國立臺灣大學電機資訊學院電機工程學系
碩士論文
Department of Electrical Engineering
College of Electrical Engineering and Computer Science
National Taiwan University
Master Thesis

促進大型開放式線上課程中的同儕互評
Encouraging Peer Grading in MOOCs

柯劭珩
Shao-Heng Ko

指導教授：陳和麟博士
Advisor: Ho-Lin Chen, Ph.D.

中華民國 106 年 5 月
May, 2017

doi:10.6342/NTU201701049

國立臺灣大學碩士學位論文
口試委員會審定書

促進大型開放式線上課程中的同儕互評
Encouraging Peer Grading in MOOCs

本論文係柯劭珩君 (R04921049) 在國立臺灣大學電機工程學系完成之碩士學位論文，於民國 106 年 5 月 31 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

所　長：

doi:10.6342/NTU201701049

誌謝

在論文完成的此刻，首先要感謝指導教授陳和麟老師。有人說，普通的老師是拉著學生往前走；而好的老師則是在一條做學問的路上先點起燈，讓學生自己走，在摸索的過程中自己成長茁壯。然而，我的指導教授不屬於以上任何一種。他會在任何時刻，為學生指引出前方每一條不同的路，其中許多都有他探訪過留下的足跡。而當你居然選擇了一條他未曾去過的路，他也會陪著你走一遭。在碩士班的兩年中，老師給了我非常大的彈性和信任，讓我能將興趣和注意力集中在許多不同的事物上，最後能從生活中找到一個新的研究題材；而在論文的發想、撰寫和修改時，總是精準的找到我的盲點。接下來要感謝口試委員們：葉丙成老師、于天立老師、孔令傑老師。三位老師在百忙中抽空審閱本篇論文、擔任口試委員，也都對本篇論文提出了非常實用的建議。然而老師們所不知道的是，，在接受邀請之前，其實他們早已對我的研究做出了深遠的影響。在大學二年級時，有緣成為葉老師前幾批翻轉教學的對象，讓我體驗到了不同的學習樂趣和成就感，進而啟發我踏入 MOOCs 的世界；而于老師的人工智慧，則是我人生中第一堂透過 MOOCs 平台修習的課程。在孔老師的資訊經濟課堂中，我學習了許多建模和分析的技巧，開始思考直接從生活中尋找新問題的可能性。本篇論文的研究主題，是我在台大 MOOCs 團隊實習過程中實際遭遇的問題，而老師們除了都和團隊有密切的關係，各自開授的實體課程，也都是我大學和碩士班生涯中，最有收穫的幾堂課。能夠邀請到老師們擔任我的口試委員，是非常圓滿的緣分。感謝台大 MOOCs 團隊的教學設計師、實習生同仁們，兩年來無論在實務或理論上的切磋討論、思辯交流；感謝我的女朋友曉亭，在這段期間不間斷的陪伴，在咖啡館中見證主要成果的誕生；感謝我的父母、家人、朋友們的支持與鼓勵，讓我能專注在學業之上。作為碩士論文，我相信這只是個開始，我盼望並且期待，能將此生虛擲在思考有趣的事物之上。

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

摘要

在大型開放式線上課程 (MOOCs) 當中，由於學習者數量極為龐大，高階學習表現通常只能透過同儕互評 (Peer Grading) 的方式來評量。在 MOOCs 中實施同儕互評時，學習者通常缺乏為其他人評分的動機，因而沒有付出足夠的心力。為改善此現象，我們考慮讓學習者的成績與其評量他人的準確度相關的機制，並建立相關的賽局理論模型，以分析學習者在此機制下的理性行為。我們發現一組能保證純粹均衡存在的條件，在此條件下，課程設計者將可透過調整機制參數的方式，促進學習者投資更多的心力在評分之上。更進一步，若學習者之間具有同質性時，我們證明在所有純粹均衡當中，所有為同一份作業評分的學習者都會付出恰相等的時間。藉由這個性質，我們能夠計算所有可能的純粹均衡點。我們將上述結果推廣到某些學習者並非採取理性策略的狀況，並討論如何在實際情況中應用本研究的結果。

關鍵字：賽局理論，大型開放式線上課程，同儕互評，納許均衡，機制設計

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

Abstract

Due to huge participant sizes in *Massive Open Online Courses* (MOOCs), *peer grading* is practically the only existing solution to grading high-level assignments. One of the main issues of utilizing peer grading in MOOCs is that learners are not motivated and do not spend enough effort in grading. To modify current peer grading mechanism to induce better grading, we focus on the idea of making the learners' grade depend on the accuracy of their grading of others' work. We built a game theoretical model to characterize the rational behavior of learners in such a mechanism. We found a set of conditions which guarantees existence of pure-strategy equilibria. When the conditions are satisfied, the course designer can encourage the learners to spend more time on grading through tuning the mechanism parameters. Furthermore, when the learners are assumed to be homogeneous, we proved that in any pure equilibrium, any submitted work will be graded with identical effort by the relevant graders. With this property all the possible pure equilibria are theoretically calculable. We also extended our result to the case where some of the learners are not strategic or rational. We discussed applications of our results in practical situations.

Keywords. Game Theory, Massive Open Online Courses, Peer Grading, Nash Equilibrium, Mechanism Design

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

Contents

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

Chapter 1

Introduction

Online learning is becoming a flourishing industry. In addition to the long-standing OpenCourseWares, learners today have MOOCs, or *Massive Open Online Courses*, to gain knowledge from. Unlike OpenCourseWares, which simply releases the course material, MOOCs are "online classes that anyone, anywhere can participate in" [13]. Furthermore, current MOOCs are trending toward formality; they are charging certification fees, offering institutional credits and even online programs. Consequently, the task of performance assessment is becoming more and more important.

It is common for a MOOC to have thousands of participants, in which the load of evaluating the grades is well beyond any course staff can afford. While automated grading techniques can easily handle multiple choices and programming tasks, they are nearly useless when it comes to grading more sophisticated assignments, like mathematical proofs, art work, written pieces and speeches. *Peer grading* is practically the only existing solution to grading high-level assignments.

The concept of peer grading comes from traditional pedagogy, where its effects were well-studied [1, 8, 9, 20, 21]. In the context of MOOCs, after any learner submits an assignment, it will be graded by several peers, and the learner himself is required to grade the others' assignments as well. The final score of an assignment is determined by some aggregation of all the scores given to it. Past work [4, 9, 15] has shown that such aggregation is decently close to the evaluation from the instructor even with a small number of graders grading each assignment. Peer grading also has the effect of deepening learner's

doi:10.6342/NTU201701049

comprehension [19], building positive learning environment [21], and even metacognitive benefits [8].

However, there are problems to be addressed when peer grading is applied in MOOCs. Learners in MOOCs, who come from all around the world and all backgrounds, feature great diversity. Empirical study shows that peer grading has weaker reliability in MOOCs than in real classes [17], and not all learners are satisfied with the mechanism [4, 17]. This leads to a main question:

How can we modify the peer grading mechanism to induce better grading precision?

The work of Piech et al. [18], with an eye on tuning the mechanism used by Coursera, investigated various probabilistic models of peer grading. A part of their work assigns weights to graders by past performance, exploiting the assumption that more adequate grading is correlated with higher grades. While the assumption is still up for debate [7], Piech et al. claim that peer grading accuracy can be improved simply by measuring the bias and reliability of graders. However, there remains an unaddressed issue: graders in practice are not spending enough time on peer grading, possibly due to lacking motivation. Thus, Piech et al. called for game theoretical research on mechanisms to incentivize learners to put more effort into grading.

Following the above work, there have been multiple work on incentivizing the learners, by both rewards and punishments. On the rewards side, de Alfaro and Shavlovsky [6] implemented a system named *Crowdgrader* that lets learners collaboratively review and grade assignments. In this system, the overall grade of a learner depends both on the aggregate grade received, and their "precision" in reviewing their peer's work, which is determined by the average error with respect to the consensus grade. While this ever-evolving platform is primarily used in college classes, the designers reported that "the number of students who complained about mis-gradings was about the same as the one we typically experience using TAs" [6].

On the punishments side, Carbonara et al. [3] tackled the problem using an audit game

doi:10.6342/NTU201701049

approach, in which learners are penalized if they are caught misgrading the others' work. They studied the problem of allocating limited auditing resources, like TA hours, to heterogeneous learners. Under the assumption that all learners spend a fixed amount of total time in doing the assignment and peer grading, they showed an algorithm to obtain an approximately optimal allocation. However, this assumption is also the biggest caveat of their model, as it is better to motivate learners to spend more time on learning.

Finally, Lu et al. [16] designed a large experiment to motivate peer graders by letting their grading performance to be examined as well, but without any rewards or punishments as consequence. While this alone seemed not to do the trick, they found that learners improved their grading accuracy by only evaluating the others' grading performances. They proposed the possibility of motivating the graders by other incentives instead of grades.

Motivated by all the above work, and the fact that most of the work about peer grading are empirical studies, we aim to investigate the behavior of peer graders under the rewards approach from a game theoretical perspective. By building a model of the peer grading mechanism, we ask the following main questions:

- *What will be the rational behaviors of graders?*
- *Under what conditions will such mechanisms work in line with our expectations?*
- *What can the course provider do to affect the graders' behaviors?*

The overall structure of our model is similar to the design in [6], in which a consensus grade of an assignment directly comes from aggregation of the graders' opinions. An evaluation of one grader's accuracy is then measured by how far away his opinions are from the related consensus grades. The final utility of a learner is realized as a linear combination of his assignment score and his grading accuracy. Instead of strategically deciding how much points they give to a particular assignment, we assume that each grader only decides the amount of time he puts into every grading task, which will not be observed

doi:10.6342/NTU201701049

by the course provider. Unlike that in [3], we do not assume a tradeoff between time put in doing assignment and in peer grading, as these two phases often have different time periods in common MOOCs settings.

Though we mainly think of the utility from precise grading as given grades, our model can capture other types of rewards like fame, self-fulfillment, or even monetary rewards and career opportunities. While [6] measures the grader's overall accuracy, we take into account every grading attempt rather separately, broadening the possible strategies for each grader. Generally, we assume the peer graders are heterogeneous, both in their grading ability and their evaluation of time, which captures the feature of a MOOCs environment.

Similar to that in [18], we assume that the grade given by a peer grader to each assignment is distributed around the true value of the assignment. Naturally, the degree of dispersion depends on the effort in grading. We do not assume a peer grader is capable of doing precise grading even with unlimited effort. We further assume that the graders do *unbiased grading*, which means they overgrade and undergrade equally often. While peer graders tend to bias toward high grades in reality [11, 15], this effect can be neutralized by pretreatments after measuring the bias.

The main result of this paper is a set of sufficient conditions, called the *encouraging conditions*, which guarantees the existence of pure Nash equilibria. Once the condition is met, the course designer can encourage or discourage the graders to spend more time on grading by tuning the mechanism parameters. We also proved that the encouraging conditions are satisfied in a series of exact situations, where the grade given by a peer grader to an assignment follows a normal distribution with the intrinsic value of the assignment being the mean and the variance being a non-increasing and convex function of the time spent.

Later, we focus on a special environment where all learners are homogeneous. This setting can be related to SPOCs, or *Small Private Online Courses*, where "MOOCs are used as a supplement to classroom teaching" [10]. We found that under this assumption, every pure Nash equilibrium has the property that each assignment is graded by the same

doi:10.6342/NTU201701049

level of effort by all the relevant graders. On the other hand, we pointed out that the pattern of equilibrium behaviors is still valid in the situation when some of the graders are assumed to be irrational players with perdetermined strategies.

The general model we use is introduced in Section 2, where we promptly simplify it into more compact models. Our main result lies in Section 3, including the encouraging conditions, the consequent existence of pure equilibria, and how the mechanism designer encourages peer grading by tuning the parameters. We also propose a series of practical settings that satisfy the encouraging conditions. In Section 4, we present the further limited homogeneous model and obtain stronger results. In Section 5 we extend our analysis to include irrational graders. Finally, discussions toward biased grading and practical implications are in Section 6.

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

Chapter 2

Model

In this chapter, we propose a model which will be used throughout the paper. We characterize those assignments that can be graded without subjectivity by assuming an intrinsic value for each submitted work. Intuitively, the more time a learner spent on grading a submission, the closer his grading will be to its intrinsic value. Our model is mainly built on such assumption, while the overall structure is similar to the common mechanism used in MOOC platforms.

2.1 Players and Actions

We assume N learners, $\{a_1, a_2, ...a_N\}$, working on some assignment in a MOOC, have already finished their submissions and are entering the peer-grading phase. The submitted work of a_i has intrinsic value $v_i \in [0, M]$, where M is the maximum score of this assignment. The grading task is fully described by **G**, an N by N matrix with boolean elements. $G_{(i,j)}$ equals to 1 if a_j is asked to peer grade the submitted work of a_i, and 0 otherwise. Each learner is asked to grade exactly k submissions, and each submission is graded by exactly k learners; hence, $\sum_i G_{(i,j)} = \sum_i G_{(j,i)} = k$. Furthermore, no learner self-grades his submitted work; hence $G_{(i,i)} = 0$. We assume that all possible grading relations are chosen equally likely beyond the learners' knowledge. Thus the learners cannot inference any information about who is grading their submissions or vice versa.

doi:10.6342/NTU201701049

The vector $T_j = [t_j^1, t_j^2, ...t_j^N]$ is the strategy of learner a_j. $t_j^i \in [0, U]$ is the amount of time he puts on grading the submission of a_i, where U is assumed to be the upper bound limit. This upper bound is natural most of the time; no player will spend longer time if the corresponding cost is larger than the maximum reward possible. While such limit does not exist in practice, peer grading cannot take forever long, and there will be a stop-loss point such that keep increasing the time will not do any good. In reality, a grader only determines how much of time he puts on each submission he receives. Therefore, $t_j^i = 0$ if $G_{(i,j)} = 0$. The course provider cannot observe any learner's strategies.

2.2 Grading Mechanism

In this model, we assume that a peer grader does not strategically give grades. Instead, the grades he gives are random variables, following distributions decided by the amount of time (effort) he puts on each submitted work, and its intrinsic value. To characterize this, we denote $f_j(\cdot)$ to be the *grading possibility function* of grader j, where $f_j(x, v, t)$ refers to the probability that grader j gives x points to an assignment with intrinsic value v, after spending t units of time on grading it. Thus, given fixed v, t, $f_j(x, v, t)$ is a probability density function, and corresponds to a cumulative density function $F_j(x, v, t)$. Clearly, $f_j(x, v, t) = 0 \, \forall j, \, \forall \, x \notin [0, M]$. Let S_j^i be the score a_j gives to the submitted work of a_i. If $G_{(i,j)} = 1$, then $S_j^i \sim F_j(\cdot, v_i, t_j)$. Furthermore, we assume unbiased grading: given v, t, $f_j(v-x, v, t) = f_j(v+x, v, t)$, $\forall x \in \mathbf{R}$. To simplify notations, we use f_{g} to represent the functions $f_1(\cdot,\cdot,\cdot), f_2(\cdot,\cdot,\cdot), ...f_N(\cdot,\cdot,\cdot)$.

Denote $\mathbf{S}_i = [S_j^i | G_{(i,j)} = 1]$ to be the vector of grades given to the submission of a_i. The aggregate score, or consensus grade, of a_i's submission, is then determined by $\hat{S}_i = f_{\text{agg}}(\mathbf{S}_i)$. Naturally, $f_{\text{agg}}(\mathbf{S}_i) = f_{\text{agg}}(\mathbf{S'}_i)$ if $\mathbf{S'}_i$ is a permutation of $\mathbf{S}_i$; that is, the consensus grade should not depend on the order of the peer grades given. Furthermore, $f_{\text{agg}}(\cdot)$ should be non-decreasing in every element in $\mathbf{S}_i$. There exist various methods in aggregating peer grades; for example, the Crowdgrader platform in [6] uses an Olympian average function, where the highest and the lowest grades are dropped before taking

doi:10.6342/NTU201701049

average. Coursera chooses a median function instead [5]. Thus we do not specify an exact method here. Since we require $f_j(\cdot, v_i, t)$ to be symmetric with respect to v_i for any t, given any strategy profile, the derived probability distribution of $\hat{S}_i$ will also be symmetric with respect to v_i.

Next, the accuracy level of a_j grading the submission of a_i is determined by $\alpha_j^i = f_{\text{accu}}(|S_j^i - \hat{S}_i|) \in [0, 1]$, while $f_{\text{accu}}(\cdot)$ must be non-increasing. Similarly to above, we only require that the function solely depends on the difference between the single grade and the aggregated grade. Finally, the average peer grading accuracy of a_j is determined by $\hat{\alpha}_j = \frac{1}{k}\sum_i(\alpha_j^i)$. Clearly, since $\alpha_j^i \in [0, 1]$, $\hat{\alpha}_j \in [0, 1]$.

2.3 Utility

We assume that grader a_i has time-to-grade ratio $r_i \geq 0$, which means he is willing to give up r_i units of time to earn one point in his grade. All values of r_i are public information. We then define the time-to-grade ratio vector $\mathbf{r} = [r_i]$. Also, we define λ to be the portion grading performance accounts for in one's final utility. This means a learner can earn up to λM points by peer grading, if we take into account only rewards in grades. The final utility of learner a_i is defined to be $\pi_i = \lambda M \hat{\alpha}_i - r_i \sum_j t_i^j$. All learners are risk-neutral. Note that a_i should get $(1 - \lambda)\hat{\mathbf{S}}_i$ points from his own work as well. However, this part is independent to any of his strategic decision. Hence we can remove this term from the utility in our model. We can define one specific game model by defining all the above parameters, functions and probablistic distributions.

Definition 2.1. An unbiased peer grading game, or UPG game, is a tuple
$G = \big(N, k, M, \mathbf{r}, U, \lambda, f_{\text{g}}, f_{\text{agg}}(\cdot), f_{\text{accu}}(\cdot)\big)$.

Note that all f_j's need not to be the same. Hence, the graders' actions, and consequently the perceived score distributions are generally heterogeneous. Nowadays, it is common for MOOCs to utilize analytic rubrics in peer grading assignments, leading to more objective and systematic grading results; however, complete objectivity is impossi-

doi:10.6342/NTU201701049

ble to be achieved. Also, how learners value their time is generally out of the instructor's control.

Parameters and functions $M, k, \lambda, f_{\text{agg}}(\cdot), f_{\text{accu}}(\cdot)$ are chosen by the mechanism designer. These information, $N, \mathbf{r}$ and f_{g} and the whole structure of the mechanism are public information to all learners, while the intrinsic value vector $\mathbf{V} = \{v_i\}$ and the relation matrix $\mathbf{G}$ are unobserved. Since the learners are risk-neutral, learner a_i rationally determines T_i in order to maximize $\mathbf{E}[\pi_i]$. Note that T_i will not be observed by the designer and will not be directed used to determine the reward; the designer can only decide the parameters and functions above, seeking to induce better overall effort and/or grading accuracy.

2.4 Decomposition of Model

First and foremost, a simplification can be directly made from our model. We can observe that an UPG game is indeed a series of independent smaller subgames, each containing only one submitted work of assignment. This is described in the theorem below.

Definition 2.2. Given the other players' strategy profile, $T_{-j} = \{T_1, ...T_{j-1}, T_{j+1}, ..., T_N\}$, we define learner a_j's *best response* T^*_j to be the optimal choice of T_j that maximizes $\mathbf{E}[\pi_j]$, conditioned on T_{-j}.

Theorem 2.1. *Given the other players' two strategy profiles, T_{-j}, T'_{-j}, the corresponding best responses t^*_j and t'^*_j satisfies the following: $t^{*i}_j = t'^{*i}_j$, if $t^i_{j'} = t'^i_{j'}, \forall j' \neq j$.*

Proof. Consider player j. We first observe that his total expected utility can be decomposed into

$$\mathbf{E}[\pi_j] = \sum_i (\frac{\lambda M}{k}\mathbf{E}[\alpha^i_j] - r_j t^i_j).$$

Therefore, maximizing the expected utility is equivalent to maximizing the total of the k nonzero terms, each corresponding to the expected utility from grading one specific submitted work. Given the other players' strategies, fixing i, the term $\frac{\lambda M}{k}\mathbf{E}[\alpha^i_j] - r_j t^i_j$

doi:10.6342/NTU201701049

only depends on $\mathbf{T^i} = [t_j^i | G_(i,j) = 1]$. Thus, maximizing the total of all k terms is equivalent to simultaneously maximizing k terms independently.

If $t_{j'}^i = t'^i_{j'}, \forall j' \neq j$ holds, then the optimal choice of t_j^i, which maximizes $\frac{\lambda M}{k}\mathbf{E}[\alpha_j^i] - r_j t_j^i$, should also maximize $\frac{\lambda M}{k}\mathbf{E}[\alpha'^i_j] - r_j t'^i_j$. □

Equivalently, the general UPG game can be decomposed into N subgames, each containing one submitted work of assignment and k relevant peer graders. The equilibrium of a general UPG game, described by all learner's effort on all submitted works they grade, is then composed of their respective effort in all subgames. Here we emphasize that there exists no total time limit for a learner, or equivalently, the maximum time possible to put in, which corresponds to the length of the peer grading phase, is much longer than kU, which is the maximum total time spent on grading. Thus, putting in more time grading one submission does not affect the grading of other submissions. With the help of this property, we can separate them and analyze one subgame at a time. For the rest of the paper, we define such subgame as follows to simplify the notations.

Definition 2.3. In a UPG subgame, only k learners are considered; each of them grades the same submission with value v. t_j represents the amount of effort learner j puts in grading the submitted work, dropping the superscript from t_j^i in the general game. $S_j, \hat{S}$ and α_j are defined analogously, inheriting the meaning of the counterparts with superscript. $\pi_j = \mathbf{E}[\frac{\lambda M}{k}\alpha_j - r_j t_j]$ is now the utility learner a_j gets directly from his peer grading effort in this subgame.

Note that N becomes irrelevant once we separate a UPG game into subgames, and we can fully describe such a subgame by specifying $(M, k, \mathbf{r}, U, \lambda, f_{\text{g}}, f_{\text{agg}}(\cdot), f_{\text{accu}}(\cdot))$. We call this tuple a *setting* of a UPG subgame. Note that here f_{g} represents only k functions.

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

Chapter 3

Analysis

3.1 Encouraging Conditions and Existence of Equilibria

In this section, we define the following encouraging conditions and prove that these conditions lead to the existence of pure Nash equilibria in a UPG subgame.

Definition 3.1. A setting of a UPG subgame $(M, k, \mathbf{r}, U, \lambda, f_{\text{g}}, f_{\text{agg}}(\cdot), f_{\text{accu}}(\cdot))$ satisfies *the Encouraging Conditions* if:

- (EC-1)

  Given the other players' strategies $T_{-j} = [t_1, t_2, ..., t_{j-1}, t_{j+1}, ..., t_k]$,

  $\mathbf{E}[\alpha_j](T_{-j}, t_j)$ is non-decreasing and strictly concave on $t_j \in [0, U]$.

- (EC-2)

  Let T_{-j} and T'_{-j} be two strategy profiles satisfying the following properties:

  – $t_p < t'_p$ for some p,

  – $t_q = t'_q \, \forall \, q \neq p$,

  Then $\frac{\partial}{\partial t_j}\mathbf{E}[\alpha_j](T_{-j}, t_j) < \frac{\partial}{\partial t_j}\mathbf{E}[\alpha_j](T'_{-j}, t_j)$, $\forall t_j$.

Intuitively, EC-1 says that if a grader puts more effort in grading, his accuracy improves, with a diminishing marginal effect. The following lemma says that, combining with a linear cost on time, EC-1 makes the decision problem for the grader straightforward.

doi:10.6342/NTU201701049

Lemma 3.1. *If a setting of a UPG subgame satisfies EC-1, then given any T_{-j}, $\pi_j(T_{-j}, t_j)$ is strictly concave on $t_j \in [0, U]$, and the best response t_j^* is unique. Also, either $\frac{\partial}{\partial t_j}\mathbf{E}[\alpha_j](T_{-j}, t_j)\Big|_{t_j^*} = \frac{kr_j}{\lambda M}$, or $t_j^* \in \{0, U\}$.*

Proof.

$$\begin{aligned}\frac{\partial^2(\pi_j)}{\partial(t_j)^2} &= \frac{\partial^2}{\partial(t_j)^2}\frac{\lambda M}{k}\mathbf{E}[\alpha_j](T_{-j}, t_j) - \frac{\partial^2(rt_j)}{\partial(t_j)^2} \\ &= \frac{\lambda M}{k}\frac{\partial^2}{\partial(t_j)^2}\mathbf{E}[\alpha_j](T_{-j}, t_j) \\ &< 0,\end{aligned}$$

which gives the first statement.

The last inequality is from the concavity of $\mathbf{E}[\alpha_j]$, required in EC-1. Since the partial utility function is concave, its global maximum either lies on the boundary or satisfies the first-order condition

$$\begin{aligned}\frac{\partial(\pi_j)}{\partial(t_j)}\Big|_{t^*_j} &= \frac{\partial}{\partial(t_j)}\frac{\lambda M}{k}\mathbf{E}[\alpha_j](T_{-j}, t_j) - r_j t_j\Big|_{t^*_j} \\ &= \frac{\lambda M}{k}\frac{\partial}{\partial(t_j)}\mathbf{E}[\alpha_j](T_{-j}, t_j)\Big|_{t^*_j} - r_j \\ &= 0,\end{aligned}$$

which gives the second statement. □

Note that in the boundary cases,

- $\frac{\partial}{\partial t_j}\mathbf{E}[\alpha_j](T_{-j}, t_j) > \frac{kr_j}{\lambda M}, \forall t_j \in [0, U]$, if $t^*_j = U$.
- $\frac{\partial}{\partial t_j}\mathbf{E}[\alpha_j](T_{-j}, t_j) < \frac{kr_j}{\lambda M}, \forall t_j \in [0, U]$, if $t^*_j = 0$.

On the other hand, EC-2 states that, an arbitrary grader unilaterally increasing his own grading effort will increase the marginal utilities for all other graders. Intuitively, although without specifications, we generally think of the grading possibility functions to be distributed closer to the intrinsic value if the effort is increased. Therefore, unilaterally increasing effort will make the concensus grade also distributed closer to the intrinsic

doi:10.6342/NTU201701049

value. This effect is generally good for graders since they can only invest effort to stand closer to the intrinsic value, instead of the concensus. If this effect can encourage graders to invest more effort, then we can expect a positive reinforcement process. The following theorem shows that pure Nash equilibria always exist if both conditions are satisfied. The positive reinforcement process can be seen in the proof.

Theorem 3.2. *In any UPG subgame that satisfies both encouraging conditions, there exists at least one pure Nash equilibrium.*

Proof. We prove the existence of pure Nash equilibria by describing a virtual algorithm that "calculates" one.

Given the setting of the UPG subgame, initially we let $T_i = 0\,\forall i$, or equivalently no grader puts in any kind of effort. We then modify the effort levels one by one in the order of i if an equilibrium is not reached yet. When modifying T_i, we fix all other effort levels T_{-i}, and move T_i to the best response of grader i, with respect to T_{-i}.

Trivially, since all effort levels are initialized to be zero, T_i either stays at zero or is raised upwards when it is modified for the first time. Also by EC-2, whenever one of the effort levels is raised upwards, all marginal utilities for all other graders will be increased.

Suppose the effort level of grader j, now temporarily set to T_j, is being modified for the second or more time, and no effort levels were decreased in the previous $k-1$ modifications.

By the previous lemma, if the marginal utility, $\frac{\partial}{\partial(t_j)}\mathbf{E}[\alpha_j](T_{-j}, t_j) - \frac{kr_j}{\lambda M}$, is negative in the whole interval $[0, U]$, then the best response for grader j will remain at zero regardless of all others' strategies. Suppose not, then in the previous round, T_j is set to make the marginal utility at zero. By the effects of EC-2, the marginal utility is now non-negative after a whole round of modifications. Suppose in the current round the other graders' strategies are T_{-j}, and T_j is about to be modified to T'_j. From EC-1, since $\frac{\partial}{\partial(t_j)}\mathbf{E}[\alpha_j](T'_{-j}, t_j) = \frac{kr_j}{\lambda M} \leq \frac{\partial}{\partial(t_j)}\mathbf{E}[\alpha_j](T_{-j}, t_j)$, we have $T'_j \geq T_j$. Combined with the fact that all effort levels can only increase in the first round, by induction we can prove that all effort levels are never decreased.

Since the effort levels cannot exceed U, our algorithm will eventually converge to a stable

doi:10.6342/NTU201701049

state when k consecutive effort levels are not modified in their turns, which gives a pure Nash equilibria. However, this algorithm does not converge in any guaranteed time limit, so it has little use of calculating the equilibria in practice. □

3.2 Encouraging Peer Grading

In the previous section we have already seen the ratio $\frac{kr_i}{\lambda M}$, a grader's marginal accuracy reward in equilibrium. Any marginal accuracy reward at least this large will justify putting in more time, hence it is the *effective time-to-grade ratio*. While r_i is an exogenous parameter, k, M and λ is specified by the mechanism designer. Thus, the designer can increase or decrease this ratio to any arbitrary extent. Below we describe its effect on the possible equilibria.

Definition 3.2. The effective time-to-grade ratio for grader i in a UPG subgame is $\bar{r}_i = \frac{kr_i}{\lambda M} = \bar{r}r_i$, where $\bar{r} = \frac{k}{\lambda M}$.

Theorem 3.3. *Assume k is fixed. Suppose both EC-1 and EC-2 are satisfied in a UPG subgame G_1 with $\bar{r} = \bar{r}_1$, and G_2, G_3 are UPG subgames that differs with G_1 only in $\bar{r}_2 > \bar{r}_1 > \bar{r}_3$. If $T_1 = [t_{1i}]$ is an equilibrium in G_1, then*

- *There exists an equilibrium T_2 in G_2 where $t_{2i} \leq t_{1i}\, \forall i$.*
  If the i-th equality holds, then $t_{1i} = 0$.

- *There exists an equilibrium T_3 in G_3 where $t_{3i} \geq t_{1i}\, \forall i$.*
  If the i-th equality holds, then $t_{1i} = U$.

Proof. We first prove the first statement and assume that $t_i > 0\, \forall\, i$.
For convenience, we denote $\delta_i(X_{-i}, x_i) = \frac{\partial}{\partial t_i}\mathbf{E}[\alpha_i](T_{-i}, t_i)\Big|_{T_{-i}=X_{-i},\, t_i=x_i}$ to be the slope of expected reward of player i, when player i spends x_i and the overall strategy profile of all other players is X_{-i}. We know that $\delta_i(X)$ is nonincreasing on x_i (which comes from EC-1) and increasing on every element in X_{-i} (which comes from EC-2), and that $\delta_i((T_1)_{-i}, t_{1i}) = \bar{r_1}r_i\, \forall i$.
Suppose $\delta_i(\mathbf{0}, 0) \leq \bar{r_2}r_i\, \forall i$, then $\delta_i(\mathbf{0}, y) \leq \delta_i(\mathbf{0}, 0) \leq \bar{r_2}r_i\, \forall i$, which means $T = \mathbf{0}$ is

doi:10.6342/NTU201701049

an equilibrium in G_2. Otherwise there exists j such that $\delta_j(\mathbf{0}, 0) > \bar{r_2} r_j$, which means there exists $z > 0$ such that $\delta_j(\mathbf{0}, z) = \bar{r_2} r_j$. Since $\bar{r_2} > \bar{r_1}$, there exists $y < t_{1j}$ such that $\delta_j((T_1)_{-j}, y) = \bar{r_2} r_j$. Consequently, $\delta_j(\mathbf{0}, y) < \delta_j((T_1)_{-j}, y) = \bar{r_2} r_j$, so $z < y$. Let Br' be the best response function in G_2. Br' is continuous, and we have that $0 < z = Br'(\mathbf{0}) < y = Br'((T_1)_{-j}) < t_{1j}$.

Though the best response function takes $k - 1$ arguments, we now limit the degree of freedom to 1, by restricting the input vector to be parallel to $(T_1)_{-j}$. Let $(\beta T_1)_{-j}$ be the strategy profile of all graders except j that satisfies $(\beta t_1)_i = \frac{\beta}{U} t_{1i} \forall i$. Denote $\tilde{Br}'(\beta) = Br'((\beta T_1)_{-j})$, then we have $0 < z = \tilde{Br}'(0) < y = \tilde{Br}'(U) < t_{1j} \leq U$. By the fixed-point theorem there exists at least one $w \in (z, y)$ s.t. $\tilde{Br}'(w) = w$. This means there exists an equilibrium in G_2 where $t_j = w < t_{1j}$ and $t_i = \frac{w}{U} t_{1i} < t_{1i} \forall i \neq j$.

By the same method we can prove the equality cases. Furthermore, the second statement can be proved analogously by comparing $\delta_i(\mathbf{U}, U)$ to $\bar{r_3} r_i$. □

Equivalently, increasing the effective time-to-grade ratio distorts the equilibrium downwards, and vice versa; extreme equilibria are preserved if the environment is made even more extreme. Suppose all functions are smooth, the equilibria will move continuously with the fluctuation of $\bar{r}$. In fact, $\bar{r}$ is what the mechanism designer can really control. Decreasing λ or M all leads to an increase of $\bar{r}$, which "discourages" grading behavior and moves all the equilibria downwards; while increasing either λ or M "encourages" grading behavior, and moves all the equilibria upwards. Certainly, in practice λ and M cannot be raised without limitation, which will be discussed in Section 6.

The above implications may seem trivial at a first glance; learners invest effort and should get reward from that. However, we should point out again that our mechanism works under a subtle limitation that instructors cannot perceive and evaluate the learners' effort, but only the grading outcomes. Therefore whole mechanism employs a "reward comes from outcomes" method. We have shown that, this method is as effective as an ideal "reward comes from effort" scenario, and the idea of encouraging learners only by rewarding their closeness to concensus is, game-theoretically, indeed effective.

Last, we can argue that all the above results in Section 3.1 and 3.2 will still hold even if

doi:10.6342/NTU201701049

the expected accuracy level mentioned in EC-1 is only weakly concave. While this means that graders generally have multiple best responses to choose from, choosing the highest effort level among those best responses does not violate rationality. Therefore the positive reinforcement still exists, and all the other results follow. The only difference is that there will be far more possible equilibria.

3.3 Settings That Meet the Encouraging Conditions

So far we still do not know under what settings of a UPG subgame that the encouraging conditions will hold. In this section, we describe a series of practical settings of a UPG subgame that satisfy both encouraging conditions. While the encouraging conditions are rather strong, we show that they can be satisfied with some rational assumptions and a simple mechanism.

The first question is how to characterize the peer-assessed grades. We assume the grade to follow a normal distribution here, which is common when characterizing an observation. The mean will simply be the intrinsic value as we assume unbiased grading. The variance depends on the amount of time. Intuitively, the more effort the grader puts in grading, the closer the grade will be to the intrinsic value. Furthermore, the marginal effect should be diminishing. Thus we assume the variance to be non-increasing and convex in the time spent on grading. Notice this does not mean a grader can become arbitrarily precise if he spends a very large amount of time on grading, since the above function may not be strictly decreasing. For the mechanism, we use an averaging method to aggregate the grades. We proved that, combined with any non-increasing piecewise continuous $f_{\text{accu}}(\cdot)$, this setting satisfies both encouraging conditions.

Definition 3.3. We define the averaging function to be $f_{\text{avg}}(\mathbf{S}) = \frac{1}{k}\sum_j(S_j)$.

Proposition 3.4. *Suppose that for every j, $F_j(\cdot, v, t_j) \sim N(v, g_j^2)$ where $g_j = g_j(t_j)$ is a non-increasing convex function. If $f_{\text{agg}}(\cdot) = f_{\text{avg}}(\cdot)$, $f_{\text{accu}}(\cdot)$ is non-increasing piecewise continuous, then for any values of $(M, k, \mathbf{r}, U, \lambda)$, $(M, k, \mathbf{r}, U, \lambda, f_{\text{g}}, f_{\text{agg}}(\cdot), f_{\text{accu}}(\cdot))$ satisfies both EC-1 and EC-2.*

doi:10.6342/NTU201701049

To prove this proposition, we start by proving a weaker proposition where $f_{\text{accu}}(\cdot)$ is limited to a threshold function, which means full mark is given once a learner gives a grade not too far off from the consensus. Then we extend the proposition to the general case.

Definition 3.4. We define the h-threshold awarding function to be $f_h(x) = 1$ if $x \leq h$ for some threshold value $0 \leq h < \infty$, and $f_h(x) = 0$ otherwise.

Proposition 3.5. *Suppose that for every j, $F_j(\cdot, v, t_j) \sim N(v, g_j^2)$ where $g_j = g_j(t_j)$ is a non-increasing convex function. If $f_{\text{agg}}(\cdot) = f_{\text{avg}}(\cdot)$, $f_{\text{accu}}(\cdot) = f_h(\cdot)$ for some $0 \leq h < \infty$, then for any values of $(M, k, \mathbf{r}, U, \lambda)$, $(M, k, \mathbf{r}, U, \lambda, f_{\text{g}}, f_{\text{agg}}(\cdot), f_{\text{accu}}(\cdot))$ satisfies both EC-1 and EC-2.*

Proof. To simplify notations let $g(\cdot) = g_j(\cdot)$. Since S_j follows a normal distribution, $\hat{S} = \frac{1}{k}\Sigma_j S_j$ also follows a normal distribution. Consider player j. Let the distribution of other graders' average score to be $F_{\text{o}} = \frac{1}{k-1}\sum\limits_{i \neq j} S_i \sim N(v, x^2)$, then the aggregated score is given by

$$\hat{S} = f_{\text{avg}}(\mathbf{S}) = \frac{1}{k}\sum_i S_i = \frac{1}{k}F_j + \frac{k-1}{k}F_{\text{o}},$$

the distance to the aggregated score is given by

$$S_j - \hat{S} = F_j - \hat{S} = \frac{k-1}{k}(F_{\text{o}} - F_j) \sim N(0, \frac{k-1}{k}(x^2 + g^2)),$$

and the expected accuracy is given by

$$\begin{aligned}
\mathbf{E}[\alpha_j](T_{-j}, t) &= \mathbf{E}[f_h(|S_j - \hat{S}|)] \\
&= Pr[|S_j - \hat{S}| \leq h] \\
&= \text{erf}(\frac{h}{\sqrt{\frac{(k-1)(x^2+g^2)}{k}}}) \\
&= \text{erf}(\frac{h\sqrt{k}}{\sqrt{(k-1)(x^2 + g^2)}}).
\end{aligned}$$

By definition, $g'(t_j) \leq 0$, $g''(t_j) \geq 0$.

doi:10.6342/NTU201701049

Let the effective error threshold $p = \frac{h\sqrt{k}}{\sqrt{(k-1)(x^2+g^2)}}$, then

$$
\begin{aligned}
\frac{\partial p}{\partial t} &= \frac{h\sqrt{k}}{\sqrt{k-1}} \frac{g(-g')}{(x^2+g^2)^{\frac{3}{2}}} \geq 0 \\
\frac{\partial}{\partial t}\mathbf{E}[\alpha_j](T_{-j}, t) &= \frac{\partial}{\partial t}\text{erf}(p) = (\frac{2}{\sqrt{\pi}}e^{-p^2})\frac{\partial p}{\partial t} \geq 0 \\
\frac{\partial^2}{\partial t^2}\mathbf{E}[\alpha_j](T_{-j}, t) &= \frac{\partial}{\partial t}(\frac{2}{\sqrt{\pi}}e^{-p^2})\frac{\partial p}{\partial t} \\
&= (\frac{2}{\sqrt{\pi}}e^{-p^2})((-2p)\frac{\partial p}{\partial t} + \frac{h\sqrt{k}}{\sqrt{k-1}}(\frac{-(g')^2+gg''}{(x^2+g^2)^{\frac{3}{2}}} - \frac{3g^2(g')^2}{(x^2+g^2)^{\frac{5}{2}}})) \\
&\leq 0.
\end{aligned}
$$

The last inequality follows from the fact that $g'(t_j)$ is non-positive and $g''(t_j)$ is non-negative. Let T_{-j} and T'_{-j} be two strategy profiles satisfying $t_p < t'_p$ for some p, and $t_q = t'_q \forall q \neq p$. Let $F_{o1} = \frac{1}{k-1}\sum_{i\neq j} S_i \sim N(v, x_1^2)$ and $F_{o2} = \frac{1}{k-1}\sum_{i\neq j} S_i \sim N(v, x_2^2)$ be respectively the other grader's average score distribution in both cases. Also let the corresponding effective error thresholds be $p_1 = \frac{h\sqrt{k}}{\sqrt{(k-1)(x_1^2+g^2)}}$ and $p_2 = \frac{h\sqrt{k}}{\sqrt{(k-1)(x_2^2+g^2)}}$. Clearly, $x_1 > x_2$, which implies $p_1 < p_2$. Then we have

$$
\begin{aligned}
\frac{\partial}{\partial t}(\mathbf{E}[\alpha_j](T_{-j}, t) - \mathbf{E}[\alpha_j](T'_{-j}, t)) &= \frac{\partial}{\partial t}(\text{erf}(h_1) - \text{erf}(h_2)) \\
&= (\frac{2}{\sqrt{\pi}}e^{-p_1^2})\frac{\partial p_1}{\partial t} - (\frac{2}{\sqrt{\pi}}e^{-p_2^2})\frac{\partial p_2}{\partial t} \\
&= (\frac{2}{\sqrt{\pi}}\frac{(k-1)g(-g')}{h^2k})(\frac{e^{-p_1^2}}{p_1^3} - \frac{e^{-p_2^2}}{p_2^3}) \\
&\leq 0.
\end{aligned}
$$

The first term, $\frac{2}{\sqrt{\pi}}\frac{(k-1)g(-g')}{h^2k}$, is non-negative since g is a non-increasing non-negative function. Since $\frac{\partial}{\partial p}\frac{e^{-p^2}}{p^3} = -\frac{e^{-p^2}(2p^2+3)}{p^4} \leq 0$, the second term is an definite integral of a non-positive function, so it is non-positive. Hence the product of the two terms is non-positive. □

Any non-increasing piecewise continuous function can be decomposed into combination of integrals of threshold functions. For each threshold function, the portion of expected utility given by it satisfies EC-1. The expected utility is now combination of

doi:10.6342/NTU201701049

integrals of non-decreasing and concave functions, thus being non-decreasing and concave itself [2]. Similarly, in the condition in EC-2, the combination of integrals of error functions from the latter strategy profile always increases more rapidly then that from the former profile.

Again, all the grading distributions need not to be the same. We showed that as long as every grader's distribution is a normal distribution with a variance non-increasing with his effort, both encouraging conditions will be satisfied.

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

Chapter 4

Homogeneous Grading

Along with the development of MOOCs, another form of online learning environments in the name of SPOCs has also emerged [10]. *Small Private Online Courses* has at most several hundreds of qualified learners. With an eye on better learning experiences, a SPOC may require learners to pass tests, complete prerequisites, or even require student status of the relating institutes, used as material in a blended-learning course. The diversity of learners is drastically decreased in this type of learning environments. While grading several hundreds of copies of assignments may not be impossible, it is still tedious and requires massive effort. Plus all its merits, peer grading can still be used in SPOCs.

Corresponding to this type of environment where the learners are more homogeneous than before, in this chapter we focus on a further limited case of our model, where the peer graders not only share identical grading abilities but also value their time identically. We show that all possible pure Nash equilibria in this scenario share a common property: a submission is graded with the same level of effort by every grader. While this is still an ideal scenario, we can learn from it what homogeneity brings to our model.

Definition 4.1. A *homogeneous unbiased peer grading subgame*, or HUPG subgame, is a UPG subgame satisfying the following properties:

- $f_1 = f_2 = \dots = f_k = f$, where f is the *shared grading probability distribution*.
- $r_i = r_2 = \dots = r_k = r$, where r is the *shared time-to-grade ratio*.

doi:10.6342/NTU201701049

When both encouraging conditions are met in an HUPG subgame, we found that in equilibrium it is impossible for two learners to put in different levels of effort. Informally, in such an equilibrium the low-effort learner's peers puts more effort in total than the high-effort one's peers. This means the low-effort learner would like to put in more effort than the high-effort learner as well, which leads to a contradiction. Consequently, we found that a submission is graded with the same level of effort by every grader, as described in the following theorem.

Theorem 4.1. *Suppose that EC-1 and EC-2 are both satisfied in an HUPG subgame. If* $T^* = \{t_1^*, t_2^*, ..., t_k^*\}$ *is a pure Nash equilibrium, then* $t_1^* = t_2^* = ... = t_k^* = t^*$.

Proof. Assume the contrary. Then there must exist $t_i^* < t_j^*$ for some $i \neq j$. Considering the strategy profiles T_{-i} and T_{-j}, we have

$$\begin{aligned}
\frac{\partial}{\partial t}\mathbf{E}[\alpha](T_{-j}, t)\bigg|_{t_i^*} &\geq \frac{\partial}{\partial t}\mathbf{E}[\alpha](T_{-j}, t)\bigg|_{t_j^*} && \text{(EC-1)} \\
&\geq \frac{kr}{\lambda M} \\
&\geq \frac{\partial}{\partial t}\mathbf{E}[\alpha](T_{-i}, t)\bigg|_{t_i^*} && \text{(Lemma 3.1)} \\
&> \frac{\partial}{\partial t}\mathbf{E}[\alpha](T_{-j}, t)\bigg|_{t_i^*}, && \text{(EC-2)}
\end{aligned}$$

which is a contradiction. □

Combining with the previous result that a pure NE always exists, we immediately have the following corollary. However we give a direct proof here.

Corollary 4.1.1. *In any HUPG subgame that satisfies both encouraging conditions, there exists at least one equilibrium where* $t_i^* = t^*, \forall i$.

Proof. Let $T_{-i}(p)$ be a strategy profile with $t_j = p \forall j \neq i$. By applying the property EC-2, we have $\frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(p), t) < \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(q), t)$ iff $p < q$. Denote $t_i^*(p)$ to be player i's

doi:10.6342/NTU201701049

best response with respect to $T_{-i}(p)$. Consider $t_i^*(0)$ and $t_i^*(U)$. If either $t_i^*(0) = 0$ or $t_i^*(U) = U$, we have a trivial equilibrium where $t^* = 0$ or $t^* = U$. Suppose not, then there are four cases:

- $0 < t_i^*(0) < U, t_i^*(U) = 0$. By lemma 3, we have

$$\begin{aligned}\frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(0),t)]\Big|_{t_i^*(0)} &= \frac{kr}{\lambda M} \\ &> \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(U),t)\Big|_{t_i^*(0)} \\ &> \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(0),t)\Big|_{t_i^*(0)},\end{aligned}$$

  which is a contradiction.

- $t_i^*(0) = U, 0 < t_i^*(U) < U$. Similarly, we have

$$\begin{aligned}\frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(0),t)\Big|_{t_i^*(U)} &> \frac{kr}{\lambda M} \\ &= \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(U),t)\Big|_{t_i^*(U)} \\ &> \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(0),t)\Big|_{t_i^*(U)},\end{aligned}$$

  which is a contradiction.

- $t_i^*(0) = U, t_i^*(U) = 0$. We have

$$\frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(0),t) > \frac{kr}{\lambda M} > \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(U),t) > \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(0),t)$$

  for all $t \in [0, U]$, which is also a contradiction.

doi:10.6342/NTU201701049

- $0 < t_i^*(0), t_i^*(U) < U$. This gives

$$\begin{aligned}\frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(0),t)\bigg|_{t_i^*(0)} &= \frac{kr}{\lambda M} \\ &= \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(U),t)\bigg|_{t_i^*(U)} \\ &> \frac{\partial}{\partial t}\mathbf{E}[\alpha_i](T_{-i}(0),t)\bigg|_{t_i^*(U)}.\end{aligned}$$

The only possible case is the last one, as the other three all result in contradictions.

Since all of the above functions are non-increasing in t, we obtain $t_i^*(U) > t_i^*(0)$ from the last inequality. Notice that in this case, $\frac{\partial}{\partial t_i}\mathbf{E}[\alpha_i](T_{-i}(x),t_i)\Big|_{t_i^*(x)} = \frac{kr}{\lambda M}$ holds for all possible x including boundaries.

Define the best response function $Br(x) = t_i^*(x)$. Since $f, f_{\text{agg}}(\cdot), f_{\text{accu}}(\cdot)$ are all continuous, $Br(\cdot)$ is continuous in x. Observed that $0 < Br(0) < Br(U) < U$, the fixed-point theorem applies. Hence there exists $x \in (Br(0), Br(U))$ s.t. $Br(x) = x$, which gives the equilibrium $t^* = x$. □

Since an HUPG subgame is a special case of a UPG subgame, Theorem 3.3 still holds in this scenario. Moreover, all the pure NEs can be directly computed. Denote $\delta(x)$ to be the slope of expected reward of player i when all players puts in x unit of effort, which is equivalent to $\delta(\mathbf{x}, x)$ in the proof of Theorem 3.3, then if $t^* = x$ is an equilibrium, $\delta(x) = \bar{r}$ or $x \in \{0, U\}$. We can find all the equilibria by finding all roots of $\delta(x) = \bar{r}$ and checking if either possible extreme equilibria exists.

doi:10.6342/NTU201701049

Chapter 5

Peer Grading with Irrational Players

So far we have assumed that all learners rationally seek to maximize their own expected utilities. However, this behavior model does not fully describe all possible behaviors of a learner, as we often see in real courses. First, dropouts are way too common in today's MOOCs that researchers use machine learning techniques to predict them [12,14]; learners who drop out will not finish grading. Next, some learners may have already accumulated sufficient points to pass the course. Extra points may not serve as motivation if they do not affect whether the course is passed. There might also be devoted learners who feel obliged to put maximum effort in each task. As proposed in [18], we may need extra graders on some particular assignments, who can be a TA, who has nothing to do with the rewards and will always put sufficient amount of effort into grading.

Here we define a modified game model, in which some of the players do not seek maximized utility. We assume that such players have different utility valuations, and will decide their effort levels independent of whatever the others do. We assume that the effort levels are known to all other players in the game. We show that given this information, the behavior of all rational graders should resemble our findings in the previous sections.

Definition 5.1. In *a UPG subgame with irrational players*, there are k total graders. Graders a_1 to $a_{k'}$ are rational players as before with time-to-grade ratios r_1 to $r_{k'}$; for $k' + 1 \leq i \leq k$, grader a_i will simply put effort c_i regardless of what the others do. All values of c_i are public information. All other parts in the model remain unchanged. We

doi:10.6342/NTU201701049

define the constant effort vector $\mathbf{C} = [c_i]$. We call it *an HUPG subgame with irrational players* if $r_1 = r_2 = ... = r_{k'}$.

Suppose the encouraging conditions are satisfied. We observe that Theorem 3.3 still holds here, as the value of $\bar{r}$ still affects the effort levels in equilibria. Besides, they are also influenced by the irrational graders' efforts. It is rather intuitive that all rational graders will raise their efforts once an irrational grader raises his effort level, and vice versa. The designer can influence the equilibria if he can influence the irrational graders as well.

For simplicity, we will now omit the constant strategies by the irrational graders and use $t = [t_1, t_2, ..., t_{k'}]$ to represent the rational graders' strategies in an equilibrium. The following theorem can be proved analogously to Theorem 3.3.

Theorem 5.1. *Assume k and k' are fixed. Suppose both EC-1 and EC-2 are satisfied in a UPG subgame G_1 with irrational players playing strategies $\mathbf{C}$, and $\bar{r} = r_1$. If $T_1 = [t_{1i}]$ is an equilibrium in G_1, then:*

- *For a UPG subgame G_2 that differs with G_1 only in $\bar{r} = r_2 > r_1$, there exists an equilibrium where $t_{2i} \leq t_{1i} \forall i$.*
  If the i-th equality holds, then $t_{1i} = 0$.

- *For a UPG subgame G_3 that differs with G_1 only in $\bar{r} = r_3 < r_1$, there exists an equilibrium where $t_{3i} \geq t_{1i} \forall i$.*
  If the i-th equality holds, then $t_{1i} = U$.

- *For a UPG subgame G' that differs with G_1 only in $c'_i < c_i$ for an irrational player a_i, there exists an equilibrium T' where $t'_i \leq t_{1i} \forall i$.*
  If the i-th equality holds, then $t_{1i} = 0$.

- *For a UPG subgame G'' that differs with G_1 only in $c''_i > c_i$ for an irrational player a_i, there exists an equilibrium T'' where $t''_i \geq t_{1i} \forall i$.*
  If the i-th equality holds, then $t_{1i} = U$.

The first two results are analogous to Theorem 3.3. In the third situation, since there exists some $c'_i < c_i$, by EC-2 we have $\delta'(x, \mathbf{C}) < \delta_1(x, \mathbf{C}) \forall x \in [0, U]$. This effect is

doi:10.6342/NTU201701049

similar to have $\bar{r}$ raised by the difference of the two functions, though this amount depends on x. However, we did not need the precise amount to get the first two results; all we needed is the fact that $\bar{r}$ is increased. As a consequence all equilibria moves downwards, as in the first situation. Analogously, the last situation is similar to the second situation.

Similarly, in the homogeneous case Theorem 4.1 still holds as follows:

Theorem 5.2. *Suppose encouraging conditions are satisfied in an HUPG subgame with k' rational players. If T^* is a Nash equilibrium, then $t_1^* = t_2^* = ... = t_{k'}^* = t^*$. There exists at least one such equilibrium.*

The proof is very similar to that in Chapter 4, except for we can only limit our comparisons among the rational players' strategies. The decision process of one grader does not take into account whether the other graders are rational, but only their actions. Analogous to the proof in Chapter 4, we can show that it is still impossible for a pair of rational players to give different level of effort. The irrational players can be viewed as a part of the "environment".

However, the equilibria level does not remain the roots of $\delta(t^*) = \bar{r}$, since the irrational players do not all gives t^* units of time. We will need to consider those effort levels. Let $\delta(x, \mathbf{C}) = \frac{\partial}{\partial t_i}\mathbf{E}[\alpha_i](T_{-i}(x), t_i)\Big|_{t_i=x}$ be the slope of expected rewards for a rational grader when all rational graders puts in x units of effort, and the irrational graders follow their predefined strategies. The non-extreme equilibria can now be calculated by solving $\delta(x, \mathbf{C}) = \bar{r}$, and the extreme ones can be found by checking the values of $\delta(0, \mathbf{C})$ and $\delta(U, \mathbf{C})$.

It appears that the mechanism designer has even more power in this scenario, since he can now influence the equilibria through the irrational players. However there is a caveat: all the irrational players' strategies must be public, which might not be practical aside of announcing what the TAs will do. Nonetheless, we can still expect the rational players to adjust their behavior based on their observations and expectations about the irrational players, like correctly forecasting some peers to stop putting effort towards the end of the course.

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

Chapter 6

Discussion and Future Work

6.1 Other Game Settings

Our model is very flexible as the mechanism designer can dictate all of the parameters and functions, i.e. $M, k, \lambda, f_{\text{agg}}(\cdot), f_{\text{accu}}(\cdot)$. While we know that the first three parameters have no effect on whether the encouraging conditions are satisfied, there are many aggregating and accuracy-calculating functions that can be plugged in. As mentioned in Chapter 2, Crowdgrader [6] utilizes the Olympic average of all scores to determine the aggregated grade, and Coursera uses the median [5]. While we proved that an average function paired with almost all reasonable accuracy-calculating function will do the trick, whether the other aggregating methods can satisfy the encouraging conditions are still unknown.

6.2 Setting Up the Parameters in Practice

We know that both increasing M and λ can encourage more effort on grading. In theory, one can make M and λ as large as possible, as this induces maximum effort possible from the learners. But maximizing M means that letting the relevant assignment account for one hundred percent of the grades in the course, while maximizing λ means the final grade of the assignment come exclusively from peer grading accuracy. Learners will simply stop doing assignments if all their utilities come from their grades this way.

Since the assessment rules should reflect the learning goals, the best ratio should be

doi:10.6342/NTU201701049

determined by how important the course provider thinks about the skills learners get from doing peer grading. While the goals are subjective, the benefits from doing peer grading should not be neglected, and the course provider can keep in mind that the more peer-grading reward weighs, the more effort it induces. Teachers and lecturers commonly offer bonus points corresponding to skills beyond their teaching goals only to stimulate learning motivation, and rewarding accurate peer grading can be viewed alike.

Finally, in some of the current MOOCs the idea of learning goals are taken to extreme so that there are no final grades in them. Learners only pass the course by finishing all the requirements or they do not. Often the requirements are still assessed to determine if a particular level is reached, and one can still utilize our ideas in them. Learners may even have a greater chance to remain a rational grader throughout the course in this situation. Experimental and empirical studies will be needed to find the real influence of the parameters on not only the grading accuracy but also the learning effect as well.

6.3 Average Amount of Graders

In the dynamics of equilibria we have always kept k (and k', if irrational players exist) fixed. While it is clear that increasing k will also increase $\bar{r}$ and trigger the subsequent effects, it influences the equilibria in another way; that is, the size of the "environment" increases as well. Even we fix all the graders' effort levels, the probabilistic distribution of the other graders' aggregated score will not remain the same if the number of graders fluctuate. To be more explicit, when k is increased, that distribution tends to become more dense around the mean, and the crowd opinion is more accurate thanks to a larger crowd size. Such effect is similar to have the others increase their efforts and will encourage the graders to put more effort as well, in contrast to its effect on $\bar{r}$. Thus the influence of tuning k is more complicated then that of M and λ. It is not clear which side of the effects outweighs another in various model settings, and it is intuitive that both situations are possible given the flexibility of our model.

The best amount of graders appointed to grade one submission is long-discussed, and the consensus suggests around three to six graders is enough [4, 9, 17], as it takes at least

doi:10.6342/NTU201701049

several graders to make the aggregation sufficiently accurate. While intuitively more graders results in more accurate aggregation, now we found that too much graders may not do well if accuracy is rewarded. There are no previous experiments focusing on this amount with accuracy rewarded. One such experiment is not very hard to setup given the existing platforms like that in [6].

6.4 Unbalanced Peer Grading Tasks

It is assumed that k is a hard rule determined by the course provider. Our model does not handle unfinished grading tasks, since they are different with giving a zero grade. Also, in current MOOCs platforms it is possible for learners to actively ask for more peer grading tasks. From the perspective of course providers, such requests should be welcomed.

One interesting question is what will happen if we ease the hard limit on k. The course provider can let learners decide themselves how many copies of assignments they would like to grade; perhaps the relevant reward can be granted in a linear way. To further investigate this scenario, we must make clear how to arrange the grading relations once the limit on k is removed. If a learner asks for a new assignment to grade, which one does the system select? How to aggregate the concensus if the assignments are graded by different amount of learners?

Suppose somehow the grading tasks can be arranged somewhat equally so that every assignment is still graded by almost identical amount of learners. What will be the learners' rational behaviors? Will everyone grade identical copies of assignments in the homogeneous case? Does the course provider still have the power to induce more efforts on grading by tuning parameters? These will be interesting questions to explore.

doi:10.6342/NTU201701049

6.5 Towards Biased Grading

So far we assume that undergrading and overgrading happen equally likely. However, Sadler and Good [19] suggested that peer-grading tends to undergrade, at least compared to self-grading; Freeman [11] suggested peer-grading can sometimes overgrade, especially on harder stuff, likely due to the learners' inability to find subtle mistakes. Piech et al. [18] tried to exploit the biases of graders to aggregate more accurate scores.

Theoretically, such biases can be measured from all the peer grading records of any given learner, as large MOOC platforms keep user data efficiently. This gives the possibility to preprocess the peer grades, hopefully to eliminate the bias before aggregating them. However, statistically the biases may take too long to stabilize. A learner can show different grading bias in different courses as well, as biases are somewhat related to learners' ability. One can even argue that it is inappropriate to use past grading performances to determine future grading bias; the learner might start putting more effort anytime, for example. The troublesome grading biases need to be considered carefully in practice.

Our analysis and results will remain intact if all the graders have homogeneous bias. This is intuitive, since all the accuracy measurements comes from comparison between a single grade and the consensus grade. However it will not be the case if the graders are biased heterogeneously. Analysis can be made if the biases of learners can be approximated by probabilistic distributions; for example, in [18] learners are assumed to have a random bias following a normal distribution. An interesting question is whether the encouraging conditions will remain satisfied, if the learners become biased, but unbiased in expectation.

6.6 Concensus Grading

We utilize concensus grading throughout our model. One can decide the definitions of concensus: the average, the median, or even the majority of a Yes/No question. But we should keep in mind that all these measurements are merely aggregates of individual grading outcomes, and individual grading outcomes are impossible to be fully objective. We

doi:10.6342/NTU201701049

cannot expect objectivity in peer grading, which is why we should use peer grading carefully and with limitation. However, most of the time crowd subjectivity is the best we can do, as instructors may not be able to offer objectivity as well.

doi:10.6342/NTU201701049

doi:10.6342/NTU201701049

Bibliography

[1] D. Boud, R. Cohen, and J. Sampson, editors. *Peer learning in higher education*. 2001.

[2] S. Boyd and L. Vandenberghe. *Convex optimization*, section 3.2, page 79. Seventh edition, 2004.

[3] A. U. Carbonara, A. Datta, A. Sinha, and Y. Zick. Incentivizing peer grading in moocs: An audit game approach. In *International Joint Conference on Artificial Intelligence*, ICJAI '15, pages 497–503, 2015.

[4] K. Cho, C. D. Schunn, and R. W. Wilson. Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Educational Psychology*, 98(4):891–901, 2006.

[5] Coursera. See feedback and grades for peer reviewed assignments, 2017.

[6] L. de Alfaro and M. Shavlovsky. Crowdgrader: a tool for crowdsourcing the evaluation of homework assignments. In *Proceedings of the 45th ACM technical symposium on Computer science education*, SIGCSE '14, pages 415–420, 2014.

[7] L. de Alfaro and M. Shavlovsky. Dynamics of peer grading: An empirical study. In *Proceedings of the 9th International Conference on Educational Data Mining*, EDM '16, pages 62–69, 2016.

[8] F. Dochy, M. Segers, and D. Sluijsmans. The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3):331–350, 1999.

doi:10.6342/NTU201701049

[9] N. Falchikov and J. Goldfinch. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3):287–322, 2000.

[10] A. Fox. From moocs to spocs. *Communications of the ACM*, 56(12):38–40, 2013.

[11] S. Freeman and J. W. Parks. How accurate is peer grading? *CBE-Life Sciences Education*, 9(4):482–488, 2010.

[12] S. Halawa, D. Greene, and J. Mitchell. Dropout prediction in moocs using learner activity features. *Experiences and best practices in and around MOOCs*, (37):7–16, 2014.

[13] M. B. Hoy. Moocs 101: an introduction to massive open online courses. *Medical reference services quarterly*, 33(1):85–91, 2014.

[14] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart. Predicting mooc dropout over weeks using machine learning methods. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pages 60–65, 2014.

[15] N.-F. Liu and D. Carless. Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, 11(3):279–290, 2006.

[16] Y. Lu, J. Warren, C. Jermaine, S. Chaudhuri, and S. Rixner. Grading the graders: Motivating peer graders in a mooc. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 680–690, 2015.

[17] H. Luo, A. C. Robinson, and J.-Y. Park. Peer grading in a mooc: Reliability, validity, and perceived effects. *Journal of Asynchronous Learning Networks*, 18(2):n2, 2014.

[18] C. Piech, J. Huang, Z. Chen, C. B. Do, A. Y. Ng, and D. Koller. Tuned models of peer assessment in moocs. In *Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013)*, EDM '13, pages 153–160, 2013.

[19] P. M. Sadler and E. Good. The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1):1–31, 2006.

doi:10.6342/NTU201701049

[20] L. A. Stefani. Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, 19(1):69–75, 1994.

[21] K. Topping. Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3):249–276, 1998.

doi:10.6342/NTU201701049