligned}$$

This is given only if $\hat{\theta}^r(\Gamma) = \hat{\theta}^{FB}$.

Finally,

$$W^r(\Gamma) = \sum_{i=1}^N q^i(\hat{\theta}^r(\Gamma)p_e^i + (1 - \hat{\theta}^r(\Gamma))p_0^i) - \frac{1}{2}\hat{\theta}^r(\Gamma)^2$$

maximizes at $\hat{\theta}^r(\Gamma) = \hat{\theta}^{FB}$. Thus, full disclosure is expected utility optimal. \square

Proof of Proposition 1. Define

$$\begin{aligned}\Delta^{\Gamma^H} &:= \hat{\theta}^r(\Gamma^H|\phi) - \hat{\theta}^{FB} \\ &= \sum_{h=1}^H v_h(\Gamma^H|\phi) \sum_{i=1+\gamma_{h-1}N}^{\gamma_H} (p_e^i - p_0^i) - \sum_{n=1}^N q^n (p_e^n - p_0^n)\end{aligned}$$

It is to show that $\Delta^{\Gamma^H} < 0$ for all $H < N$ and $\Gamma^H \subset \Psi_H$. The proof is by induction.

First step: consider some rule $\tilde{\Gamma}^{N-1} \subset \Psi_{N-1}$ such that grade $C_t(\cdot)$ pools scores q^t and q^{t+1} .

$$\Delta^{\tilde{\Gamma}^{N-1}} = (v_t(\cdot) - q^t)(p_e^t - p_0^t) - (q^{t+1} - v_t(\cdot))(p_e^{t+1} - p_0^{t+1})$$

Defining $z^n = (\bar{\phi}p_e^n + (1 - \bar{\phi})p_0^n)$ and using $v_t(\cdot)$ as given in equation (3) gives

$$\begin{aligned}\Delta^{\tilde{\Gamma}^{N-1}} \frac{z^t + z^{t+1}}{q^{t+1} - q^t} &= (z^{t+1}(p_e^t - p_0^t) - z^t(p_e^{t+1} - p_0^{t+1})) \\ &= p_0^{t+t} p_e^t - p_0^t p_e^{t+1} = p_0^t p_0^{t+1} (\tau_t - \tau_{t+1}) < 0.\end{aligned}$$

Second step: consider some $\Delta^{\tilde{\Gamma}^H} \subset \Psi_H$ and assume $\Delta^{\tilde{\Gamma}^H} < 0$. Now consider some $\Delta^{\tilde{\Gamma}^{H-1}} \subset \Psi_{H-1}$ which pools two neighbouring grades $C_{t'}(\cdot)$ and $C_{t'+1}(\cdot)$ into a new grade C_s , all other grades remain unchanged.

$$\begin{aligned}\Delta^{\tilde{\Gamma}^{H-1}} - \Delta^{\tilde{\Gamma}^H} &= \hat{\theta}^r(\Gamma^{\tilde{H}-1}|\hat{\theta}^{FB}) - \hat{\theta}^r(\Gamma^{\tilde{H}}|\hat{\theta}^{FB}) \\ &= (v_s(\cdot) - v_{t'}(\cdot)) \sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} (p_e^i - p_0^i) - (v_{t'+1}(\cdot) - v_s(\cdot)) \sum_{i=1+\gamma_{t'}N}^{\gamma_{t'+1}N} (p_e^i - p_0^i)\end{aligned}$$

with

$$\begin{aligned}v_s(\cdot) - v_{t'}(\cdot) &= \frac{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j) + \sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i \sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i} \\ &= \frac{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i \sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i} \\ v_{t'+1} - v_s &= \frac{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i z^j (q^i - q^j) + \sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i \sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i} \\ &= \frac{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i \sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i}\end{aligned}$$

Therefore, $\Delta^{\tilde{\Gamma}^{H-1}} - \Delta^{\tilde{\Gamma}^H} < 0$ if and only if

$$\begin{aligned} & \frac{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} \sum_{j=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i z^j (q^i - q^j)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'+1}N} z^i} \left(\frac{\sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} (p_e^i - p_0^i)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i} - \frac{\sum_{i=1+\gamma_{t'}N}^{\gamma_{t'+1}N} (p_e^i - p_0^i)}{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i} \right) < 0 \\ \Leftrightarrow & \left(\frac{\sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} (p_e^i - p_0^i)}{\sum_{i=\gamma_{t'-1}N+1}^{\gamma_{t'}N} z^i} - \frac{\sum_{i=1+\gamma_{t'}N}^{\gamma_{t'+1}N} (p_e^i - p_0^i)}{\sum_{i=\gamma_{t'}N+1}^{\gamma_{t'+1}N} z^i} \right) < 0 \\ \Leftrightarrow & \sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} \sum_{j=1+\gamma_{t'}N}^{\gamma_{t'+1}N} p_0^j p_e^i - p_0^i p_e^j = \sum_{i=1+\gamma_{t'-1}N}^{\gamma_{t'}N} \sum_{j=1+\gamma_{t'}N}^{\gamma_{t'+1}N} p_0^i p_0^j (\tau_i - \tau_j) < 0, \end{aligned}$$

which is given due to the monotone likelihood ratio property. \square

Proof of Lemma 2. Analogous to the proof of Lemma 1. \square

Proof of Lemma 3. The problem can be written as a maximization problem where p_0 is being fixed. It is to show that there the maximum is not strictly positive. For technical convenience, we consider $\tau_n = 0$, i.e. $p_0^n = 0$ for some small n and $\tau_n = \tau_{n+1}$ for some n (in contradiction to (A1) where we assumed strictly increasing τ_n).

$$\begin{aligned} \max_{\tau} f(\tau) & := -(N-1) \frac{\tau_N - \tau_1}{\tau_N + \tau_1} + \sum_{n=1}^N n p_0^n (\tau_n - 1) \\ \text{s.t. } \tau_n & \geq \tau_{n-1} \geq 0 \text{ for all } n, \quad \sum_{n=1}^N p_0^n \tau_n = 1 \end{aligned}$$

τ_N can be written as a function of τ_{-N} . Then, for every $2 \leq n \leq N-1$,

$$\frac{\partial f(\tau_{-N})}{\partial \tau_n} = p_0^n (n - N + \frac{2(N-1)\tau_1}{(\tau_N + \tau_1)^2 p_0^N})$$

Possible solutions to the program are:

(a) $\tau_1 \rightarrow 0$. Define $E_e^n := \sum_{n=1}^N n p_e^n = \sum_{n=1}^N n p_0^n \tau_n$ and $E_0^n := \sum_{n=1}^N n p_0^n$. Then:

$$\lim_{\tau_1 \rightarrow 0} f(\tau) = -(N-1) + (E_e^n - E_0^n) < 0.$$

(b) $1 > \tau_1 > 0$. Since $\partial f(\tau_{-N})/\partial \tau_n$ is monotonically increasing in n , a candidate solution has the property that there exists some critical n , denoted by \hat{n} such that $\tau_{\hat{n}}^1 := \tau_1 = \tau_2 = \dots = \tau_{\hat{n}} < \tau_{\hat{n}+1} = \dots = \tau_N := \tau_{\hat{n}}^N$. Denote $p^{\hat{n}} := \sum_{i=\hat{n}+1}^N p_0^i$.

Then $\tau_{\hat{n}}^N = (1 - (1 - p^{\hat{n}})\tau_{\hat{n}}^1)/p^{\hat{n}}$ and

$$f(\tau) = -\frac{(N-1)(1 - \tau_{\hat{n}}^1)}{p^{\hat{n}}(\tau_{\hat{n}}^N + \tau_{\hat{n}}^1)} - (1 - \tau_{\hat{n}}^1) \sum_{i=1}^{\hat{n}} i p_0^i + (1 - \tau_{\hat{n}}^1) \frac{1 - p^{\hat{n}}}{p^{\hat{n}}} \sum_{i=\hat{n}+1}^N i p_0^i.$$

$f(\tau) < 0$ iff

$$-(N-1) + (\tau_{\hat{n}}^N + \tau_{\hat{n}}^1)(1-p^{\hat{n}}) \sum_{i=\hat{n}+1}^N ip_0^i - (\tau_{\hat{n}}^N + \tau_{\hat{n}}^1)p^{\hat{n}} \sum_{i=1}^{\hat{n}} ip_0^i < 0.$$

It holds that $\sum_{i=1}^{\hat{n}} ip_0^i > (1-p^{\hat{n}})$ and $\sum_{i=\hat{n}+1}^N ip_0^i < Np^{\hat{n}}$. Therefore, $f(\tau) < 0$ if

$$\begin{aligned} 1 - (\tau_{\hat{n}}^N + \tau_{\hat{n}}^1)(1-p^{\hat{n}})p^{\hat{n}} &> 0. \\ \Leftrightarrow 1 > (1-p^{\hat{n}})(1 + \tau_{\hat{n}}^1(2p^{\hat{n}} - 1)) \end{aligned}$$

which can be checked to hold for all $\tau_{\hat{n}}^1 \in (0, 1)$ and $p^{\hat{n}} \in (0, 1)$.

(c) $\tau_1 \rightarrow 1$ yields $\lim_{\tau_1 \rightarrow 1} f(\tau) = 0$. □

Proof of Proposition 2. From the analysis in the text, it follows that for any Γ^H , $H \geq 3$ with $\gamma_1 = q^1$ and $\gamma_{H-1} = q^{N-1}$, it holds that $\omega_{s_{1H}}(\cdot) = \omega_{s_{H1}}(\cdot) = (p_0^1 p_e^N + p_0^N p_e^N)^{-1}$ and $\omega_{s_{hg}}(\cdot) = 0$ for all other states, thereby:

$$\lim_{\delta \rightarrow 0} \hat{\theta}^s(\Gamma^H | \phi, \delta) = \lim_{\delta \rightarrow 0} \hat{\theta}^s(\Gamma^N | \phi, \delta),$$

which is strictly greater than $\hat{\theta}^{FB}$ by Lemma 3. From Proposition 1, these rules are strictly smaller than $\hat{\theta}^{FB}$ at $\delta = 1$. By continuity of $\hat{\theta}^s(\Gamma^H | \phi, \delta)$ in δ , there exists some δ for which $\hat{\theta}^s(\Gamma^H | \phi, \delta) = \hat{\theta}^{FB}$. □

Proof of Proposition 3. First, if no information is revealed, i.e. $H = 1$, then $\hat{\theta}^s(\Gamma^1 | \phi, \delta) = 0$ for all tuples (ϕ, δ) .

Some probabilistic rule in \mathcal{G} establishes continuity between any $\hat{\theta}^s(\Gamma^1 | \phi, \delta)$ and $\hat{\theta}^s(\Gamma^2 | \phi, \delta)$ for any $\Gamma^2 \subset \Psi_2$. Consider the following rule in Ψ_2 : $\tilde{\Gamma}^2 = q^1$ and the probabilistic rule in \mathcal{G} , $\tilde{G}^2 = \{q^1, q^N, (r, 1-r), (0, 1)\}$ with $r \in (0, 1]$. That is, score q^1 is awarded grade $C_1(\cdot)$ with probability r and $C_2(\cdot)$ with probability $1-r$. All other scores are awarded grade $C_2(\cdot)$.

The function $\hat{\theta}(\tilde{G}^2 | \phi, \delta)$ is clearly continuous in r . When $r = 1$, $\tilde{G}^2 = \tilde{\Gamma}^2$.

Also, when $r \rightarrow 0$, $\tilde{G}^2 \rightarrow \Gamma^1$, and $\hat{\theta}(\tilde{G}^2 | \phi, \delta) \rightarrow \hat{\theta}(\Gamma^1 | \phi, \delta) = 0$, as can be verified by inspection of equation (6). Then, all sum elements of equation (6) which contain $g = 1$ or $h = 1$ approach 0, while the value of grade $C_2(\cdot)$ limits the value of $C_2(\cdot)$ of Γ^0 . Further, the probability that grade $C_2(\cdot)$ occurs if \mathcal{L}_0 is chosen, is given by $\sum_{i=q^2}^{q^N} p_0^i + (1-r)p_0^1$ (analog presentation for \mathcal{L}_e), while the probability that grade $C_1(\cdot)$ occurs is rp_0^1 , which is clearly zero for $r \rightarrow 0$. As a result, the weighting functions as given in 5 also limit those of Γ^0 , i.e. 1.

The same procedure links all higher coarseness categories up to the full disclosure rule

Γ^N . More specifically, the new probabilistic rules are all derived from the deterministic rules $\tilde{\Gamma}^H = \{\gamma_1 = q^1, \gamma_2 = q^2, \dots, \gamma_{H-1} = q^{H-1}, \gamma_H = q^N\}$. Then, the rules $\tilde{G}^H = \{q^1, \dots, q^{H-1}, q^N, (1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 1, 0, 0), (0, \dots, 0, r, 1-r), (0, \dots, 1)\}$, $r \in (0, 1]$ establish the link between $\tilde{\Gamma}^{H-1}$ and $\tilde{\Gamma}^H$ for all H . The last step is the transition from $\tilde{\Gamma}^{N-1}$ to the full disclosure rule Γ^N .

By continuity of $\hat{\theta}^s(\tilde{G}^H|\phi, \delta)$ in r for any tuple (ϕ, δ) and any rule \tilde{G}^H , continuity is established between $\hat{\theta}^s(\Gamma^0|\hat{\theta}^{FB}, \delta)$ – student behavior when no information disclosure – and $\hat{\theta}^s(\Gamma^N|\hat{\theta}^{FB}, \delta)$ – student behavior for full disclosure. As a result, first best optimal rules exist for all δ for which $\hat{\theta}^s(\Gamma^N|\hat{\theta}^{FB}, \delta) \geq \hat{\theta}^{FB}$. \square

References

- Betts, J. R.**, “The Impact of Educational Standards on the Level and Distribution of Earnings,” *The American Economic Review*, 1998, *88* (1), 266–275.
- **and F. Grogger**, “The impact of grading standards on student achievement, educational attainment, and entry-level earnings,” *Economics of Education Review*, 2003, *22* (4), 343–352.
- Boleslavsky, R. and C. Cotton**, “Grading standards and education quality,” *American Economic Journal: Microeconomics*, Forthcoming.
- Bordalo, P., N. Gennaioli, and A. Shleifer**, “Salience Theory Of Choice Under Risk,” *The Quarterly Journal of Economics*, 2012, *127* (3), 1243–1285.
- , – , **and** – , “Salience and Consumer Choice,” *Journal of Political Economy*, 2013, *121* (5), 803–843.
- Cameron, J. and R. Eisenberger**, “Detrimental Effects of Reward, Reality or Myth?,” *American Psychologist*, 1996, *51* (11), 1153–1166.
- Chan, W., L. Hao, and W. Suen**, “A Signaling Theory of Grade Inflation,” *International Economic Review*, 2007, *48* (3), 1065–90.
- Costrell, R. M.**, “A Simple Model of Educational Standards,” *The American Economic Review*, 1994, *84* (4), 956–971.
- Covington, Martin P.**, “Goal Theory, Motivation, and School Achievement: An Integrative Review,” *Annual Review of Psychology*, 2000, *51*, 171–200.
- Dubey, P. and J. Geanakoplos**, “Grading exams: 100, 99, 98,... or A, B, C?,” *Games and Economic Behavior*, 2010, *69* (1), 72–94.
- Figlio, D. N. and M. E. Lucas**, “Do high grading standards affect student performance?,” *Journal of Public Economics*, 2004, *88* (9–10), 1815–1834.
- Harbaugh, R. and E. Rasmusen**, “Coarse Grades: Informing the Public by Withholding Information,” *working paper*, 2013.
- Healy, C. and S. Rojstaczer**, “Where A Is Ordinary: The Evolution of American College and University Grading, 1940-2009,” *Teachers College Record*, 2012, *114* (7), 1–23.

- Kőszegi, B. and A. Szeidl**, “A Model of Focusing in Economic Choice,” *The Quarterly Journal of Economics*, 2013, 128 (1), 53–104.
- Lizzeri, A.**, “Information revelation and certification intermediaries,” *The RAND Journal of Economics*, 1999, pp. 214–231.
- Mas-Colell, A., M. D. Whinston, and J. G. Green**, *Microeconomic Theory*, Oxford University Press, 1995.
- Oettinger, G. S.**, “The Effect of Non-linear Incentives on Performance: Evidence from ”ECON 101”,” *The Review of Economics and Statistics*, 2002, 84 (3), 509–517.
- Ostrovsky, M. and M. Schwarz**, “Information Disclosure and Unraveling in Matching Markets,” *American Economic Journal: Microeconomics*, 2010, 2 (2), 34–63.
- Rayo, L. and I. Segal**, “Optimal Information Disclosure,” *Journal of Political Economy*, 2010, 118 (5), 949–987.
- Tversky, A.**, “Intransitivity of Preferences,” *Psychological Review*, 1969, 76 (1), 31–48.
- Zubrickas, R.**, “Optimal Grading,” *International Economic Review*, Forthcoming.