Encouraging Peer Grading in MOOCs

Shao-Heng Ko

Collective Algorithms Lab
National Taiwan University
shaohengko@gmail.com

Jan 17, 2020

About Us

Shao-Heng

About Us

Shao-Heng

Cheng-Yu

About Us

Shao-Heng

Cheng-Yu

- We are from Collective Algorithms Lab, EE dept.

About Us

Shao-Heng

Cheng-Yu

- We are from Collective Algorithms Lab, EE dept.
  - led by Prof. Ho-Lin Chen

About Us

Shao-Heng

Cheng-Yu

- We are from Collective Algorithms Lab, EE dept.
  - led by Prof. Ho-Lin Chen
  - we are interested in computation with multiple parties

About Us

Shao-Heng

Cheng-Yu

- We are from Collective Algorithms Lab, EE dept.
  - led by Prof. Ho-Lin Chen
  - we are interested in computation with multiple parties
    - Distributed Computing

About Us

Shao-Heng

Cheng-Yu

- We are from Collective Algorithms Lab, EE dept.
  - led by Prof. Ho-Lin Chen
  - we are interested in computation with multiple parties
    - Distributed Computing
    - Molecular Computing

About Us

Shao-Heng

Cheng-Yu

- We are from Collective Algorithms Lab, EE dept.
  - led by Prof. Ho-Lin Chen
  - we are interested in computation with multiple parties
    - Distributed Computing
    - Molecular Computing
    - Computational aspect of Games and Mechanisms

Algorithmic Game Theory and Mechanism Design

- Understanding and developing efficient algorithms for:
  - analyzing games: Algorithmic Game Theory
  - designing mechanisms: Algorithmic Mechanism Design

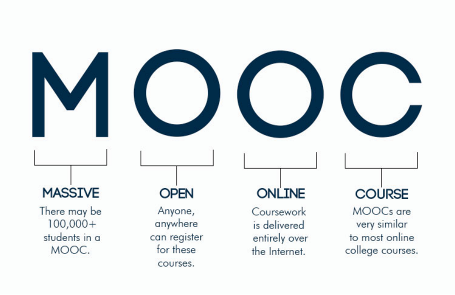
Algorithmic Game Theory and Mechanism Design

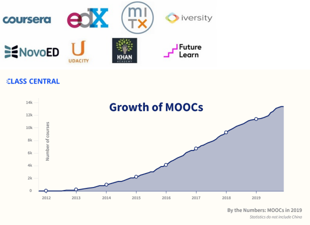
- Understanding and developing efficient algorithms for:
  - analyzing games: Algorithmic Game Theory
  - designing mechanisms: Algorithmic Mechanism Design
- Algorithmic Game Theory vs. Game Theory
  - Inefficiency of equilibria
  - Computational complexity of finding equilibria

Algorithmic Game Theory and Mechanism Design

- Understanding and developing efficient algorithms for:
  - analyzing games: Algorithmic Game Theory
  - designing mechanisms: Algorithmic Mechanism Design
- Algorithmic Game Theory vs. Game Theory
  - Inefficiency of equilibria
  - Computational complexity of finding equilibria
- Algorithmic Mechanism Design vs. Mechanism Design
  - Care about universal and worst-case results
  - Care about implementability, e.g., polynomial-time computable mechanisms

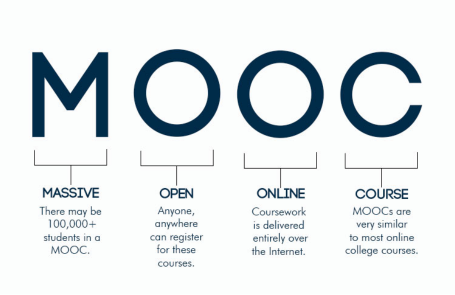
Outline

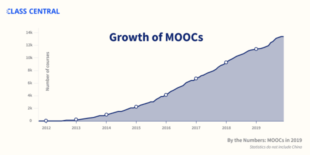




- Emerged as a mainstream trend of learning in 2012 (the "year of MOOCs")

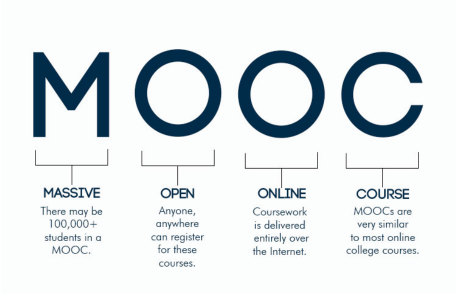
[1] https://www.classcentral.com/report/mooc-stats-2019/.

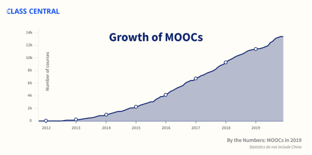




- Emerged as a mainstream trend of learning in 2012 (the "year of MOOCs")
- Number of engaged learners reached 110M in 2019.[1]

[1] https://www.classcentral.com/report/mooc-stats-2019/.





- Emerged as a mainstream trend of learning in 2012 (the "year of MOOCs")
- Number of engaged learners reached 110M in 2019.[1]
- 13.5K courses by 900+ universities by 2019.

[1] https://www.classcentral.com/report/mooc-stats-2019/.

Peer Grading/Assessment in MOOCs

Peer Grading/Assessment in MOOCs

- Extremely high amount of learners overload course staff

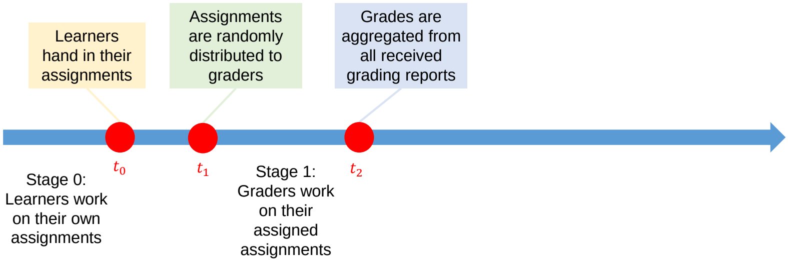
Peer Grading/Assessment in MOOCs

- Extremely high amount of learners overload course staff
- Deployed and analyzed in Coursera.org in large-scale as early as [Piech et al., 2013]

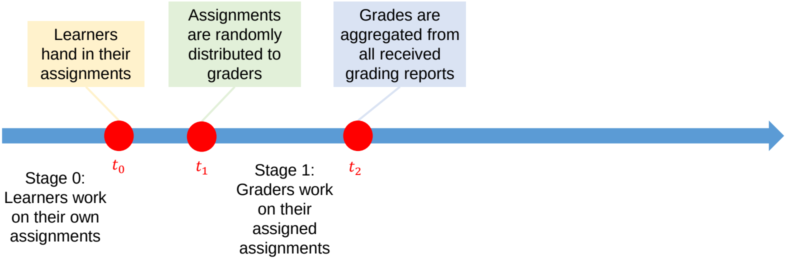
Peer Grading/Assessment in MOOCs

- Extremely high amount of learners overload course staff
- Deployed and analyzed in Coursera.org in large-scale as early as [Piech et al., 2013]
- Still the only practical solution to grading high-level assignments in MOOCs.

Basic Peer Grading Mechanism

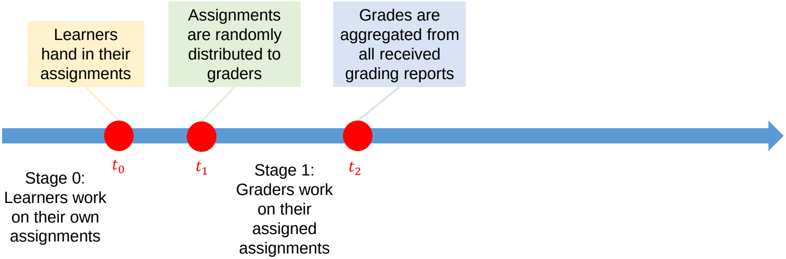


Basic Peer Grading Mechanism



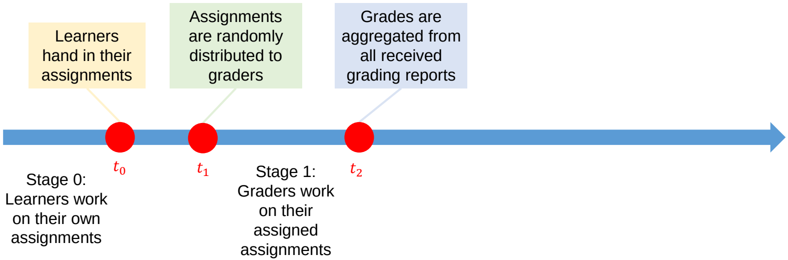
- People studied
  - Bias [Piech et al., 2013]

Basic Peer Grading Mechanism



- People studied
  - Bias [Piech et al., 2013]
  - Aggregation Methods, e.g., cardinal v.s. ordinal [Raman and Joachims, 2014, Mi and Yeung, 2015, Caragiannis et al., 2015]

Basic Peer Grading Mechanism



- People studied
  - Bias [Piech et al., 2013]
  - Aggregation Methods, e.g., cardinal v.s. ordinal [Raman and Joachims, 2014, Mi and Yeung, 2015, Caragiannis et al., 2015]
  - Incentives

A Moral Hazard Problem

- Lack of effort in grading [Lu et al., 2015].

A Moral Hazard Problem

- Lack of effort in grading [Lu et al., 2015].
- Efforts put in grading others' work were not reflected in concrete reward

A Moral Hazard Problem

- Lack of effort in grading [Lu et al., 2015].
- Efforts put in grading others' work were not reflected in concrete reward
- Only absent graders, instead of ineffective ones, were punished

A Moral Hazard Problem

- Lack of effort in grading [Lu et al., 2015].
- Efforts put in grading others' work were not reflected in concrete reward
- Only absent graders, instead of ineffective ones, were punished
- An improved mechanism that incentivizes effort in peer grading is necessary.

Proposed approaches

- Through punishment: a Stackelberg game that limited TA resources are efficiently allocated to double check the grading [Carbonara et al., 2015]

[2] https://www.crowdgrader.org/

Proposed approaches

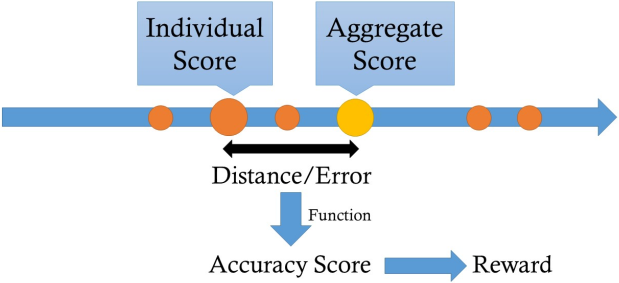
- Through reward: Crowdgrader[2]: incentivizing accurate graders directly [de Alfaro and Shavlovsky, 2014]

crowdgrader

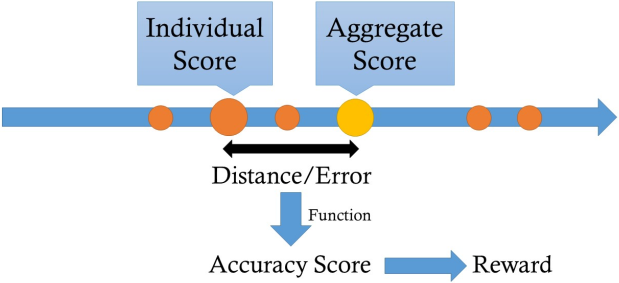
CrowdGrader lets students submit and collaboratively grade their solutions to homework assignments.

[2] https://www.crowdgrader.org/

Crowdgrader in a Nutshell

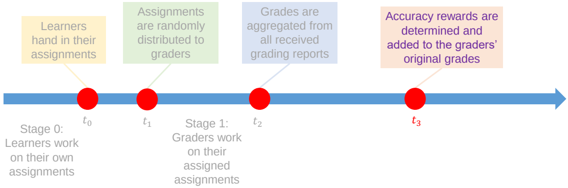


Crowdgrader in a Nutshell

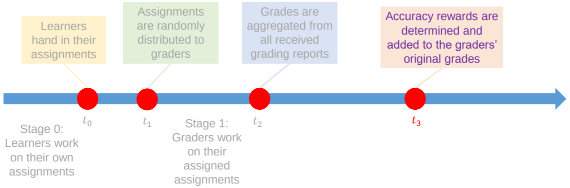


- Crowd-grading is found as effective and satisfactory (to the learners) as TA-grading [de Alfaro and Shavlovsky, 2014]

Revamped Peer Grading Mechanism

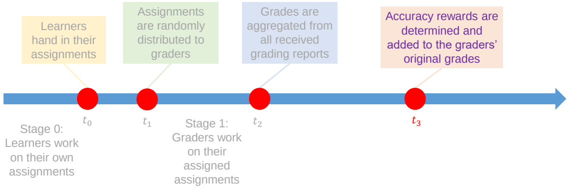


Revamped Peer Grading Mechanism



- Natural Questions:
  - What are the best ways to determine the accuracy (and consequent reward) to effectively incentivize efforts?

Revamped Peer Grading Mechanism



- Natural Questions:
  - What are the best ways to determine the accuracy (and consequent reward) to effectively incentivize efforts?
  - While the settings in Crowdgrader work empirically, are they theoretically robust as well?

Our Contribution

- We build and analyze a model of the Crowdgrading mechanism

Our Contribution

- We build and analyze a model of the Crowdgrading mechanism, and then explore the following:
  - Characterizing the rational behaviors of graders

Our Contribution

- We build and analyze a model of the Crowdgrading mechanism, and then explore the following:
  - Characterizing the rational behaviors of graders
  - Characterizing necessary conditions that such mechanism meet expectation

Our Contribution

- We build and analyze a model of the Crowdgrading mechanism, and then explore the following:
  - Characterizing the rational behaviors of graders
  - Characterizing necessary conditions that such mechanism meet expectation, which means
    - Effort is incentivized

Our Contribution

- We build and analyze a model of the Crowdgrading mechanism, and then explore the following:
  - Characterizing the rational behaviors of graders
  - Characterizing necessary conditions that such mechanism meet expectation, which means
    - Effort is incentivized, and
    - Course staff can gain direct control (of the outcome) by tuning mechanism parameters.

Our Contribution

- We build and analyze a model of the Crowdgrading mechanism, and then explore the following:
  - Characterizing the rational behaviors of graders
  - Characterizing necessary conditions that such mechanism meet expectation, which means
    - Effort is incentivized, and
    - Course staff can gain direct control (of the outcome) by tuning mechanism parameters.
  - Finding real settings that satisfy the necessary conditions

Outline

Motivation

Model

Players and Actions

Mechanism

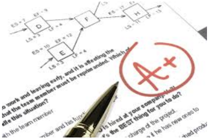
Results

Encouraging Conditions and Existence of Pure Equilibria

Controlling Pure Equilibria from Mechanism Parameters

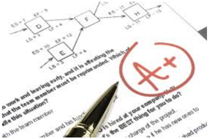
Players

- N learners, $\{a_1, a_2, ...a_N\}$



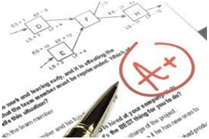
Players

- N learners, $\{a_1, a_2, ...a_N\}$
- We analyze one-shot
  (i.e., non-repetitive) scenario:
  1 cycle of peer grading



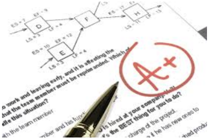
Players

- N learners, $\{a_1, a_2, ...a_N\}$
- We analyze one-shot
  (i.e., non-repetitive) scenario:
  1 cycle of peer grading
- Work of a_i has intrinsic value $v_i \in [0, M]$



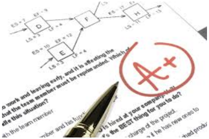
Players

- N learners, $\{a_1, a_2, ...a_N\}$
- We analyze one-shot (i.e., non-repetitive) scenario: 1 cycle of peer grading
- Work of a_i has intrinsic value $v_i \in [0, M]$
- Each learner grades exactly k submissions



Players

- N learners, $\{a_1, a_2, ...a_N\}$
- We analyze one-shot (i.e., non-repetitive) scenario: 1 cycle of peer grading
- Work of a_i has intrinsic value $v_i \in [0, M]$
- Each learner grades exactly k submissions
- Each submission is graded by exactly k learners



Actions

- Graders decide effort level (i.e., amount of time spent) on each submission

Actions

- Graders decide effort level (i.e., amount of time spent) on each submission
- Grader a_j's action is thus

$$T_j = [t_j^1, t_j^2, \ldots t_j^N], t_j^i \in [0, U]$$

(only k among these might be nonzero)

Actions

- Graders decide effort level (i.e., amount of time spent) on each submission
- Grader a_j's action is thus

$$T_j = [t_j^1, t_j^2, ...t_j^N], t_j^i \in [0, U]$$

(only k among these might be nonzero)

- Efforts are unobservable

Perceived Grade

- $S_j^i \sim \mathrm{f_g}(x, v_i, t_j)$: the grade a_j perceives w.r.t. a_i's work

Perceived Grade

- $S_j^i \sim \mathrm{f_g}(x, v_i, t_j)$: the grade a_j perceives w.r.t. a_i's work
  - depends on both effort level and intrinsic value

Perceived Grade

- $S_j^i \sim \mathrm{f_g}(x, v_i, t_j)$: the grade a_j perceives w.r.t. a_i's work
  - depends on both effort level and intrinsic value
  - we assume $\mathrm{f_g}$ to be unbiased, i.e., $\mathrm{f_g}(v_i + x, v_i, t_j) = \mathrm{f_g}(v_i - x, v_i, t_j)$, $\forall x \in \mathbb{R}, \forall t_j$

Perceived Grade

- $S_j^i \sim f_g(x, v_i, t_j)$: the grade a_j perceives w.r.t. a_i's work
    - depends on both effort level and intrinsic value
    - we assume f_g to be unbiased, i.e.,
      $f_g(v_i + x, v_i, t_j) = f_g(v_i - x, v_i, t_j), \forall x \in \mathbb{R}, \forall t_j$
- We assume truthfulness, i.e., a_j directly reports S_j^i to the mechanism without manipulation

Concensus Grade

- $\mathbf{S}_i = [S_j^i | G_{(i,j)} = 1]$: grades given to a_i's work

Concensus Grade

- $\mathbf{S}_i = [S_j^i | G_{(i,j)} = 1]$: grades given to a_i's work
- $\hat{S}_i = \mathrm{f}_{\mathrm{agg}}(\mathbf{S}_i)$: concensus grade

Concensus Grade

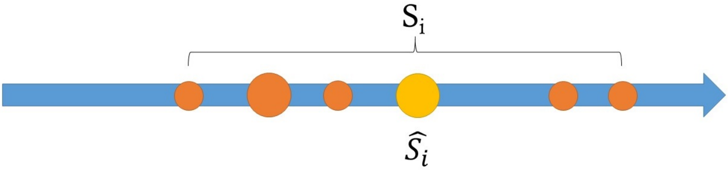
- $\mathbf{S}_i = [S_j^i | G_{(i,j)} = 1]$: grades given to a_i's work
- $\hat{S}_i = \mathrm{f}_{\mathrm{agg}}(\mathbf{S}_i)$: concensus grade
  - order-independent: $\mathrm{f}_{\mathrm{agg}}(\mathbf{S}_i) = \mathrm{f}_{\mathrm{agg}}(\mathbf{S}_i')$ if $\mathbf{S}_i'$ permutes $\mathbf{S}_i$

Concensus Grade

- $\mathbf{S}_i = [S^i_j | G_{(i,j)} = 1]$: grades given to a_i's work
- $\hat{S}_i = \mathrm{f_{agg}}(\mathbf{S}_i)$: concensus grade
  - order-independent: $\mathrm{f_{agg}}(\mathbf{S}_i) = \mathrm{f_{agg}}(\mathbf{S}'_i)$ if $\mathbf{S}'_i$ permutes $\mathbf{S}_i$
  - non-decreasing in every element in $\mathbf{S}_i$

Concensus Grade

- $\mathbf{S}_i = [S_j^i | G_{(i,j)} = 1]$: grades given to a_i's work
- $\hat{S}_i = f_{agg}(\mathbf{S}_i)$: concensus grade
  - order-independent: $f_{agg}(\mathbf{S}_i) = f_{agg}(\mathbf{S}_i')$ if $\mathbf{S}_i'$ permutes $\mathbf{S}_i$
  - non-decreasing in every element in $\mathbf{S}_i$
  - Examples: Average, Median, Olympian Average, etc.



Accuracy Level

- $\alpha_j^i = \mathsf{f}_{\mathsf{accu}}(|S_j^i - \hat{S}_i|) \in [0, 1]$: accuracy of a_j to a_i's work

Accuracy Level

- $\alpha_j^i = \mathrm{f}_{\mathrm{accu}}(|S_j^i - \hat{S}_i|) \in [0, 1]$: accuracy of a_j to a_i's work
  - depends on only the distance $|S_j^i - \hat{S}_i|$ between the perceived grade and the concensus grade

Accuracy Level

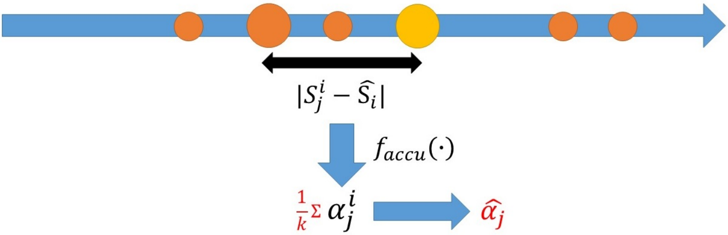
- $\alpha_j^i = f_{\text{accu}}(|S_j^i - \hat{S}_i|) \in [0, 1]$: accuracy of a_j to a_i's work
  - depends on only the distance $|S_j^i - \hat{S}_i|$ between the perceived grade and the concensus grade
  - non-increasing in $|S_j^i - \hat{S}_i|$

Accuracy Level

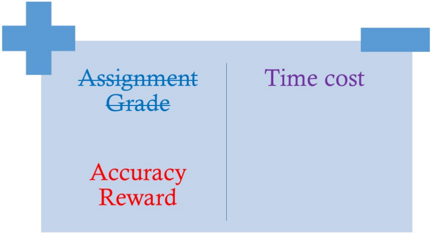
- $\alpha_j^i = \mathrm{f}_{\mathrm{accu}}(|S_j^i - \hat{S}_i|) \in [0,1]$: accuracy of a_j to a_i's work
  - depends on only the distance $|S_j^i - \hat{S}_i|$ between the perceived grade and the concensus grade
  - non-increasing in $|S_j^i - \hat{S}_i|$
- $\hat{\alpha}_j = \frac{1}{k}\sum_i(\alpha_j^i)$: a_j's average accuracy

Accuracy Level

- $\alpha_j^i = f_{accu}(|S_j^i - \hat{S}_i|) \in [0, 1]$: accuracy of a_j to a_i's work
  - depends on only the distance $|S_j^i - \hat{S}_i|$ between the perceived grade and the concensus grade
  - non-increasing in $|S_j^i - \hat{S}_i|$
- $\hat{\alpha}_j = \frac{1}{k}\sum_i(\alpha_j^i)$: a_j's average accuracy

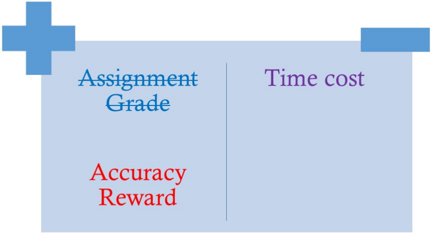


Utility



- $r_i \geq 0$: a_i's time-to-grade ratio

Utility



- $r_i \geq 0$: a_i's time-to-grade ratio
- $\pi_i = \cancel{(1-\lambda)M\hat{S}_i} + \lambda M\hat{\alpha}_i - r_i \sum_j t_i^j$.: final utility of learner i

Outline

Motivation

Model

Players and Actions

Mechanism

Results

Encouraging Conditions and Existence of Pure Equilibria

Controlling Pure Equilibria from Mechanism Parameters

Definition: EC-1

A setting satisfies **EC-1** if:
Given the others' 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 concave on $t_j \in [0, U]$.

More effort means more accuracy, with diminishing marginal increment.

Definition: EC-2

A setting satisfies **EC-2** if:
For any pair of two strategy profiles T_{-j} and T'_{-j} s.t.:

- $t_p < t'_p$ for some p,
- $t_q = t'_q \, \forall \, q \neq p$,

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

- One learner increasing effort increases everyone's marginal expected accuracy.

Definition: EC-2

A setting satisfies **EC-2** if:
For any pair of two strategy profiles T_{-j} and T'_{-j} s.t.:

- $t_p < t'_p$ for some p,
- $t_q = t'_q \, \forall \, q \neq p$,

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

- One learner increasing effort increases everyone's marginal expected accuracy.
- A positive reinforcement.

Existence of Pure NE

Theorem

In any UPG subgame that satisfies both EC-1 and EC-2, there exists at least one pure Nash equilibrium.

Existence of Pure NE

Theorem

In any UPG subgame that satisfies both EC-1 and EC-2, there exists at least one pure Nash equilibrium.

- We actually give an algorithm for the pure NE.

Existence of Pure NE

Theorem

In any UPG subgame that satisfies both EC-1 and EC-2, there exists at least one pure Nash equilibrium.

- We actually give an algorithm for the pure NE.
- We also find some real-world settings satisfying EC:

Existence of Pure NE

Theorem

In any UPG subgame that satisfies both EC-1 and EC-2, there exists at least one pure Nash equilibrium.

- We actually give an algorithm for the pure NE.
- We also find some real-world settings satisfying EC:

Proposition

Define $\mathrm{f}_{\mathrm{avg}}(\mathbf{S}) = \frac{1}{k}\sum_j (S_j)$. Suppose that $\mathrm{f_g} \sim N(v, g^2)$ where $g = g(t_j)$ is a non-increasing convex function. If $\mathrm{f}_{\mathrm{agg}}(\cdot) = \mathrm{f}_{\mathrm{avg}}(\cdot)$, $\mathrm{f}_{\mathrm{accu}}(\cdot)$ is non-increasing piecewise continuous, then for any values of $(M, k, \mathbf{r}, U, \lambda)$, $(M, k, \mathbf{r}, U, \lambda, \mathrm{f_g}(\cdot), \mathrm{f}_{\mathrm{agg}}(\cdot), \mathrm{f}_{\mathrm{accu}}(\cdot))$ satisfies both EC-1 and EC-2.

Effective time-to-grade ratio

- $\bar{r}_i = \frac{kr_i}{\lambda M} = \bar{r} r_i$, where $\bar{r} = \frac{k}{\lambda M}$

Effective time-to-grade ratio

- $\bar{r}_i = \frac{kr_i}{\lambda M} = \bar{r} r_i$, where $\bar{r} = \frac{k}{\lambda M}$:
- increases when k(# of assignments to grade) increases

Effective time-to-grade ratio

- $\bar{r}_i = \frac{kr_i}{\lambda M} = \bar{r}r_i$, where $\bar{r} = \frac{k}{\lambda M}$:
- increases when k(# of assignments to grade) increases
- decreases when λ(peer grading importance) increases

Effective time-to-grade ratio

- $\bar{r}_i = \frac{kr_i}{\lambda M} = \bar{r}r_i$, where $\bar{r} = \frac{k}{\lambda M}$:
- increases when k(# of assignments to grade) increases
- decreases when λ(peer grading importance) increases
- decreases when M(the full grade) increases

Effective time-to-grade ratio

- $\bar{r}_i = \frac{kr_i}{\lambda M} = \bar{r} r_i$, where $\bar{r} = \frac{k}{\lambda M}$:
- increases when k(# of assignments to grade) increases
- decreases when λ(peer grading importance) increases
- decreases when M(the full grade) increases
- Closely related to Nash equilibria

Encouraging Grading

Theorem

- Assume k is fixed, both EC satisfied in G_1 with $\bar{r} = \bar{r}_1$
- G_3 differs with G_1 only in $\bar{r}_3 < \bar{r}_1$
- $T_1 = [t_{1\,i}]$ is an equilibrium in G_1

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

- Decreasing $\bar{r}$ distorts the entire equilibria upwards!

Extensions

- We analyzed both special cases and general cases:

Extensions

- We analyzed both special cases and general cases:
  - If the learners are homogeneous
    (think of they are all of the same type),
    then every grader gives identical level of effort in pure NE.

Extensions

- We analyzed both special cases and general cases:
  - If the learners are homogeneous
    (think of they are all of the same type),
    then every grader gives identical level of effort in pure NE.
  - If some graders have fixed strategies (think of TA paddings),
    then the equilibria holds as long as
    irrational players' strategies are public information.

Conclusion

- We (in 2017) proposed the Encouraging Conditions (EC)

Conclusion

- We (in 2017) proposed the Encouraging Conditions (EC), which:
  - Guarantees existence of pure NE

Conclusion

- We (in 2017) proposed the Encouraging Conditions (EC), which:
  - Guarantees existence of pure NE
  - Describes an algorithm for computing the pure NE

Conclusion

- We (in 2017) proposed the Encouraging Conditions (EC), which:
  - Guarantees existence of pure NE
  - Describes an algorithm for computing the pure NE
  - Connects effort levels in NE with mechanism parameters

Conclusion

- We (in 2017) proposed the Encouraging Conditions (EC), which:
  - Guarantees existence of pure NE
  - Describes an algorithm for computing the pure NE
  - Connects effort levels in NE with mechanism parameters
- Proved that some practical settings meet EC

Conclusion

- We (in 2017) proposed the Encouraging Conditions (EC), which:
  - Guarantees existence of pure NE
  - Describes an algorithm for computing the pure NE
  - Connects effort levels in NE with mechanism parameters
- Proved that some practical settings meet EC
- Special case: homogeneous grading

Conclusion

- We (in 2017) proposed the Encouraging Conditions (EC), which:
  - Guarantees existence of pure NE
  - Describes an algorithm for computing the pure NE
  - Connects effort levels in NE with mechanism parameters
- Proved that some practical settings meet EC
- Special case: homogeneous grading
- Extension: irrational graders, TAs, biased grading, etc

Remark

- Let's review peer grading through Contract Theory

Remark

- Let's review peer grading through Contract Theory:
  - Course staff delegates the task of grading to the learners.

Remark

- Let's review peer grading through Contract Theory:
  - Course staff delegates the task of grading to the learners.
    - With anonymity and randomized distribution, the grading task is no different then a general task.

Remark

- Let's review peer grading through Contract Theory:
  - Course staff delegates the task of grading to the learners.
    - With anonymity and randomized distribution, the grading task is no different then a general task.
  - Adverse selection: learners' grading ability may vary

Remark

- Let's review peer grading through Contract Theory:
  - Course staff delegates the task of grading to the learners.
    - With anonymity and randomized distribution, the grading task is no different then a general task.
  - Adverse selection: learners' grading ability may vary
  - Moral hazard: learners' efforts are unobservable, but they cannot refuse to participate.

Remark

- Let's review peer grading through Contract Theory:
  - Course staff delegates the task of grading to the learners.
    - With anonymity and randomized distribution, the grading task is no different then a general task.
  - Adverse selection: learners' grading ability may vary
  - Moral hazard: learners' efforts are unobservable, but they cannot refuse to participate.
    - We have a continuous range of effort.

Remark

- Let's review peer grading through Contract Theory:
  - Course staff delegates the task of grading to the learners.
    - With anonymity and randomized distribution, the grading task is no different then a general task.
  - Adverse selection: learners' grading ability may vary
  - Moral hazard: learners' efforts are unobservable, but they cannot refuse to participate.
    - We have a continuous range of effort.
    - The output (grading outcome) is more complicated since grading is a task of estimation with two-sided error.

Remark

- Let's review peer grading through Contract Theory:
  - Course staff delegates the task of grading to the learners.
    - With anonymity and randomized distribution, the grading task is no different then a general task.
  - Adverse selection: learners' grading ability may vary
  - Moral hazard: learners' efforts are unobservable, but they cannot refuse to participate.
    - We have a continuous range of effort.
    - The output (grading outcome) is more complicated since grading is a task of estimation with two-sided error.
    - The ground truth (production) is also not perfectly observable.

Remark

- Let's review peer grading through Contract Theory:
  - Course staff delegates the task of grading to the learners.
    - With anonymity and randomized distribution, the grading task is no different then a general task.
  - Adverse selection: learners' grading ability may vary
  - Moral hazard: learners' efforts are unobservable, but they cannot refuse to participate.
    - We have a continuous range of effort.
    - The output (grading outcome) is more complicated since grading is a task of estimation with two-sided error.
    - The ground truth (production) is also not perfectly observable.
    - Therefore, aggregation significantly affects the peer grading game.

Pitfalls and limitations of this work

- Range of effort, intrinsic value, and range of grade are all bounded continuous, which makes the analysis hard and the model artificial

Pitfalls and limitations of this work

- Range of effort, intrinsic value, and range of grade are all bounded continuous, which makes the analysis hard and the model artificial
- Truthfulness is an assumption

Pitfalls and limitations of this work

- Range of effort, intrinsic value, and range of grade are all bounded continuous, which makes the analysis hard and the model artificial
- Truthfulness is an assumption
- Application is limited, while other similar scenarios exist: peer grading is actually very similar to other peer assessment scenarios, e.g., peer reviewing in academics, or crowdsourcing.


Q
&
A

Caragiannis, I., Krimpas, G. A., and Voudouris, A. A. (2015).
Aggregating partial rankings with applications to peer grading in massive online open courses.
In *AAMAS*, pages 675–683. ACM.

Carbonara, A. U., Datta, A., Sinha, A., and Zick, Y. (2015).
Incentivizing peer grading in MOOCS: an audit game approach.
In *IJCAI*, pages 497–503. AAAI Press.

de Alfaro, L. and Shavlovsky, M. (2014).
Crowdgrader: a tool for crowdsourcing the evaluation of homework assignments.
In *SIGCSE*, pages 415–420. ACM.

Lu, Y., Warren, J., Jermaine, C. M., Chaudhuri, S., and Rixner, S. (2015).

Grading the graders: Motivating peer graders in a MOOC.
In *WWW*, pages 680–690. ACM.

Mi, F. and Yeung, D. (2015).
Probabilistic graphical models for boosting cardinal and ordinal peer grading in moocs.
In *AAAI*, pages 454–460. AAAI Press.

Piech, C., Huang, J., Chen, Z., Do, C. B., Ng, A. Y., and Koller, D. (2013).
Tuned models of peer assessment in moocs.
In *EDM*, pages 153–160. International Educational Data Mining Society.

Raman, K. and Joachims, T. (2014).
Methods for ordinal peer grading.
In *KDD*, pages 1037–1046. ACM.