Article:

Shimoji, Makoto orcid.org/0000-0002-7895-9620 (2022) Setting an exam as an information design problem. International Journal of Economic Theory. ISSN: 1742-7363

https://doi.org/10.1111/ijet.12368

Received: 25 August 2021 | Revised: 27 November 2022 | Accepted: 5 December 2022

DOI: 10.1111/ijet.12368

ORIGINAL ARTICLE

International Journal of Economic Theory IJET WILEY

Setting an exam as an information design problem

Makoto Shimoji

Department of Economics and Related Studies, University of York, Heslington, York, UK

Correspondence
Makoto Shimoji, Department of Economics and Related Studies, University of York, Heslington, York YO10 5DD, UK.
Email: makoto.shimoji@york.ac.uk

Funding information
Department of Economics and Related Studies, University of York, Grant/Award Number: RIS-12

Abstract

We take a teacher's exam-setting task as an information design problem. Specifically, the teacher chooses a conditional distribution of grades given students' types. After observing their exam results, each student updates her belief regarding her type via Bayes' rule and chooses an action. Students' reactions to the same exam result could be different, depending on their heterogeneous prior beliefs. The teacher's objective is to persuade students to take a certain action (e.g., applying to universities), which some may not choose without an exam. The teacher adopts different grade distributions, depending on the teacher's and the students' heterogeneous prior beliefs.

KEYWORDS

Bayesian persuasion, grade distribution, heterogeneous beliefs, multiple receivers

JEL CLASSIFICATION

C72, D83, I21

> … a pessimist gets nothing but pleasant surprises, an optimist nothing but unpleasant. Fer-de-Kance, Rex Stout (2008, p. 3).[1]

[1]The novel was originally published in 1934.

1 | INTRODUCTION

In this paper, we analyze a strategic aspect of exam grade distributions. There are mainly two strategic contexts, depending on who uses the information. First, schools attempt to control the informational contents of grade distributions to influence the behavior of those who use the information to assess students' types (e.g., potential employers). Students' grades signal their types, and potential employers look at applicants' transcripts to distinguish them. Therefore, it could be schools' interest to control grade distributions and their informational contents. Second, exam grade distributions would affect students' decisions, ex ante or ex post. One theoretical contribution is Dubey and Geanakoplos (2010) where grading is rather taken as games of status, that is, students care about their ranking and choose their effort accordingly.[2]

1.1 | Effect of information on students' decisions

Our focus is on the latter. More specifically, we are interested in an effect of information given by not only the grades themselves but also the grade distributions on students' future decisions. The information students receive through their grades and the grade distributions would naturally influence their beliefs on their types and thus their future decisions not only in the short run (e.g., effort for next exams) but also in the long run (e.g., which universities to apply to).[3] The importance of the informational contents of the grade distributions to students cannot be understated.

Such effects even in the short run have been documented. Using the data from introductory economics courses, Oettinger (2002) documented how students' performance on final exams is influenced by their performance on earlier exams:

> … [S]tudents who are closer to a grade boundary going into the final exam tend to perform better on the final exam, after controlling for prefinal exam performance. (p. 509)

Similar arguments can also be applied to different contexts. For example, consider journal submissions where "grades" are binary (acceptance or rejection). If one receives rejections from top journals, she may decide not to submit her future manuscripts to them. That is, while editors would still like to encourage future submissions even when they reject current manuscripts, editorial decisions could influence potential contributors' future submission decisions. Journals thus may adjust their responses to different authors.[4]

[2]Students' effort choice may also depend on other factors, such as cultural differences (Gneezy et al., 2019) and exams' stakes (Schlosser et al., 2019).

[3]For example, see Zafar (2011) and Stinebrickner and Stinebrickner (2012) who examined how students update their beliefs regarding abilities and outcomes.

[4]Card and Della Vigna (2020) examined the data of the editorial processes from four top journals. One of their findings indicates "the observed penalty on expected citations at the R&R stage more likely reflects a higher bar for prolific authors than a discount for the fact that their papers get too many citations, conditional on quality" (p. 197).

1.2 | Our approach

We consider a teacher's task to set an exam as an information design problem. Specifically, we assume that the teacher chooses a conditional distribution of exam grades given the students' types. The teacher is aware that the students' realized grades would affect their later decisions. We have two key assumptions. One is that not only the teacher but also the students are not fully aware of the students' underlying types. This implies that the students learn their types through the exam. We also assume that while the students' preferred actions depend on their types, the teacher wants the students to choose a certain action (e.g., applying to universities) *independently of their types*. These assumptions make the teacher's task challenging. How does the teacher design the grade distribution?

The difficulty the teacher faces stems from the fact that the students' decisions depend on their realized grades, which in turn depend on their types; that is, the students' decisions would indirectly depend on their types. The students' choices of actions would therefore depend on how they interpret their grades. For example, if the grades fully reveal the students' types, their choices of actions will entirely depend on their realized grades. Although the students can choose the best actions for their own types in this case, this is not necessarily the best scenario from the teacher's point of view since the realized grades may discourage the students take the action which the teacher wants them to take.

As noted above, we assume that the teacher chooses a conditional distribution of grades given the students' types. Thus, it is not necessarily the case that students' realized grades fully reveal their types. The grades the students receive would still reveal some information about their types, and thus influence their decisions. Through the grade distribution, the teacher attempts to encourage students to choose a certain action. The key is that students' decisions are based on their *updated* beliefs about their types after receiving their grades. Since the belief update depends not only on the realized grades but also the grade distribution set by the teacher, the teacher can influence students' decisions by appropriately choosing a grade distribution.

Our model is based on the notion of *Bayesian persuasion* by Kamenica and Gentzkow (2011). In their framework, the sender chooses a distribution of signals conditional on the states. Once the receiver observes the signal, she updates her belief regarding the state and chooses an action, which would affect both the sender's and the receiver's payoffs. The sender maximizes her expected payoffs by appropriately choosing the signal distribution and thus by influencing the receiver's behavior via the belief update. Shimoji (2022) further extended their framework by allowing heterogeneous beliefs and multiple receivers with the assumption that there is no strategic interaction among receivers. We adopt the latter framework.[5]

We employ a simple setting—one teacher (sender), two students (receivers) $\{1, 2\}$, two actions $\{a^L, a^H\}$, two types $\{\theta^L, \theta^H\}$, and two grades (signals) $\{fail, pass\}$.[6] Despite its simplicity, the binary signal model is still relevant. In the UK, although individual grades at universities use numerical scales in general, students' final degree classification (such as "first," "upper second (2:1)," "lower second (2:2)," and "third" which are based on students' overall grades)

[5]One theoretical contribution regarding heterogeneous beliefs is Alonso and Câmara (2016a) which considers the case of one receiver. Regarding multiple receivers, Alonso and Câmara (2016b) and Chan et al. (2019) consider voting models with multiple voters where the sender's objective is to maximize the probability that one proposal (out of two) is accepted via voting. It is assumed that all voters share the same initial belief in Alonso and Câmara (2016b) and Chan et al. (2019). Hernández and Neeman (2022) consider resource allocation problems with multiple receivers (with the common initial belief).

[6]It is possible to analyze more involved settings with the expense of computational complexities.

matters as reported by Coughlan (2018), which in turn suggests that it could be viewed as a binary signal model:

> [H]ead of the Institute of Student Employers ... says about two-thirds of graduate recruiters expect a 2:1 as a minimum requirement. ... many are not that bothered whether students get a first or a 2:1, ...[7]

1.3 | What we show

We assume that a student wants to choose (i) a^L if her type is θ^L and (ii) a^H if her type is θ^H while the teacher would like them to take a^H independently of their types. We also assume that student 1 believes that it is more likely that her type is θ^L while student 2 believes that it is more likely that her type is θ^H. Therefore, without an exam, student 1 chooses a^L and student 2 chooses a^H. As a benchmark, we show that the teacher would set a different exam for each student if she can do so.[8] However, the teacher needs to set the same exam for all students, prohibiting her from having personalized exams.

Given the grade distribution the teacher chooses, each student updates her belief regarding her type after observing the exam result via Bayes' rule. The teacher wants student 1 to choose a^H instead if she receives *pass*, which is possible only when *fail* can also be realized. This however implies that there is also a chance that each student receives *fail*, implying that not only student 1 but also student 2 may choose a^L after observing *fail*. This dilemma, due to the students' heterogeneous priors, clearly makes the teacher's task to set the exam challenging.

We show that there are two types of exams the teacher will set, depending on the teacher's and the students' initial beliefs about the students' types. In both exams, student 1 chooses (i) a^L after observing *fail* and (ii) a^H after observing *pass*. The difference stems from student 2's response to the exam results. The first is the exam we call "exam A"—student 2 also chooses a^L after observing *fail* and a^H after observing *pass*. The other is the exam we call "exam B"—student 2 chooses a^H independently of the exam result.

The comparison of exams A and B shows that for any set of the teacher's and the students' beliefs, exam A has a higher chance to produce *pass* for each type (Corollary). We also show that while student 1 is indifferent between exams A and B, student 2 strictly prefers exam A to exam B (Proposition 1). This is not because students have a higher chance to obtain *pass*, but because exam A provides more information about her type from student 2's point of view. We identify the condition under which the teacher sets either exam A or exam B (Proposition 2). The teacher's and the student's beliefs crucially influence the teacher's decision regarding which exam the teacher employs. These results cannot be obtained without heterogeneous beliefs.

In reality, growing concerns about grade inflation may prevent teachers from setting exams with higher averages (as in exam A) even if they find such exams appropriate. Such hesitations could be due to institutional reasons. For example, Butcher et al. (2014) discussed an anti-grade-inflation policy at Wellesley college. If exams with higher averages would never be desirable at all, such restrictions would have no adverse effect. In Section 5, we refer to recent

[7] "Does it matter what degree grade you get?" (https://www.bbc.co.uk/news/education-45939993).

[8] The teacher would not need to set an exam for student 2 since she would choose a^H without exam.

studies which documented potential positive effects of grade inflation. They imply that having an option to administer exams which may result in higher averages could be important. These empirical findings would suggest us to consider the choice of grade distributions as information design problems.

2 | PRELIMINARIES

We have one teacher and two students. Each student has one of two possible types, $\{\theta^L, \theta^H\}$, and chooses her action from the set $\{a^L, a^H\}$. Each student's payoff depends on the action she chooses and her type: $u_i(a_H|\theta^L) = u_i(a^L|\theta^H) = 0$ and $u_i(a^L|\theta^L) = u_i(a^H|\theta^H) = 1$. That is, each student wants to avoid a "mismatch." This implies that student $i \in \{1, 2\}$ chooses (i) a^H only if she believes that the probability of her being θ^H is at least $\frac{1}{2}$ and (ii) a^L only if she believes that the probability of her being θ^L is at least $\frac{1}{2}$. The teacher however has a different objective and would like both students to choose a^H *irrespective of their types*. The teacher's payoffs are (i) -1 if both students choose a^L, (ii) 0 if only one student chooses a^H, and (iii) 1 if both students choose a^H.

Each student $i \in \{1, 2\}$ has her initial belief, $p_i^0 \in (0, 1)$, representing the probability of her type being θ^H. We allow for heterogeneous beliefs and assume that $0 < p_1^0 < \frac{1}{2} < p_2^0 < 1$. The implications are (i) student 2 is more confident (that she is θ^H) than student 1 and (ii) no student knows her true type. For each student $i \in \{1, 2\}$, the teacher believes (i) that the two students' types are independent and (ii) that the probability of student i's type being θ^H is $p_T^i \in (0, 1)$. Note that the teacher's beliefs do not need to coincide with that of the students.

With the assumption on the students' initial beliefs, we have the following observation if there is no exam.

Observation. With their initial beliefs, student 1 chooses a^L and student 2 chooses a^H. The teacher's payoff is 0.

Our interest is in whether it is possible for the teacher to persuade each student to choose a^H after observing her exam result, and if so, how the teacher's exam choice would depend on the teacher's and the students' initial beliefs. We will show below how the teacher designs the exam.

2.1 | Timing

Before the students choose their actions, the teacher sets an exam from which each student receives one of the two results $\{fail, pass\}$. The result from the exam will partially or fully reveal each student's type. To focus on the informational aspect of the exam, we assume that the exam result stochastically depends on the student's type.[9] More specifically, the exam is a distribution over $\{fail, pass\}$ conditional on the student's type $\{\theta^L, \theta^H\}$, denoted by π.[10] We assume (i) that both students take the same exam and (ii) that π is observable to the students.

[9]We do not explicitly model the students' actions (e.g., effort) for the exam the teacher sets. One possible approach is to assume that students with the same initial belief choose the same action so that the choice of action is practically embedded in each student's type. It is straightforward to incorporate such settings in the current model.

[10]Two extreme examples are: (i) $\pi(pass|\theta^H) = \pi(fail|\theta^L) = 1$ perfectly reveals the student's type and (ii) $\pi(pass|\theta^H) = \pi(pass|\theta^L)$ provides no additional information.

Each student's choice of action depends on her exam result since she will receive some information about her type through the exam result, which would then affect her belief about her own type via Bayes' rule. Namely, after observing the exam result, $x \in \{fail, pass\}$, each student $i \in \{1, 2\}$ updates her belief regarding her type via Bayes' rule:

$$p_i(\theta^H|x) = \frac{\pi(x|\theta^H)p_i^0}{\pi(x|\theta^H)p_i^0 + \pi(x|\theta^L)\left(1 - p_i^0\right)} = 1 - p_i(\theta^L|x). \tag{1}$$

After updating her belief regarding her type, each student chooses her action from $\{a^L, a^H\}$. That π appears in the expression above implies that the teacher can influence the students' behavior via their belief update reflecting the results from the exam. The timeline is shown in Figure 1.

2.2 | Benchmark—One student

In this subsection, we consider the case with only one student. We demonstrate that choosing a conditional distribution of grades given students' types can be seen as a constrained maximization problem from the teacher's point of view. While maintaining the assumption for the student's payoffs, we assume that the teacher payoffs are (i) −1 if the student chooses a^L and (ii) 1 if the student chooses a^H.[11] For this subsection, we have $p_S^0 \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right)$ and $p_T \in (0, 1)$ as the initial beliefs for the student and the teacher, respectively. In Section 3.3, we demonstrate that the presence of two students makes the analysis more involved since we allow for heterogeneous beliefs.

First, if $p_S^0 > \frac{1}{2}$, the teacher has no incentive to administer an exam (with two distinct exam results) since the student chooses a^H even without the exam.

Suppose instead $p_S^0 < \frac{1}{2}$, in which case the student chooses a^L without the exam. Unlike the case of $p_S^0 > \frac{1}{2}$, this observation itself means that the teacher would employ the exam with the two exam results; that is, $\pi(x|\theta) \in (0, 1)$ for each $x \in \{pass, fail\}$ with at least one $\theta \in \{\theta^L, \theta^H\}$. After observing $x \in \{pass, fail\}$, the student chooses

$$\begin{Bmatrix} a^H \\ a^L \end{Bmatrix} \text{ only if } \underbrace{\frac{\pi(x|\theta^H)p_S^0}{\pi(x|\theta^H)p_S^0 + \pi(x|\theta^L)\left(1 - p_S^0\right)}}_{p_S(\theta^H|x)} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} \underbrace{\frac{\pi(x|\theta^L)\left(1 - p_S^0\right)}{\pi(x|\theta^H)p_S^0 + \pi(x|\theta^L)\left(1 - p_S^0\right)}}_{p_S(\theta^L|x)}$$

or equivalently

$$\begin{Bmatrix} a^H \\ a^L \end{Bmatrix} \text{ only if } \pi(x|\theta^L) \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \left(\frac{p_S^0}{1 - p_S^0}\right)\pi(x|\theta^H). \tag{2}$$

[11]See Example 2 of Shimoji (2022) where the payoffs for the student (receiver) are qualitatively the same. The current example is different from Example 2 of Shimoji (2022) since the payoffs for the teacher (sender) are different.

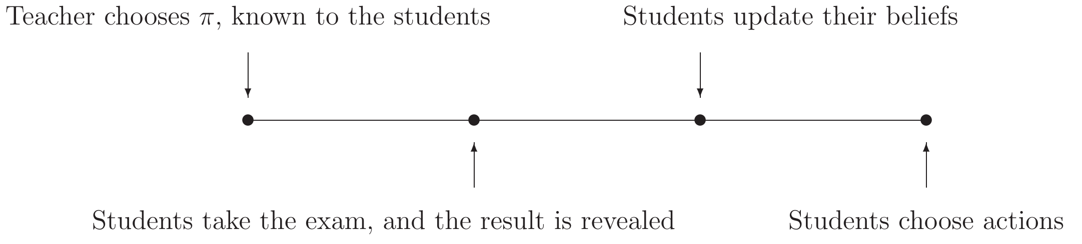

FIGURE 1 Timeline

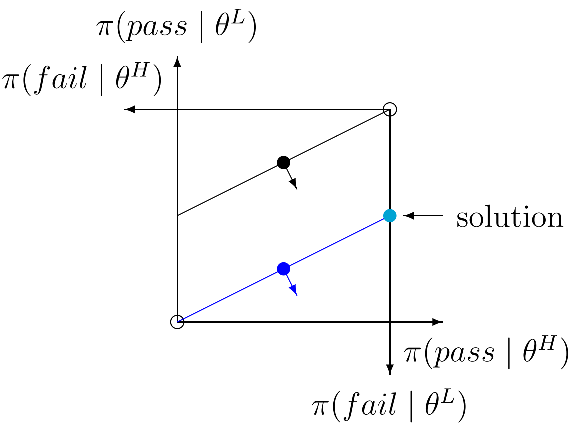

FIGURE 2 One student with $p_S^0 < \frac{1}{2}$ [Color figure can be viewed at wileyonlinelibrary.com]

Without loss of generality, our focus is on the case where the student chooses a^H after observing *pass* and a^L after observing *fail*. Figure 2 displays the box diagram such that (i) the length of each side is equal to one and (ii) any π can be represented as a point in the diagram.[12] The two inequalities (2) with $p_S^0 < \frac{1}{2}$ are displayed in the diagram; the line from the southwest origin corresponds to *pass* and the line from the northeast origin corresponds to *fail*.

Since the teacher would like the student to choose a^H, she chooses the highest possible values for $\pi(pass|\theta^H)$ and $\pi(pass|\theta^L)$ while maintaining the two inequalities in (2). That is, the teacher's expected payoff is higher as π moving towards the northeast origin.[13] The teacher thus chooses

$$\pi(pass|\theta^H) = 1, \qquad \pi(fail|\theta^H) = 0,$$
$$\pi(pass|\theta^L) = \frac{p_S^0}{1 - p_S^0} \in (0, 1), \qquad \pi(fail|\theta^L) = \frac{1 - 2p_S^0}{1 - p_S^0} \in (0, 1).$$

The student's updated belief after observing each exam result is $p_S(\theta^H|pass) = \frac{1}{2}$ and $p_S(\theta^H|fail) = 0$.

[12] That is, $\pi(pass|\theta) + \pi(fail|\theta) = 1$ for each $\theta \in \{\theta^L, \theta^H\}$. Note that the two origins correspond to the exam with one exam result, which is excluded under the current scenario.

[13] The teacher's expected payoff is $p_T[\pi(pass|\theta^H) - \pi(fail|\theta^H)] + (1 - p_T)[\pi(pass|\theta^L) - \pi(fail|\theta^L)]$.

With the presence of two students with different initial beliefs as in the original scenario, the teacher administers the same exam for both students. Our analysis below shows that the exam which maximizes the teacher's expected payoffs depends on the students' initial beliefs as well as that of the teacher.

2.3 | Two cases

The students' choices of actions depend on their updated beliefs (i.e., their initial beliefs, the realized grades, and π). Remember that given their initial beliefs, student 1 chooses a^L and student 2 chooses a^H with no exam (see Observation). Given this observation, we use the two terms "ignore" and "follow" hereafter in the following manner.

Definition We say

- student 1 **ignores** the result if she chooses a^L independently of the result,
- student 2 **ignores** the result if she chooses a^H independently of the result, and
- student $i \in \{1, 2\}$ **follows** the result if she chooses (i) a^H after observing *pass* and (ii) a^L after observing *fail*.[14]

In other words, we say (i) that a student ignores the result if her choice of action is the same as the one without exam independently of the result and (ii) that a student follows the result if her choice of action depends on the exam result.

Given the students' initial beliefs, for any π the teacher chooses, it is not possible for the teacher to make (i) student 1 always choose a^H or (ii) student 2 always choose a^L. In other words, for each student, it is either (i) that she ignores the result or (ii) that she follows the result. Therefore, depending on π the teacher chooses, one of the following four possible cases will happen:

A. Each student follows the result.
B. Student 1 follows the results while student 2 ignores the result.
C. Student 1 ignores the results while student 2 follows the result.
D. Each student ignores the result.

Among them, our focus is on Cases A and B. In both Cases C and D, student 1 ignores the result, as in the case without exam (Observation). From the teacher's point of view, this implies (i) that Case C is worse than the case without exam since student 2 may choose a^L by following the result and (ii) that Case D is equivalent to the case without an exam.[15]

In the following analysis, we consider Cases A and B. For each case, the teacher designs a different exam, and we thus consider them separately. For each case, we first identify the sets of the constraints on π's with which the case in consideration will happen. We then show the exam π such that the teacher maximizes her expected payoff given the constraints on π's. We call these payoff-maximizing exams in Cases A and B *exam* A and *exam* B, denoted π^A and π^B,

[14] If a student chooses different actions depending on the result, our focus is on the scenario where she follows the result. Not considering the case where a student chooses a^L after observing *pass* and a^H after observing *fail* is without loss of generality.
[15] Note that if $\pi(pass|\theta^H) = \pi(pass|\theta^L) = 1$, Cases C and D are equivalent.

respectively. We then discuss the teacher's optimal strategy by comparing her expected payoffs π^A and π^B.

3 | ANALYSIS

We first identify exams A and B. We then compare exams A and B. In particular, we demonstrate how the teacher's decision on which exam to employ depends on the parameters (beliefs).

3.1 | Exam A

We first identify the optimal exam for Case A—both students follow the results. Student 1 chooses a^H after observing *pass* only if

$$p_1(\theta^H|pass) = \frac{\pi(pass|\theta^H)p_1^0}{\pi(pass|\theta^H)p_1^0 + \pi(pass|\theta^L)\left(1 - p_1^0\right)} \geq \frac{1}{2} \Leftrightarrow p_1^0 \geq \frac{\pi(pass|\theta^L)}{\pi(pass|\theta^H) + \pi(pass|\theta^L)}. \tag{3}$$

Since $p_1^0 \in \left(0, \frac{1}{2}\right)$, (3) implies $\pi(pass|\theta^H) > \pi(pass|\theta^L)$. The expression (3) also implies that student 2 chooses a^H after observing *pass* since $p_2^0 > p_1^0$.[16] The teacher maximizes her expected payoffs subject to (3):[17]

$$\begin{aligned}&\left\{[\pi(pass|\theta^H)]^2 - [\pi(fail|\theta^H)]^2\right\}p_T^1 p_T^2\\ &+ \{\pi(pass|\theta^H)\pi(pass|\theta^L) - \pi(fail|\theta^H)\pi(fail|\theta^L)\}\left[p_T^1\left(1 - p_T^2\right) + \left(1 - p_T^1\right)p_T^2\right]\\ &+ \left\{[\pi(pass|\theta^L)]^2 - [\pi(fail|\theta^L)]^2\right\}\left(1 - p_T^1\right)\left(1 - p_T^2\right).\end{aligned}$$

Lemma 1. *The optimal exam under Case A, π^A, is*

$$(\pi^A(pass|\theta^H), \pi^A(pass|\theta^L)) = \left(1, \frac{p_1^0}{1 - p_1^0}\right).$$

Proof. The teacher's expected payoff is increasing in $\pi(pass|\theta^H)$ and decreasing in $\pi(fail|\theta^H)$ while an increase in $\pi(pass|\theta^H)$—and thus a decrease in $\pi(fail|\theta^H)$—makes

[16] We do not explicitly consider the constraint for student 2 to choose a^L after observing *fail*. We demonstrate that student 2 nevertheless follows the result under π^A.

[17] The first, second, and third terms correspond to the states where both are θ^H, only one of them is θ^H, and both are θ^L, respectively. In each bracket, the first term corresponds to the case where both receive *pass* (and the teacher's payoff is 1) while the second term corresponds to the case where both receive *fail* (and the teacher's payoff is −1).

it easier for (3) to hold. Hence, $\pi(pass|\theta^H) = 1$ and $\pi(fail|\theta^H) = 0$. The teacher's expected payoff is also increasing in $\pi(pass|\theta^L)$ and decreasing in $\pi(fail|\theta^L)$, implying that (3) binds and hence $\pi(pass|\theta^L) = \frac{p_1^0}{1-p_1^0} \in (0,1)$ and $\pi(fail|\theta^L) = \frac{1-2p_1^0}{1-p_1^0} \in (0,1)$ given $p_1^0 \in \left(0, \frac{1}{2}\right)$. □

The solution is qualitatively equivalent to the one identified in Section 2.2 for the student with $p_S^0 \in \left(0, \frac{1}{2}\right)$. As we discuss below, student 2 would also follow the result under π^A.

The students' updated beliefs after observing *pass* and *fail* are

$$p_1^A(\theta^H|pass) = \frac{1\cdot p_1^0}{1\cdot p_1^0 + \left(\frac{p_1^0}{1-p_1^0}\right)\left(1-p_1^0\right)} = \frac{1}{2},$$

$$p_2^A(\theta^H|pass) = \frac{1\cdot p_2^0}{1\cdot p_2^0 + \left(\frac{p_1^0}{1-p_1^0}\right)\left(1-p_2^0\right)} = \frac{\left(1-p_1^0\right)p_2^0}{2\left(1-p_1^0\right)p_2^0 - \left(p_2^0 - p_1^0\right)} > \frac{1}{2},$$

$$p_i^A(\theta^L|fail) = \frac{\left(1-\frac{p_1^0}{1-p_1^0}\right)\left(1-p_i^0\right)}{0\cdot p_i^0 + \left(1-\frac{p_1^0}{1-p_1^0}\right)\left(1-p_i^0\right)} = 1 \quad \text{for } i = 1, 2.$$

After observing *fail*, each student $i \in \{1,2\}$ chooses a^L since she is certain that her type is θ^L. The expressions above confirm that each student follows the result, as expected.[18]

The teacher's maximized expected payoff in Case A is

$$\begin{aligned} &p_T^1 p_T^2 + \left(\frac{p_1^0}{1-p_1^0}\right)\left[p_T^1\left(1-p_T^2\right) + \left(1-p_T^1\right)p_T^2\right] + \left\{\left(\frac{p_1^0}{1-p_1^0}\right)^2 - \left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\right\} \\ &\left(1-p_T^1\right)\left(1-p_T^2\right) \\ &= \left(\frac{1}{1-p_1^0}\right)^2\left[\left(1-p_1^0\right)p_T^1 + p_1^0\left(1-p_T^1\right)\right]\left[\left(1-p_1^0\right)p_T^2 + p_1^0\left(1-p_T^2\right)\right] \\ &\quad - \left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\left(1-p_T^1\right)\left(1-p_T^2\right). \end{aligned} \tag{4}$$

See Appendix A for the derivation of (4).

[18]Note that after observing *pass*, student 1 is indifferent. We assume that if a student is indifferent between the two actions, she will choose a^H. For such cases, Kamenica and Gentzkow (2011) use the notion of *sender-preferred* equilibrium to sidestep this issue.

3.2 | Exam B

We now identify the optimal exam under Case B—while student 1 follows the result, student 2 ignores the result. We still need (3) so that both students choose a^H after observing *pass*. After observing *fail*, student 2 chooses a^H only if

$$p_2(\theta^H|fail) = \frac{\pi(fail|\theta^H)p_2^0}{\pi(fail|\theta^H)p_2^0 + \pi(fail|\theta^L)\left(1 - p_2^0\right)} \geq \frac{1}{2} \Leftrightarrow p_2^0 \geq \frac{\pi(fail|\theta^L)}{\pi(fail|\theta^H) + \pi(fail|\theta^L)}. \tag{5}$$

Since student 2 ignores the result in Case B, the teacher's expected payoff is at least zero. Our focus is thus on student 1. The teacher maximizes

$$\pi(pass|\theta^H)p_T^1 + \pi(pass|\theta^L)\left(1 - p_T^1\right)$$

subject to (3) and (5).

Note that (3) and (5) can be rewritten as

$$\pi(pass|\theta^L) \leq \left[\frac{p_1^0}{1 - p_1^0}\right]\pi(pass|\theta^H) \quad \text{and} \quad \pi(fail|\theta^L) \leq \left[\frac{p_2^0}{1 - p_2^0}\right]\pi(fail|\theta^H),$$

where $\frac{p_1^0}{1-p_1^0} < 1 < \frac{p_2^0}{1-p_2^0}$, which provides the corresponding box diagram—the square with each side equal to one—in Figure 3.[19] The bottom-left origin corresponds to the first inequality (3) while the top-right origin corresponds to the second inequality (5), implying that any point in the lower triangle (except the bottom-left origin) satisfies two inequalities. Since the teacher wants higher $\pi(pass|H)$ and $\pi(pass|L)$ and lower $\pi(fail|H)$ and $\pi(fail|L)$, that is, towards northeast in the box diagram, both (3) and (5) bind. The solution thus corresponds to the point where the two lines intersect.

Lemma 2. *The optimal exam under Case B is*

$$(\pi^B(pass|\theta^H), \pi^B(pass|\theta^L)) = \left(\left(\frac{2p_2^0 - 1}{p_2^0 - p_1^0}\right)\left(1 - p_1^0\right), \left(\frac{2p_2^0 - 1}{p_2^0 - p_1^0}\right)p_1^0\right).$$

Refer to Appendix B for the proof. Note $0 < \frac{2p_2^0 - 1}{p_2^0 - p_1^0}p_1^0 < \frac{2p_2^0 - 1}{p_2^0 - p_1^0}\left(1 - p_1^0\right) < 1$, implying that θ^H has a higher chance to have *pass* than θ^L does.[20]

[19] The two origins are excluded from consideration since we assume that each result will be realized with a positive probability while each origin corresponds to the case where only one result will be realized.

[20] Note $\frac{2p_2^0-1}{p_2^0-p_1^0} \in \left(0, \frac{1}{1-p_1^0}\right)$ since $\lim_{p_2^0 \to \frac{1}{2}}\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right) = 0$ and $\lim_{p_2^0 \to 1}\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right) = \frac{1}{1-p_1^0}$.

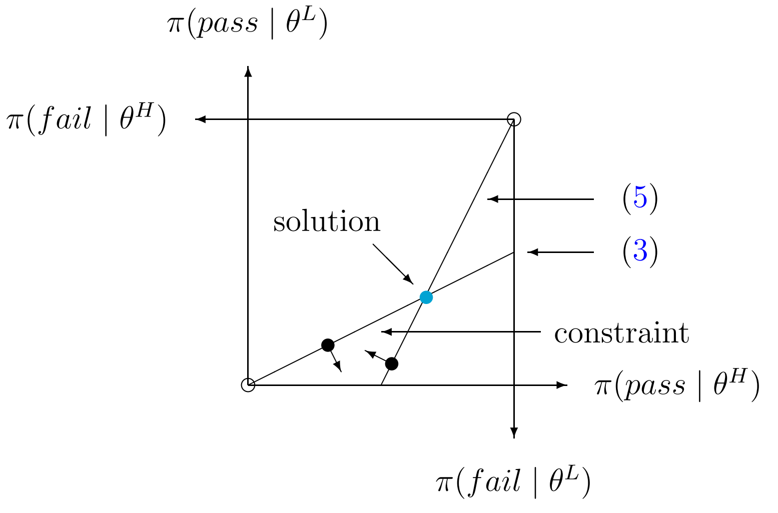

FIGURE 3 Box diagram—Case B [Color figure can be viewed at wileyonlinelibrary.com]

Students' updated beliefs after observing *pass* and *fail* are

$$p_1^B(\theta^H|pass) = \frac{\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)p_1^0}{\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)p_1^0+\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)p_1^0\left(1-p_1^0\right)} = \frac{1}{2},$$

$$p_1^B(\theta^L|fail) = \frac{\left[1-\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)p_1^0\right]\left(1-p_1^0\right)}{\left[1-\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)\right]p_1^0+\left[1-\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)p_1^0\right]\left(1-p_1^0\right)} > \frac{1}{2},$$

$$p_2^B(\theta^H|pass) = \frac{\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)p_2^0}{\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)p_2^0+\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)p_1^0\left(1-p_2^0\right)} > \frac{1}{2},$$

$$p_2^B(\theta^H|fail) = \frac{\left[1-\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)\right]p_2^0}{\left[1-\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)\right]p_2^0+\left[1-\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)p_1^0\right]\left(1-p_2^0\right)} = \frac{1}{2}.$$

These expressions show that while student 1 follows the result, student 2 ignores the result.

The teacher's corresponding expected payoff is

$$\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)p_T^1+\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)p_1^0\left(1-p_T^1\right)=\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]. \tag{6}$$

3.3 | Comparison of exams

In this subsection, we compare exams A and B from different viewpoints. One natural question is with which exam a student has a higher chance to receive *pass*. The direct comparison of exams A and B for each type gives

$$\pi^B(pass|\theta^L)=\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)p_1^0<\pi^A(pass|\theta^L)=\frac{p_1^0}{1-p_1^0},$$

$$\pi^B(pass|\theta^H)=\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)<\pi^A(pass|\theta^H)=1$$

and leads to the following result.[21]

Corollary *Irrespective of her type, a student has a higher chance to receive pass with exam A than with exam B.*

Students' preference: The best scenario from students' point of view in the current model is to know their types, leading to a payoff of one irrespective of the type realization since they would then make the appropriate choices. While student 1's optimal action is to follow the result in both exams A and B, would student 2 still want to be pushed to be more ambitious by choosing a^H even after receiving *fail*?

With "no exam," student 1's expected payoff is $1-p_1^0$ (by choosing a^L) while it is p_2^0 for student 2 (by choosing a^H). The expected payoffs for student 1 (who will follow the result in both scenarios) from exams A and B are, respectively,

$$\text{exam } A \qquad p_1^0+\left[1-\left(\frac{p_1^0}{1-p_1^0}\right)\right]\left(1-p_1^0\right)=1-p_1^0,$$

$$\text{exam } B \qquad \left[\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left(1-p_1^0\right)\right]p_1^0+\left[1-\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)p_1^0\right]\left(1-p_1^0\right)=1-p_1^0.$$

The expected payoffs for student 2 (who would follow the result in exam A and would ignore the result in exam B) from exams A and B are, respectively,

$$\text{exam } A \qquad p_2^0+\left[1-\left(\frac{p_1^0}{1-p_1^0}\right)\right]\left(1-p_2^0\right),$$

$$\text{exam } B \qquad p_2^0.$$

[21]For the first inequality, see footnote 19.

While student 1 is indifferent between exams A and B, student 2 strictly prefers exam A to exam B since $p_1^0 < \frac{1}{2}$.

Proposition 1. *Student 1's expected payoffs from exams A, B and "no exam" are identical. While student 2 is indifferent between exam B and "no exam," she strictly prefers exam A to exam B.*

That student 2 prefers exam A to exam B is *not* because the chance of having *pass* is higher in exam A than in exam B, implied by Corollary. Indeed, student 2 prefers to receive *fail* if her type is θ^L. This is because student 2 would know that she has θ^L if she receives *fail* under exam A so that she chooses a^L—the best action for θ^L—and this happens with a positive probability.

Teacher's preference: It is ultimately the teacher who decides which exam to administer. Remember that while the teacher's expected payoff is zero if students make their decisions without any information while the teacher's expected payoff from exam B is at least zero. This implies that an exam will take place. The teacher compares (4) and (6) to decide which exam to use.

Proposition 2. *The teacher chooses*

$$\text{exam} \begin{Bmatrix} \text{A} \\ \text{B} \end{Bmatrix} \text{ only if } \frac{1 - p_2^0}{1 - p_T^2} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} \frac{p_2^0 - p_1^0}{\left(1 - p_1^0\right)p_T^1 + p_1^0\left(1 - p_T^1\right)}. \tag{7}$$

Refer to Appendix C for the proof.

Several observations regarding the expression in (7):

- p_1^0 *and* p_2^0: The difference $p_2^0 - p_1^0 \in (0, 1)$ on the right-hand expression of (7) reflects the gap between the two students' initial beliefs. If p_2^0 is sufficiently close to 1, the teacher employs exam B.

- p_1^0 *and* p_T^1: The expression $\left(1 - p_1^0\right)p_T^1 + p_1^0\left(1 - p_T^1\right) \in (0, 1)$ relates p_1^0 and p_T^1. While the expression is strictly increasing in p_T^1 since $p_1^0 \in (0, \frac{1}{2})$, it is (i) increasing in p_1^0 if $p_T^1 < \frac{1}{2}$ or (ii) decreasing if $p_T^1 > \frac{1}{2}$. Therefore, a higher p_T^1 would favor exam A.
- p_2^0 *and* p_T^2: The ratio $\frac{1 - p_2^0}{1 - p_T^2} \in (0, \infty)$ reflects how confident student 2 is about her being Θ^H relative to the teacher's belief. The ratio is (i) equal to one if $p_2^0 = p_T^2$, (ii) less than one if $p_2^0 > p_T^2$ (i.e., student 2 is more confident than teacher that she is θ^H), and (iii) greater than one if $p_2^0 < p_T^2$ (i.e., teacher is more confident than student 2 that student 2 is θ^H).
- p_T^1 *and* p_T^2: The difference between p_T^1 and p_T^2 is crucial. The inequality above can be rewritten as

$$\frac{1 - p_2^0}{p_2^0 - p_1^0} \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} \frac{1 - p_T^2}{\left(1 - p_1^0\right)p_T^1 + p_1^0\left(1 - p_T^1\right)}. \tag{8}$$

Increasing either p_T^1 or p_T^2 (or both) lowers the expression on the right-hand side of (8), favoring exam A. Likewise, decreasing either p_T^1 or p_T^2 (or both) increases the expression on the right-hand side of (8), favoring exam B.

The observations above cannot be obtained if the teacher and the students share the same initial beliefs. Restricting our analysis to the case of the common initial beliefs therefore limits our analysis. These observations highlight the importance of allowing multiple students and heterogeneous beliefs.

For example, as the fourth observation suggests, if the teacher is sufficiently confident that both students are θ^H, the teacher chooses exam A. Another important observation is that which inequality holds also depends on the relative sizes of p_2^0 and p_T^2, as implied by the third observation mentioned above. Consider (8) again. Note that $\frac{1-p_2^0}{p_2^0-p_1^0} \in \left(0, \frac{1}{1-2p_1^0}\right)$ is strictly decreasing in p_2^0. Exam A is optimal if p_T^2 is sufficiently close to one while exam B is optimal if p_2^0 is sufficiently close to one. In other words, the teacher chooses exam A if she is sufficiently confident that student 2 is θ^H while she chooses exam B if student 2 is confident that she is θ^H.

The intuition for this observation is as follows: A higher chance that a student receives *pass* with exam A means that receiving *fail* is sufficiently convincing that she is θ^L. Thus, even student 2 chooses a^L after observing *fail*. Likewise, a lower chance that a student receives *pass* with exam B means that receiving *fail* is not convincing enough that she is θ^L. Student 2 then chooses a^H even after observing *fail*. From the teacher's point of view, exam A is desirable if she is confident that student 2 is θ^H—the teacher believes that the chance that student 2 receives *pass* (and thus chooses a^H) is high. However, exam B is desirable if she believes that student 2 is θ^L since student 2 chooses a^H independently of the realized grade with exam B.

4 | DISCUSSION

In the current setting with two students and heterogeneous beliefs, we show that the teacher would set one of the two exams which have different grade distributions. We observed (i) that given a set of the teacher's and the students' initial beliefs, a student has a higher chance to have *pass* with exam A independently of her type (Corollary) and (ii) while student 1 is indifferent between the two exams, student 2 strictly prefers exam A to exam B (Proposition 1). The latter observation is simply because exam A is more informative. The teacher's decision depends on the teacher's and the students' initial beliefs regarding the students' types (Proposition 2).

In reality, it is not necessarily the case that teachers can freely choose grade distributions. For example, some institutions may have grading policies in response to growing concerns regarding grade inflation. Given negative views on glade inflation, one may think that setting exams with higher means, such as exam A in the current context, should be discouraged. Recent studies however have documented positive effects of glade inflation on students' decisions.

Using a unique data set at Wesley College, Butcher et al. (2014) documented how a cap of B + on the average grades in the introductory and intermediate level courses affected students' choices of courses and majors. This cap had no effect on departments with lower average grades such as most of science departments as well as Economics while it affected the average grades in other departments (including humanities and other social sciences). In other words, the average grades of the former became relatively *higher* in the introductory and intermediate courses. Butcher et al. (2014) observed that more students chose Economics as their major (and sciences in a lesser extent) while the number declined for other social sciences.[22]

Lower university graduation rates have been a major issue in the United States. Using the data from nine large public universities, Denning et al. (2022) documented that graduation rates have increased since the 1990s, and showed that this observation can be explained by grade inflation. They showed that GPA is an indicator of graduation and suggested two possible reasons for this effect. First, grade inflation simply means that more students can exceed the threshold for the minimum GPA. Second, GPA could be viewed as a signal about a student's ability.

These empirical observations indicate that grade inflation may influence the decisions of certain groups of students in positive manners: some students who would have not chosen Economics and sciences did so in the former, and some students who would have dropped out actually completed universities in the latter. It could be interpreted that students are persuaded to make certain decisions via relatively higher grades (as in exam A), which they would not have done otherwise. These studies thus suggest that the task of exam setting can be viewed as an information design problem, as demonstrated in the current study.

There are potential extensions. With the current approach, one is that we could have more than two types, two actions, and two students. The teacher's preferences could be altered to reflect various contexts, possibly leading to different observations.[23] Other extensions would require more involved approaches. For example, we could introduce another strategic aspect due to the presence of multiple teachers.[24] The extension in this direction reflects an additional real-life aspect since students typically take multiple courses simultaneously. In addition, as implied by empirical studies above, there may be targeted groups of students whose decisions can be positively influenced by the design of grade distributions. Further theoretical contributions in this direction would benefit not only empirical studies but also have potential policy implications.

ACKNOWLEDGMENTS

I appreciate the associate editor and the reviewer for their valuable comments and suggestions. I would like to thank Pierpaolo Battigalli and Joel Sobel for their comments and suggestions on Shimoji (2022), and Alex Teytelboym for inspiring this study. I would also like to thank Anindya Bhattacharya, Subir Chattopadhyay, Alan Krause, Michael Shallcross, and Michael Thornton for discussion. Financial support from Research and Impact Support Fund of the Department of Economics and Related Studies at the University of York (RIS-12) is gratefully acknowledged.

[22]Using the data from the University of Kentucky, Ahn et al. (2022) studied the students' decisions regarding STEM (Science, Technology, Engineering, and Mathematics) enrollment. The data shows (i) lower grades in STEM compared to non-STEM, (ii) higher grades for female students in both STEM and non-STEM, and (iii) a smaller fraction of female students in STEM. Their counterfactual analysis showed that equating the average grade between STEM and non-STEM not only increases overall STEM enrollment but also shrinks the gender gap in STEM.

[23]For example, one can consider the following scenario: There is only one student who may have either p_1^0 or p_2^0, to which the teacher assigns equal probability. We would still need to have the same discussion regarding the student's response to the exam results (analogous to Cases A and B above), and the teacher's expected payoffs for these cases are modified. The comparison of the two cases produces the inequalities similar to the one in (7). I appreciate the associate editor for pointing this out.

[24]Gentzkow and Kamenica (2017) analyze Bayesian persuasion with multiple receivers.

CONFLICT OF INTEREST

The author declares no conflict of interest.

ORCID

Makoto Shimoji http://orcid.org/0000-0002-7895-9620

REFERENCES

Ahn, T., Arcidiacono, P., Hopson, A. & Thomas, J. (2022) *Equilibrium grade inflation with implications for female interest in STEM majors*. Working Paper. Available from: http://public.econ.duke.edu/%7Epsarcidi/

Alonso, R. & Câmara, O. (2016a) Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory*, 165, 672–706.

Alonso, R. & Câmara, O. (2016b) Persuading voters. *American Economic Review*, 106(1), 3590–3605.

Butcher, K. F., McEwan, P. J. & Weerapana, A. (2014) The effects of an anti-grade-inflation policy at Wellesley College. *Journal of Economic Perspectives*, 28(3), 189–204.

Card, D. & Della Vigna, S. (2020) What do editors maximize? Evidence from four leading economic journals. *Review of Economics and Statistics*, 102(1), 195–217.

Chan, J., Gupta, S., Li, F. & Wang, Y. (2019) Pivotal persuasion. *Journal of Economic Theory*, 180, 178–202.

Coughlan, S. (2018) *Does it matter what degree grade you get?* BBC. Available at https://www.bbc.co.uk/news/education-45939993 [Accessed 18 December 2022].

Denning, J. T., Eide, E. R., Mumford, K. J., Patterson, R. W. & Warnick, M. (2022) Why have college completion rates increased? *American Economic Journal: Applied Economics*, 14(3), 1–29.

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

Gentzkow, M. & Kamenica, E. (2017) Competition in persuasion. *Review of Economic Studies*, 84(1), 300–322.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S. & Xu, Y. (2019) Measuring success in education: the role of effort on the test itself. *American Economic Review: Insights*, 1(3), 291–308.

Hernández, P. & Neeman, Z. (2022) How Bayesian persuasion can help reduce illegal parking and other socially undesirable behavior. *American Economic Journal: Microeconomics*, 14(1), 186–215.

Kamenica, E. & Gentzkow, M. (2011) Bayesian persuasion. *American Economic Review*, 101(6), 2590–2615.

Oettinger, G. S. (2002) The effect of nonlinear incentives on performance: evidence from "Econ 101". *Review of Economics and Statistics*, 84(3), 509–517.

Schlosser, A., Neeman, Z. & Attali, Y. (2019) Differential performance in high versus low stakes tests: evidence from the GRE test. *Economic Journal*, 129(623), 2916–2948.

Shimoji, M. (2022) Bayesian persuasion in unlinked games. *International Journal of Game Theory*, 51(3–4), 451–481.

Stinebrickner, T. & Stinebrickner, R. (2012) Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30(4), 707–748.

Stout, R. (2008) *Fer-de-Lance & the league of frightened men*. New York: Bantam Dell.

Zafar, B. (2011) How do college students form expectations? *Journal of Labor Economics*, 29(2), 301–348.

How to cite this article: Shimoji, M. (2022) Setting an exam as an information design problem. *International Journal of Economic Theory*, 1–21. https://doi.org/10.1111/ijet.12368

APPENDIX A: TEACHER'S PAYOFF UNDER CASE A

Given π^A identified in Lemma 1, we derive the teacher's maximized expected payoff under Case A.

$$p_T^1 p_T^2 + \left(\frac{p_1^0}{1-p_1^0}\right)\left[p_T^1\left(1-p_T^2\right) + \left(1-p_T^1\right)p_T^2\right] + \left\{\left(\frac{p_1^0}{1-p_1^0}\right)^2 - \left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\right\}$$
$$\left(1-p_T^1\right)\left(1-p_T^2\right)$$
$$= \left(\frac{1}{1-p_1^0}\right)^2 \left\{\left(1-p_1^0\right)^2 p_T^1 p_T^2 + p_1^0\left(1-p_1^0\right)\left[p_T^1\left(1-p_T^2\right) + \left(1-p_T^1\right)p_T^2\right]\right.$$
$$\left. + \left[p_1^0\right]^2\left(1-p_T^1\right)\left(1-p_T^2\right)\right\}$$
$$- \left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\left(1-p_T^1\right)\left(1-p_T^2\right)$$
$$= \left(\frac{1}{1-p_1^0}\right)^2 \left\{\left(1-p_1^0\right)p_T^1\left[\left(1-p_1^0\right)p_T^2 + p_1^0\left(1-p_T^2\right)\right] + p_1^0\left(1-p_T^1\right)\right.$$
$$\left.\left[\left(1-p_1^0\right)p_T^2 + p_1^0\left(1-p_T^2\right)\right]\right\}$$
$$- \left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\left(1-p_T^1\right)\left(1-p_T^2\right).$$

The expression (4) can be directly obtained from the last line above.

APPENDIX B: PROOF OF LEMMA 2

In this subsetion, we solve for the optimal exam π^B under Case B. As shown in Figure 3, both inequalities (3) and (5) bind. We thus solve for $\pi(pass|\theta^H)$ and $\pi(pass|\theta^L)$ given (3) and (5) with equalities.

$$\frac{\pi(pass|\theta^H)p_1^0}{\pi(pass|\theta^H)p_1^0 + \pi(pass|\theta^L)\left(1-p_1^0\right)} = \frac{1}{2} \Leftrightarrow \pi(pass|\theta^H)p_1^0 - \pi(pass|\theta^L)\left(1-p_1^0\right)$$
$$= 0$$

and

$$\frac{\pi(fail|\theta^H)p_2^0}{\pi(fail|\theta^H)p_2^0 + \pi(fail|\theta^L)\left(1-p_2^0\right)} = \frac{1}{2} \Leftrightarrow \pi(fail|\theta^H)p_2^0 - \pi(fail|\theta^L)\left(1-p_2^0\right) = 0$$
$$\Leftrightarrow [1-\pi(pass|\theta^H)]p_2^0 - [1-\pi(pass|\theta^L)]$$
$$\left(1-p_2^0\right) = 0$$
$$\Leftrightarrow \pi(pass|\theta^H)p_2^0 - \pi(pass|\theta^L)\left(1-p_2^0\right)$$
$$= 2p_2^0 - 1.$$

We then have

$$\begin{bmatrix} p_1^0 & -\left(1-p_1^0\right) \\ p_2^0 & -\left(1-p_2^0\right) \end{bmatrix} \begin{bmatrix} \pi(pass|\theta^H) \\ \pi(pass|\theta^L) \end{bmatrix} = \begin{bmatrix} 0 \\ 2p_2^0 - 1 \end{bmatrix}$$

and thus

$$\begin{bmatrix} \pi(pass|\theta^H) \\ \pi(pass|\theta^L) \end{bmatrix} = \frac{1}{-p_1^0\left(1-p_2^0\right)+p_2^0\left(1-p_1^0\right)} \begin{bmatrix} -\left(1-p_2^0\right) & 1-p_1^0 \\ -p_2^0 & p_1^0 \end{bmatrix} \begin{bmatrix} 0 \\ 2p_2^0 - 1 \end{bmatrix} = \frac{2p_2^0-1}{p_2^0-p_1^0} \begin{bmatrix} 1-p_1^0 \\ p_1^0 \end{bmatrix}.$$

APPENDIX C: PROOF OF PROPOSITION 2

In this subsection, we compare the teacher's expected payoffs with π^A and π^B, namely (4) and (6) to derive the condition under which one exam leads to a higher expected payoff than the other.

$$\begin{aligned}
(4)-(6) = & \left(\frac{1}{1-p_1^0}\right)^2 \left[\left(1-p_1^0\right)p_T^1 + p_1^0\left(1-p_T^1\right)\right]\left[\left(1-p_1^0\right)p_T^2 + p_1^0\left(1-p_T^2\right)\right] - \left(\frac{1-2p_1^0}{1-p_1^0}\right)^2 \\
& \left(1-p_T^1\right)\left(1-p_T^2\right) \\
& -\left(\frac{2p_2^0-1}{p_2^0-p_1^0}\right)\left[\left(1-p_1^0\right)p_T^1 + p_1^0\left(1-p_T^1\right)\right] \\
= & \frac{\left(1-p_1^0\right)p_T^1 + p_1^0\left(1-p_T^1\right)}{\left(p_2^0-p_1^0\right)\left(1-p_1^0\right)^2}\left\{\left(p_2^0-p_1^0\right)\left[\left(1-p_1^0\right)p_T^2 + p_1^0\left(1-p_T^2\right)\right]\right. \\
& \left. -\left(1-p_1^0\right)^2\left(2p_2^0-1\right)\right\} \\
& -\left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\left(1-p_T^1\right)\left(1-p_T^2\right) \\
= & \frac{\left(1-p_1^0\right)p_T^1 + p_1^0\left(1-p_T^1\right)}{\left(p_2^0-p_1^0\right)\left(1-p_1^0\right)^2}\left\{\left(p_2^0-p_1^0\right)\left[p_1^0 + \left(1-2p_1^0\right)p_T^2\right] - \left(1-p_1^0\right)^2\left(2p_2^0-1\right)\right\} \\
& -\left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\left(1-p_T^1\right)\left(1-p_T^2\right) \\
= & \frac{\left(1-p_1^0\right)p_T^1 + p_1^0\left(1-p_T^1\right)}{\left(p_2^0-p_1^0\right)\left(1-p_1^0\right)^2}\left\{\left(p_2^0-p_1^0\right)\left(1-2p_1^0\right)p_T^2 + \left(p_2^0-p_1^0\right)p_1^0\right. \\
& \left. -\left(1-p_1^0\right)^2\left(2p_2^0-1\right)\right\} \\
& -\left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\left(1-p_T^1\right)\left(1-p_T^2\right).
\end{aligned}$$

By noting

$$\begin{aligned}\left(p_2^0-p_1^0\right)p_1^0-\left(1-p_1^0\right)^2\left(2p_2^0-1\right)&=\left(p_2^0-p_1^0\right)p_1^0-\left[1-2p_1^0+\left(p_1^0\right)^2\right]\left(2p_2^0-1\right)\\&=p_1^0\left[\left(p_2^0-p_1^0\right)-p_1^0\left(2p_2^0-1\right)\right]-\left(1-2p_1^0\right)\\&\quad\left(2p_2^0-1\right)\\&=p_1^0p_2^0\left(1-2p_1^0\right)-\left(1-2p_1^0\right)\left(2p_2^0-1\right)\\&=\left(1-2p_1^0\right)\left[1-2p_2^0+p_1^0p_2^0\right]\end{aligned}$$

the expression above becomes

$$\begin{aligned}&\frac{\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)}{\left(p_2^0-p_1^0\right)\left(1-p_1^0\right)^2}\Big\{\left(p_2^0-p_1^0\right)\left(1-2p_1^0\right)p_T^2+\left(1-2p_1^0\right)\\&\left[1-2p_2^0+p_1^0p_2^0\right]\Big\}\\&-\left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\left(1-p_T^1\right)\left(1-p_T^2\right)\\&=\frac{\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-2p_1^0\right)}{\left(p_2^0-p_1^0\right)\left(1-p_1^0\right)^2}\left[\left(p_2^0-p_1^0\right)p_T^2+1-2p_2^0+p_1^0p_2^0\right]\\&\quad-\left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\left(1-p_T^1\right)\left(1-p_T^2\right).\end{aligned}$$

By noting

$$\begin{aligned}\left(p_2^0-p_1^0\right)p_T^2+1-2p_2^0+p_1^0p_2^0&=\left(p_2^0-p_1^0\right)p_T^2+1-2p_2^0+\left(p_2^0\right)^2-\left(p_2^0\right)^2+p_1^0p_2^0\\&=\left(p_2^0-p_1^0\right)p_T^2+\left(1-p_2^0\right)^2-\left(p_2^0-p_1^0\right)p_2^0\\&=\left(1-p_2^0\right)^2-\left(p_2^0-p_1^0\right)\left(p_2^0-p_T^2\right)\end{aligned}$$

the expression above becomes

$$
\begin{aligned}
&\frac{\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-2p_1^0\right)}{\left(p_2^0-p_1^0\right)\left(1-p_1^0\right)^2}\left[\left(1-p_2^0\right)^2-\left(p_2^0-p_1^0\right)\left(p_2^0-p_T^2\right)\right]-\left(\frac{1-2p_1^0}{1-p_1^0}\right)^2\\
&\quad\left(1-p_T^1\right)\left(1-p_T^2\right)\\
&=\underbrace{\frac{\left(1-2p_1^0\right)}{\left(p_2^0-p_1^0\right)\left(1-p_1^0\right)^2}}_{(+)}\left\{\begin{matrix}\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left[\left(1-p_2^0\right)^2-\left(p_2^0-p_1^0\right)\left(p_2^0-p_T^2\right)\right]\\ -\left(1-2p_1^0\right)\left(p_2^0-p_1^0\right)\left(1-p_T^1\right)\left(1-p_T^2\right)\end{matrix}\right\}\\
&\propto\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)^2-\left(p_2^0-p_1^0\right)\left\{\begin{matrix}\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(p_2^0-p_T^2\right)\\ +\left(1-2p_1^0\right)\left(1-p_T^1\right)\left(1-p_T^2\right)\end{matrix}\right\}\\
&=\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)^2\\
&\quad-\left(p_2^0-p_1^0\right)\left\{\begin{matrix}\left(1-p_1^0\right)p_T^1\left(p_2^0-p_T^2\right)+p_1^0\left(1-p_T^1\right)\left(p_2^0-p_T^2\right)\\ +\left(1-p_1^0\right)\left(1-p_T^1\right)\left(1-p_T^2\right)-p_1^0\left(1-p_T^1\right)\left(1-p_T^2\right)\end{matrix}\right\}\\
&=\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)^2\\
&\quad-\left(p_2^0-p_1^0\right)\Big\{\left(1-p_1^0\right)\left[p_T^1\left(p_2^0-p_T^2\right)+\left(1-p_T^1\right)\left(1-p_T^2\right)\right]+p_1^0\left(1-p_T^1\right)\\
&\quad\left[\left(p_2^0-p_T^2\right)-\left(1-p_T^2\right)\right]\Big\}\\
&=\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)^2\\
&\quad-\left(p_2^0-p_1^0\right)\left\{\left(1-p_1^0\right)\left[\left(1-p_T^2\right)-p_T^1\left(1-p_2^0\right)\right]-p_1^0\left(1-p_2^0\right)\left(1-p_T^1\right)\right\}\\
&=\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)^2\\
&\quad-\left(p_2^0-p_1^0\right)\left\{\left(1-p_1^0\right)\left(1-p_T^2\right)-\left(1-p_2^0\right)\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\right\}\\
&=\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)\left[\left(1-p_2^0\right)+\left(p_2^0-p_1^0\right)\right]-\left(p_2^0-p_1^0\right)\left(1-p_1^0\right)\left(1-p_T^2\right)\\
&=\underbrace{\left(1-p_1^0\right)}_{(+)}\left\{\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)-\left(p_2^0-p_1^0\right)\left(1-p_T^2\right)\right\}\\
&\propto\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)-\left(p_2^0-p_1^0\right)\left(1-p_T^2\right).
\end{aligned}
$$

We then have

$$
\begin{Bmatrix}A\\B\end{Bmatrix}\text{ is preferred only if }\left[\left(1-p_1^0\right)p_T^1+p_1^0\left(1-p_T^1\right)\right]\left(1-p_2^0\right)-\left(p_2^0-p_1^0\right)\left(1-p_T^2\right)
\begin{Bmatrix}\geq\\ \leq\end{Bmatrix}0
$$

from which (7) can be directly obtained.