

A decorative network graph pattern in the top-left corner, featuring a complex web of interconnected nodes and edges. Some nodes are highlighted with blue circles, and others with blue dots. The pattern is composed of light gray lines and dots, with a few blue accents.

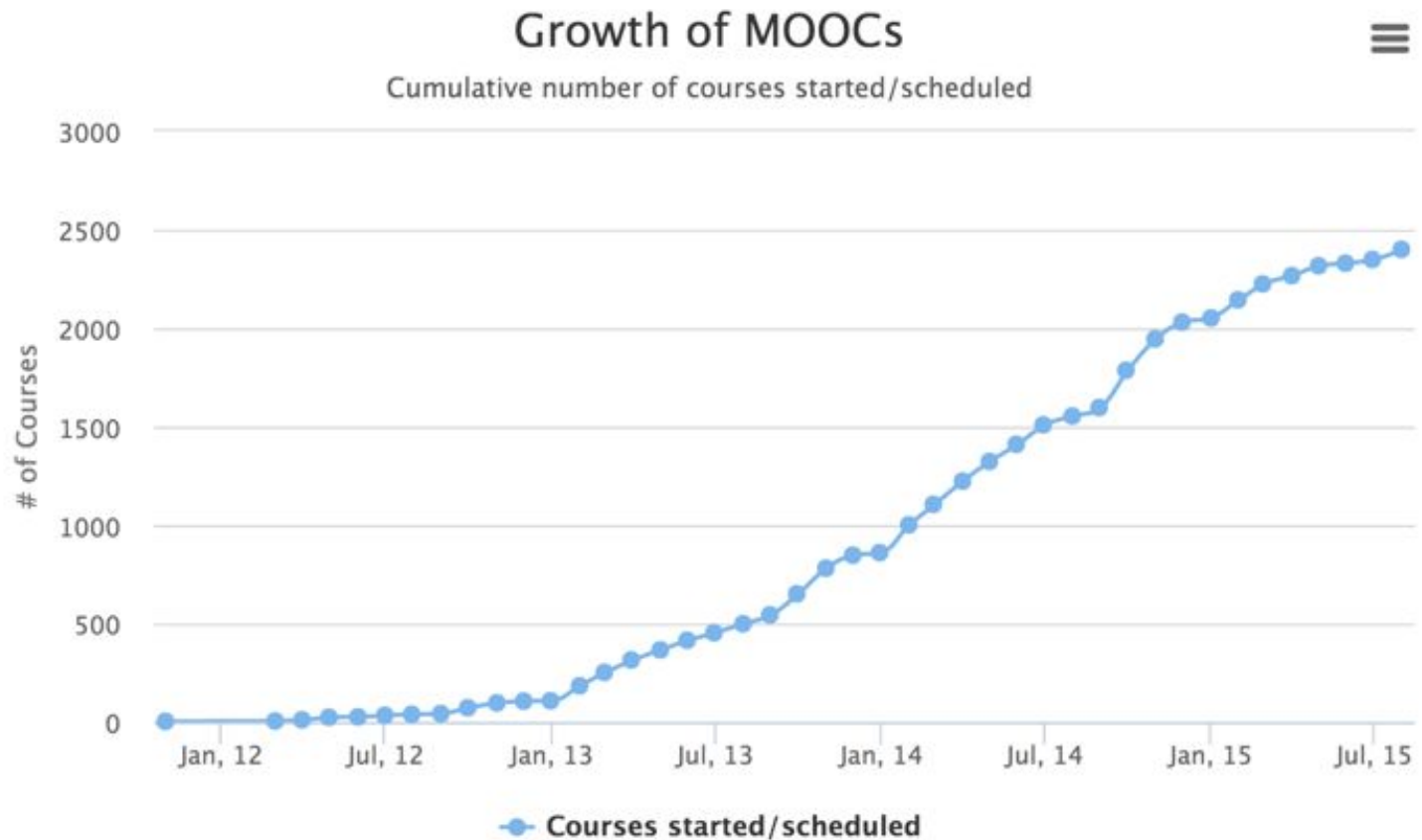
Learning at Scale

Weichen (Wilson) Liu

A decorative network graph pattern in the bottom-right corner, similar to the one in the top-left. It shows a network of nodes and edges, with some nodes highlighted by blue circles and others by blue dots. The pattern is made of light gray lines and dots, with blue highlights.

MOOC

Massive Open Online Course



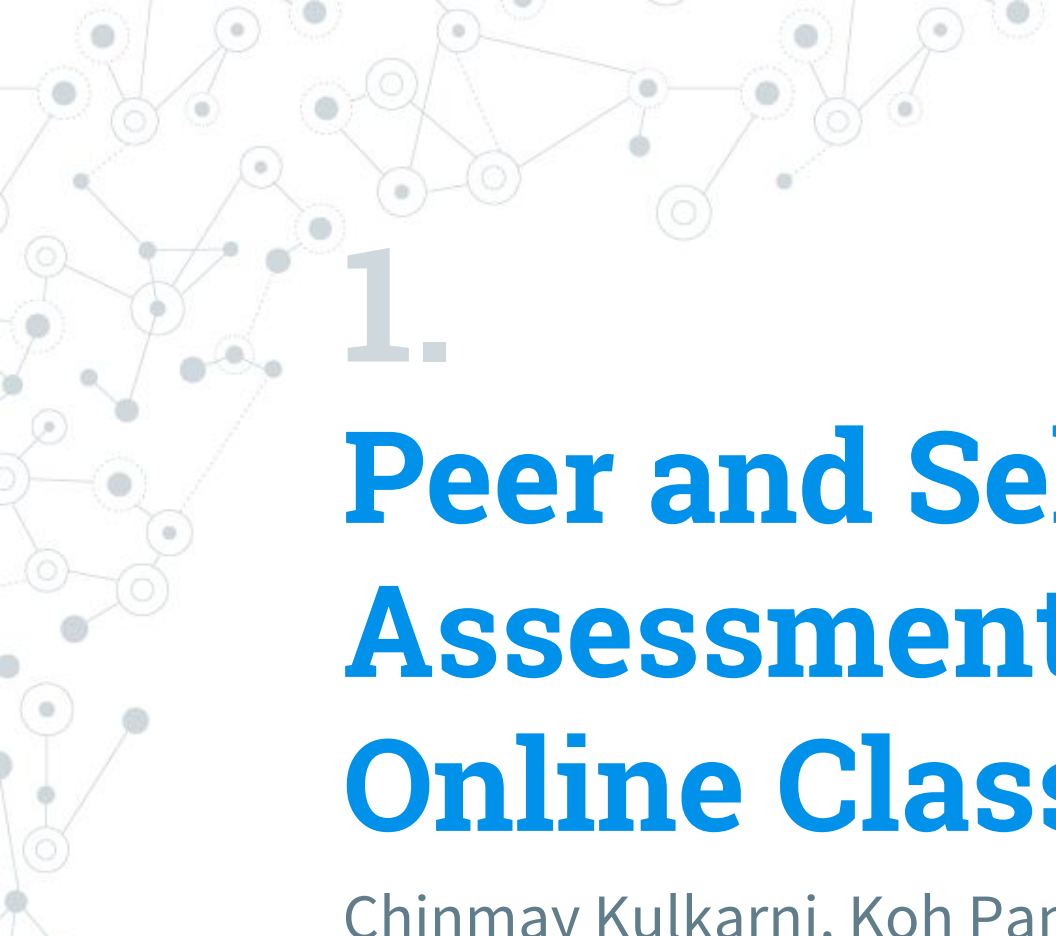


35,081

students who watched videos

2788

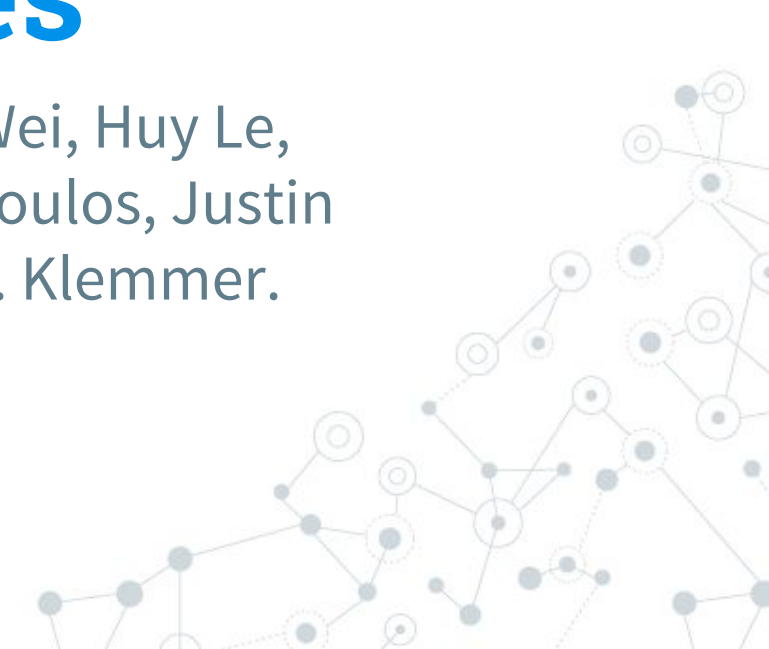
submissions of the first assignment



1.

Peer and Self Assessment in Massive Online Classes

Chinmay Kulkarni, Koh Pang Wei, Huy Le,
Daniel Chia, Kathryn Papadopoulos, Justin
Cheng, Daphne Koller, Scott R. Klemmer.
TOCHI 2013.



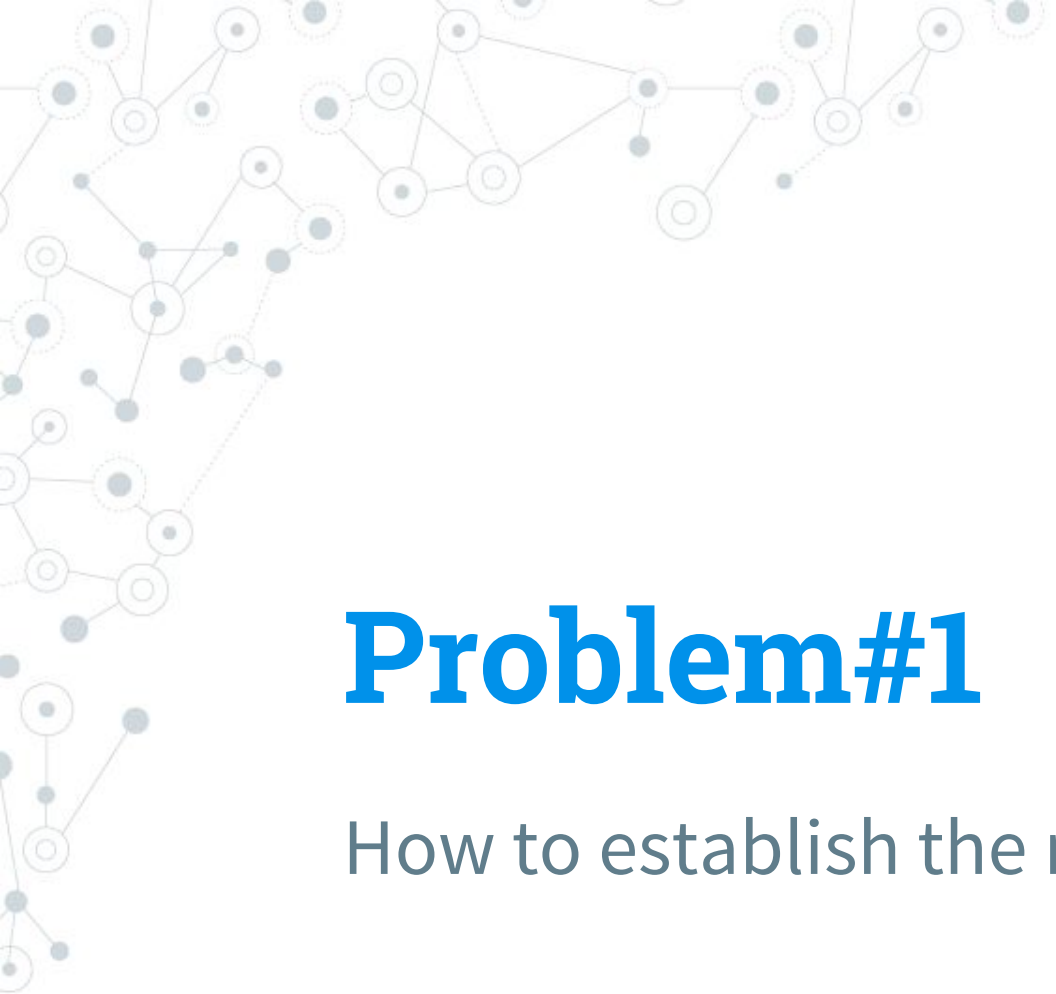
Learning Goals

- ◎ Understand peer and self assessment
- ◎ How the experiment was done
 - The accuracy Analysis
- ◎ Three approaches to improve accuracy

Peer Assessment

Viewing and critiquing other's work plays a key pedagogical role.



A decorative network diagram in the top-left corner, consisting of a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are solid dark grey, while others are hollow with a light grey outline. The lines connecting them are thin and light grey. The overall shape of the network is irregular and organic, resembling a molecular structure or a neural network.

Problem#1

How to establish the rule of grading?

A decorative network diagram in the bottom-right corner, similar in style to the one in the top-left. It features a cluster of interconnected nodes and lines. The nodes are small circles, some solid dark grey and some hollow with light grey outlines. The lines are thin and light grey, forming a complex, web-like structure.

Rubric

Guiding questions	Bare minimum	Satisfactory effort & performance	Above & Beyond
Alternate redesign—Extra credit. Have you created a fully functional alternate prototype?	0: No URL to functional prototype	3: URL present, but prototype only partially functional.	5: URL present, Alternative prototype is complete.
User testing. Photographs—extra credit. Did you submit photos from all three user testing sessions?	0: No photographs were uploaded.	3: Some photographs were uploaded (but less than 3), OR photos don't show an interesting moment in the experiment (e.g. photograph of participant signing consent form is not an interesting photo).	5: At least 3 photographs are uploaded and all photographs show interesting moments in the evaluation. Photos have meaningful captions

Category	Unsatisfactory	Bare minimum	Satisfactory effort & performance	Above & Beyond
Extra Credit: Electronic Prototype of Redesign	0: No URL to functional prototype	1: The prototype is incomplete and barely interactive.	3: The prototype is somewhat interactive, but not ready for user testing.	5: The alternative prototype is fully interactive and ready for user testing.
Photos/Sketches	0: No photographs were submitted that showed interesting moments in the user testing process.	1: 1 photograph was submitted that showed an interesting moment in the user testing process.	3: 2 photographs were submitted that showed interesting moments in the user testing process..	5: 3 or more photographs were submitted that showed interesting moments in the user testing process.

A decorative network diagram in the top-left corner, consisting of various sized circles (nodes) connected by thin lines (edges). Some nodes are solid grey, while others are hollow with a grey outline. The network is dense and irregular.

Problem#2

How to design the grading process?

A decorative network diagram in the bottom-right corner, similar to the one in the top-left, featuring a cluster of interconnected nodes and edges.

Process

Staff evaluated
12 assignments



Student:
Calibrated Peer
Assessment

Calibrated Peer assessment



Discussion (2 min, group of 2-3)

- In what ways are peer and self assessment useful respectively?
- What's the point of putting self-assess after peer-assess?

Peer assessment

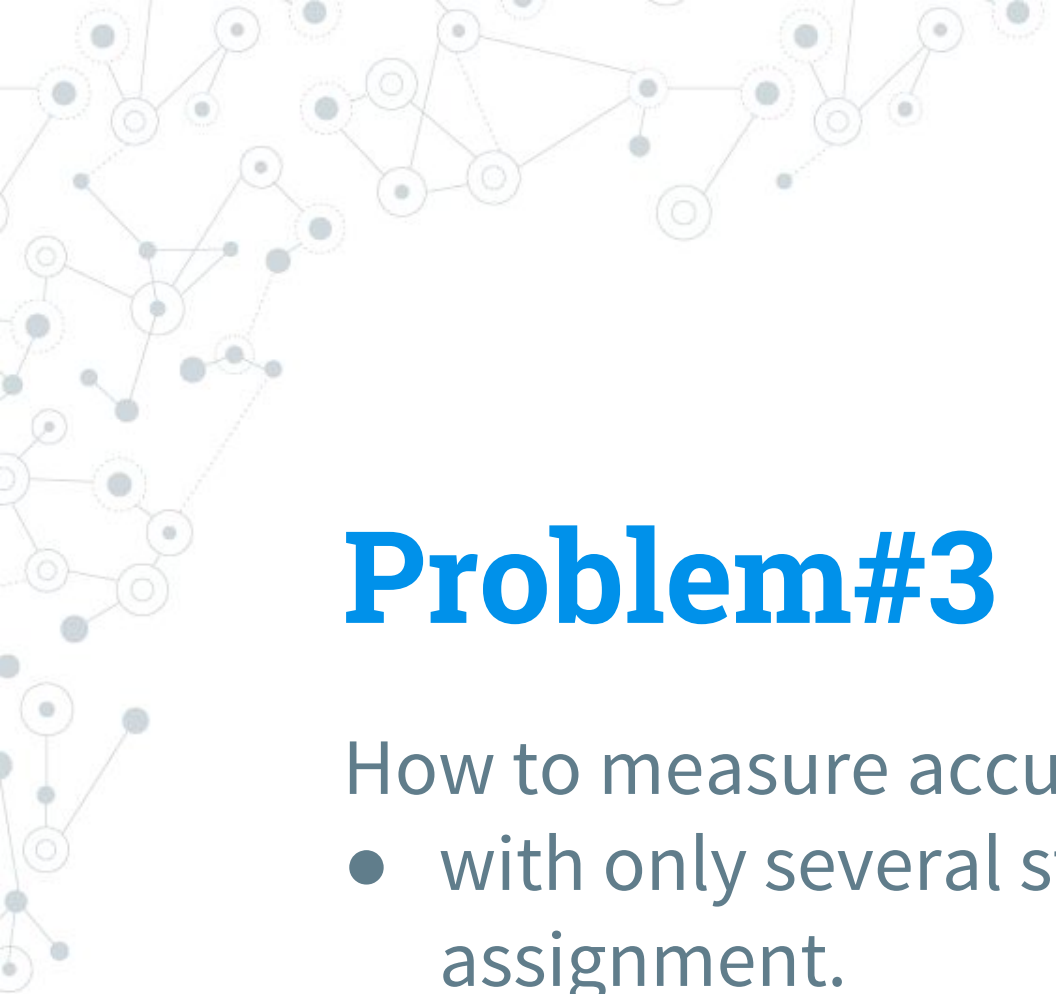


Calibrated

Peer assessment



**How the score
of the assignment
is calculated?**

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The overall structure is organic and sprawling.

Problem#3

How to measure accuracy?

- with only several staff-graded assignment.



Accuracy

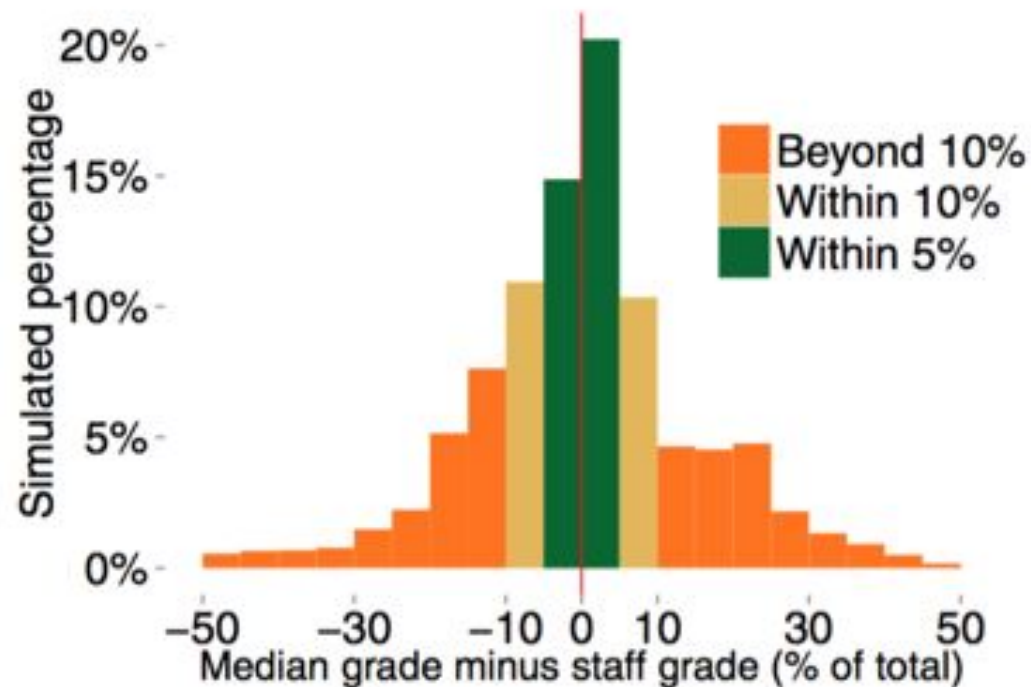
- Method
 - Ground truth submissions
 - Median-grade approach

Peer assessment



Accuracy

◎ Result



(a) Iteration 1: 34.0% of samples within 5% of the staff grade, and 56.9% within 10%.



Problem#4

Can we do better?





Improve Accuracy & Provide Qualitative Feedback

- ◎ Providing Feedback
 - ◎ Fortune Cookies
 - ◎ Data-driven Rubric Revisions
- 

Feedback

- ◎ About 800 participants
- ◎ Two conditions
 - no-feedback control
 - feedback



Feedback



You graded your peers' work **a little low** on Assignment 4. The grading rubrics are useful if you're unsure about what scores you should assign.

[What's this?](#)[Leave Feedback](#)

You graded your peers' work **a little high** on Assignment 4. The grading rubrics are useful if you're unsure about what scores you should assign.

[What's this?](#)[Leave Feedback](#)

You graded your peers' work **accurately** on Assignment 4! Keep it up!

[What's this?](#)[Leave Feedback](#)

Feedback

Word ONLINE
Human-Computer Interaction

Associate Profes

[Peer Assessments](#) / HCI Assignment 4 - Ready for Testing



You graded your peers' work accurately on Assignment 3! Keep it up!

[What's this?](#)

[Leave Feedback](#)

1. Do assignment ✓

2. Learn to evaluate

3. Evaluate your classmates ⚙

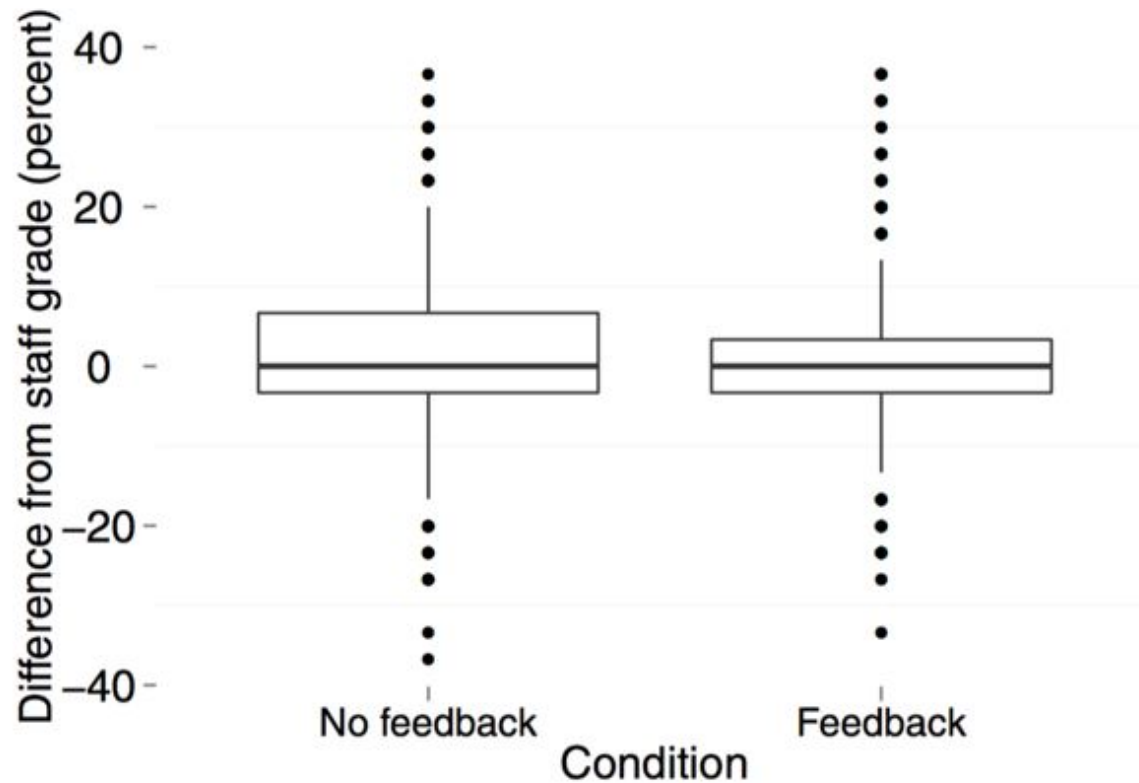
4. Evaluate yourself ⚙

5. See results ⚙

[← Return to list](#)

[Save draft](#)

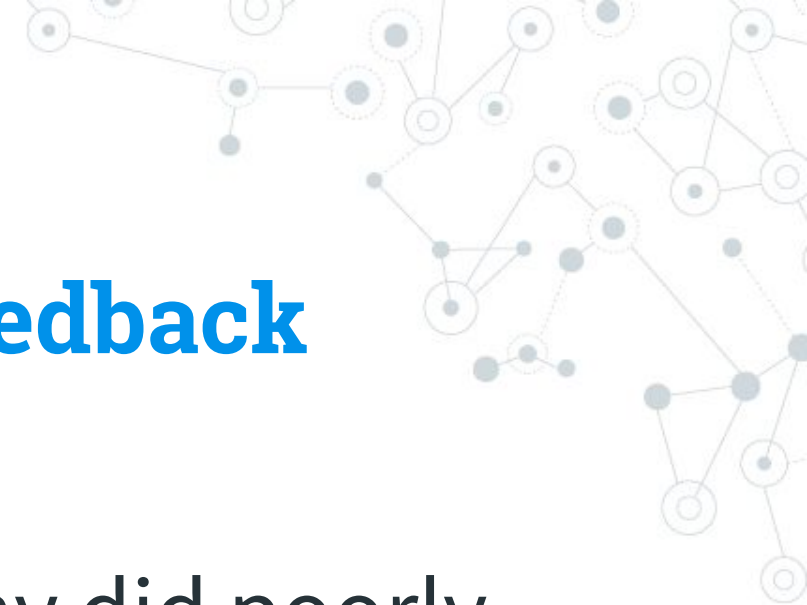
Feedback



Provide Qualitative Feedback



Peer ← Grader



Provide Qualitative Feedback

- ◎ Rubric Limitations
 - not clear exactly why did poorly on some topic
 - lack pointing out how to improve
- 

Fortune Cookie



... because _____

Fortune Cookie

Overall evaluation/feedback

Note: this section can only be filled out during the evaluation phase.

Overall feedback:

How could this student best improve his/her submission? From among the following, copy one or more pieces of advice that would help the student. Paste your advice in the feedback box below.

- Clarify the concerns, goals, and expectations of the user tests.
- Make the user tests more structured.
- ~~Make the user tests more consistent across participants.~~
- Make the prototype more interactive so the user test represents a more real-life interaction.
- ~~Determine the implications of the user succeeding (or not) on each task on the prototype.~~
- Make fewer assumptions about users/Reduce bias in user test.
- Other

Copy, then paste

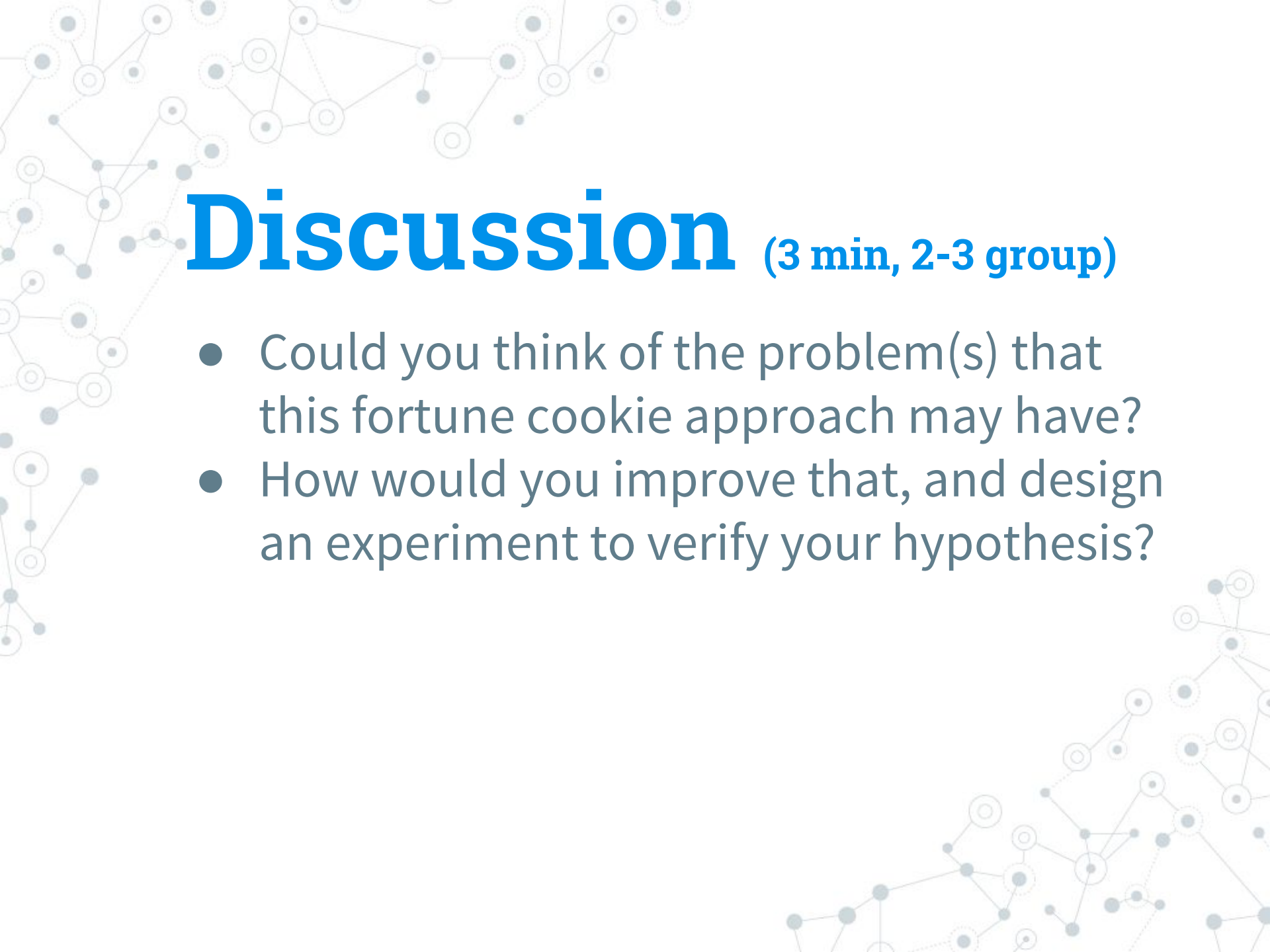
Make the prototype more interactive so the user test represents a more real-life interaction: The prototype does everything you're testing, but it couldn't hurt to make it more interactive. If the user can't possibly stray from the things you want to test, how do you know that the user can actually use the full application without making mistakes?

Fortune Cookie

- ② $\frac{2}{3}$ contained fortune cookie
- ② Do not encourage more students to leave feedback (36.2% v.s. 36.4%)

Fortune Cookie

- ◎ However
 - Reduce feedback cost
 - Encourage brainstorming



Discussion (3 min, 2-3 group)

- Could you think of the problem(s) that this fortune cookie approach may have?
- How would you improve that, and design an experiment to verify your hypothesis?

A network diagram with nodes and edges, overlaid with a grid of green and orange squares and a red vertical line. The network consists of various sized circles (nodes) connected by thin lines (edges). Some nodes are solid grey, while others are hollow with a grey outline. The grid is composed of small squares, some green and some orange, arranged in a pattern that suggests a data matrix or a spatial layout. A single red vertical line runs through the center of the grid.

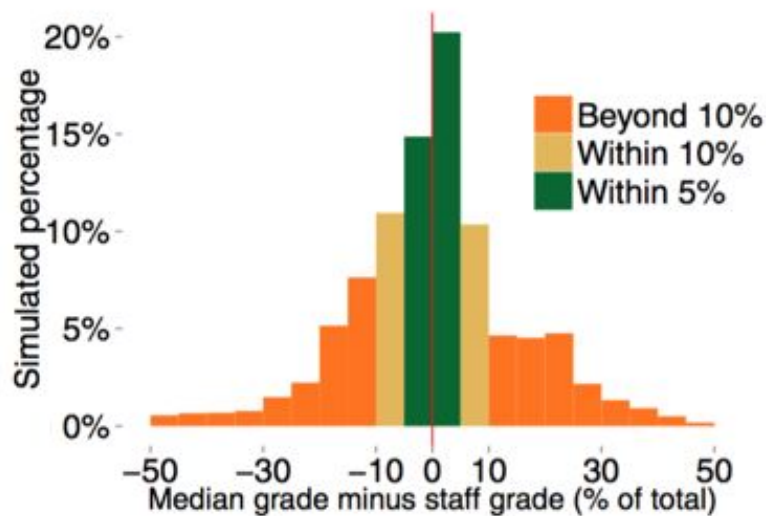


Data-driven Rubric Revisions

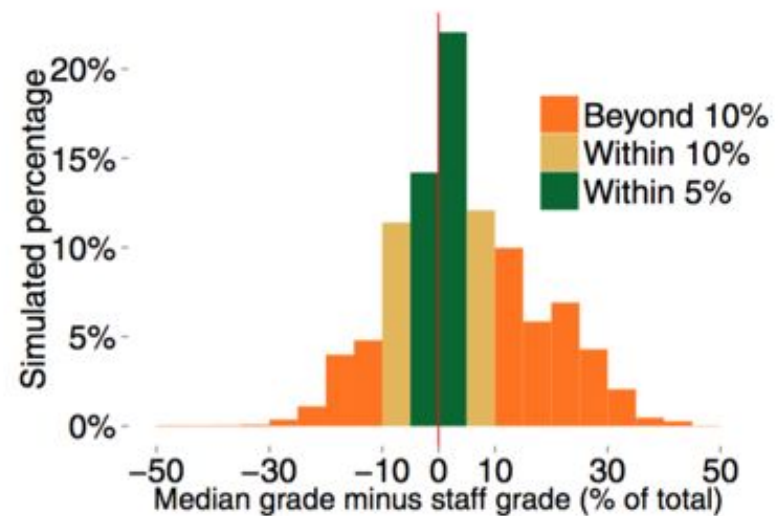
- ◎ Parallel sentence structure
- ◎ Splitting up complex rubric items
- ◎ Using less ambiguous words

Accuracy

Result



(a) Iteration 1: 34.0% of samples within 5% of the staff grade, and 56.9% within 10%.



(b) Iteration 2: 42.0% of samples within 5% of the staff grade, and 65% within 10%.

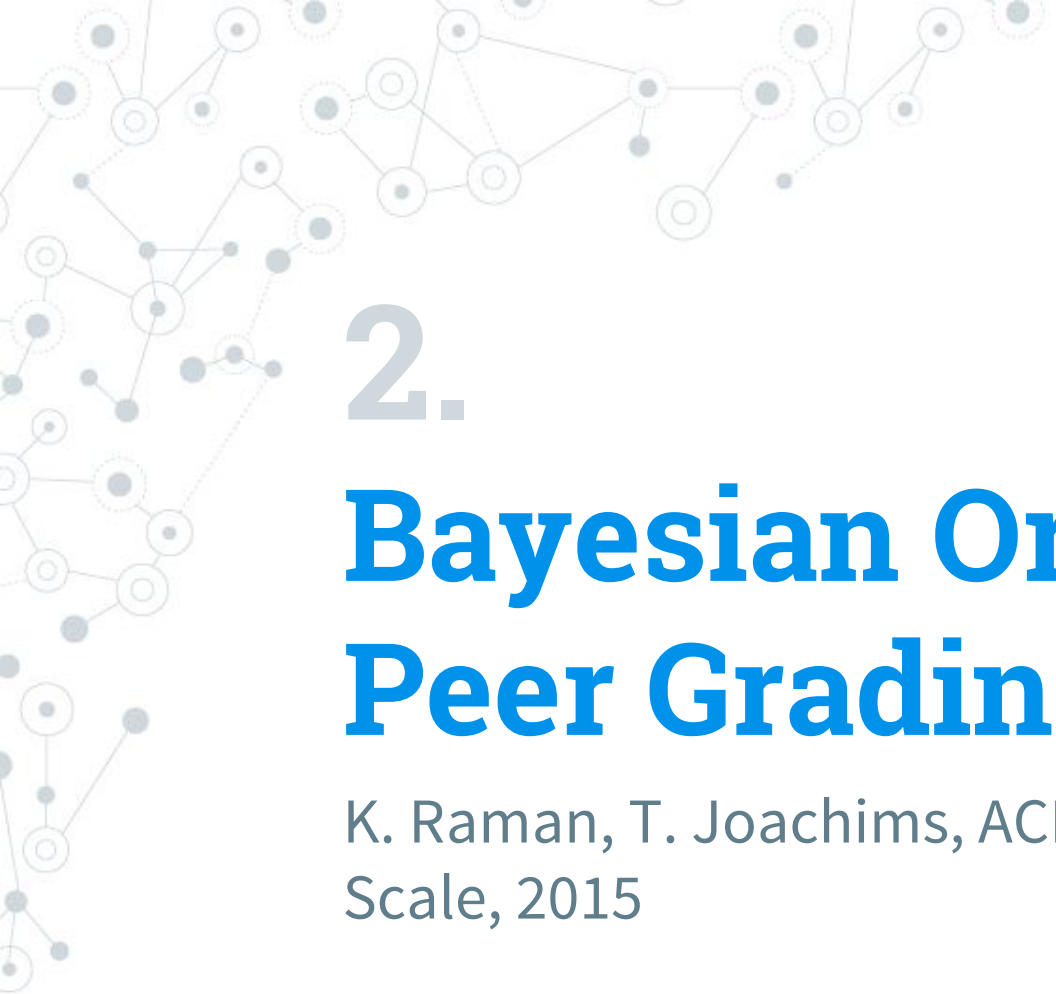
Students Reaction

- ◎ Giving feedback & self assessment are valuable learning
- ◎ 20% students voluntarily did more than required assessments



Discussion (2 min, 2-3 group)

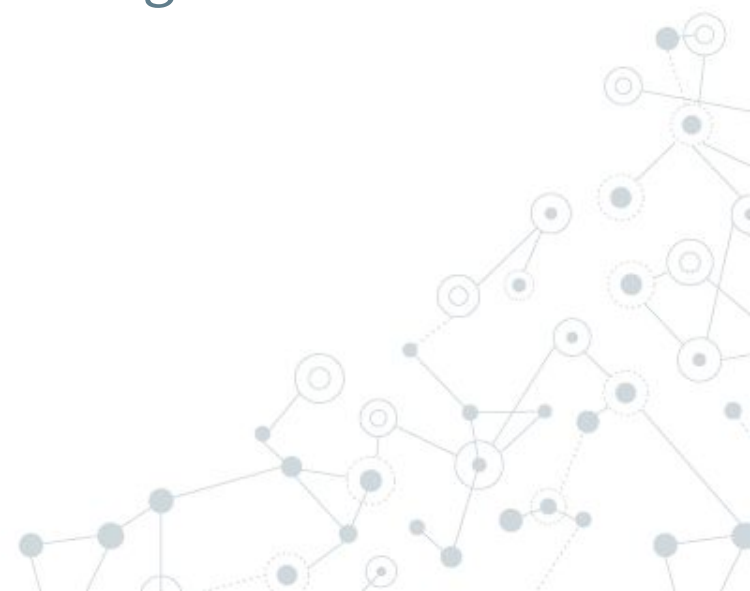
- What other domains or fields could the calibrated peer assessment technique apply to?



2.

Bayesian Ordinal Peer Grading

K. Raman, T. Joachims, ACM Learning at
Scale, 2015

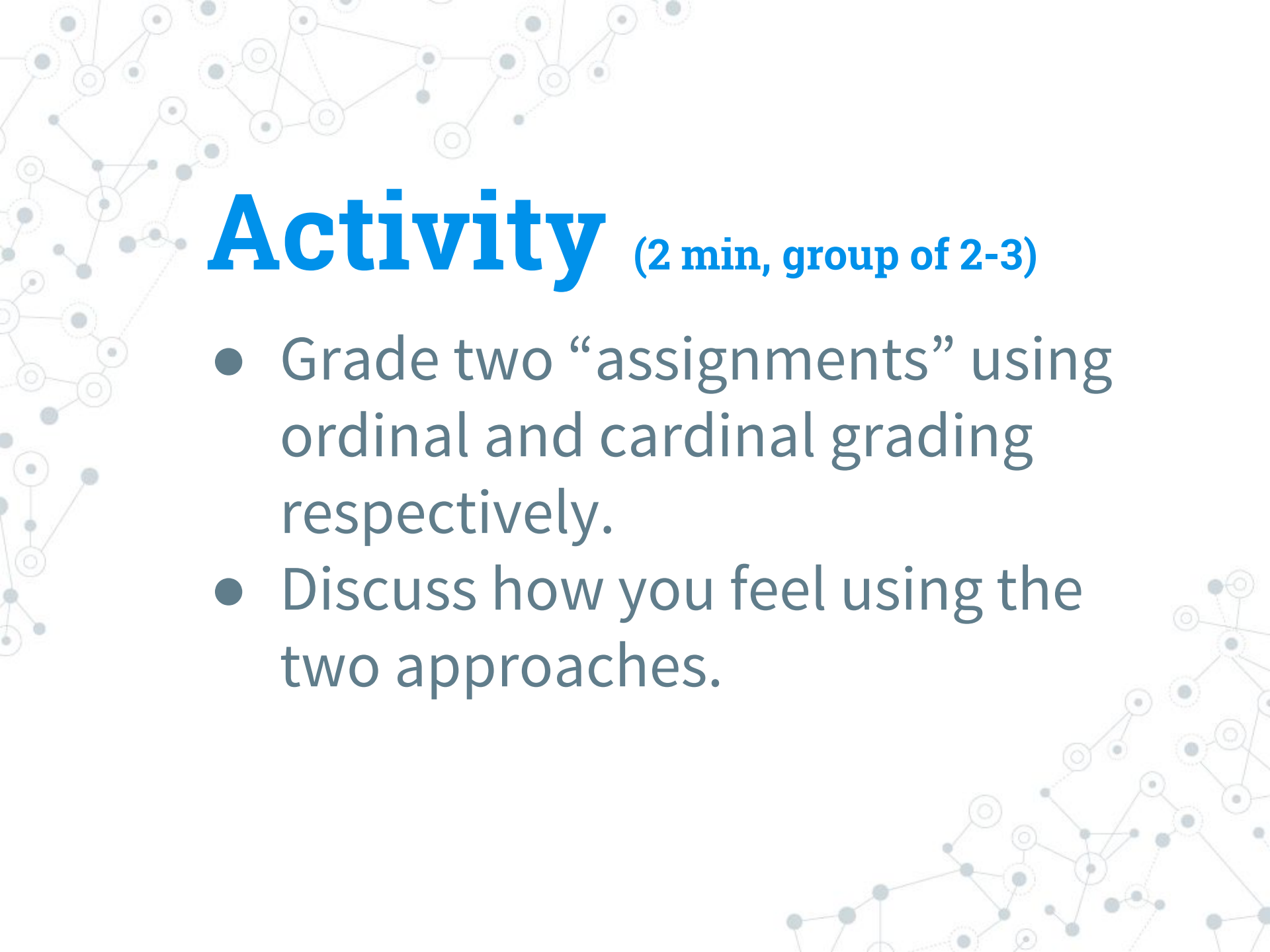


Learning Goals

- ◎ Understand ordinal and cardinal grading
- ◎ Know the uncertainty problem

Ordinal & Cardinal

- ◎ Ordinal Grading
 - Project X is better than project Y
- ◎ Cardinal Grading
 - Project X is a B-



Activity

(2 min, group of 2-3)

- Grade two “assignments” using ordinal and cardinal grading respectively.
- Discuss how you feel using the two approaches.

Ordinal v.s. Cardinal

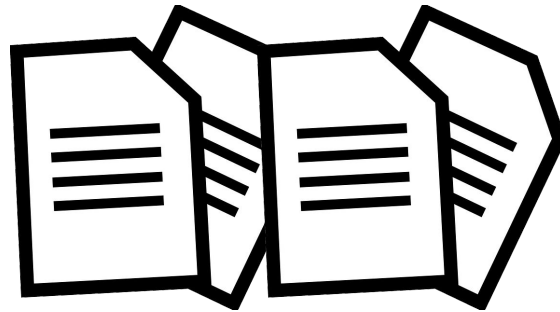
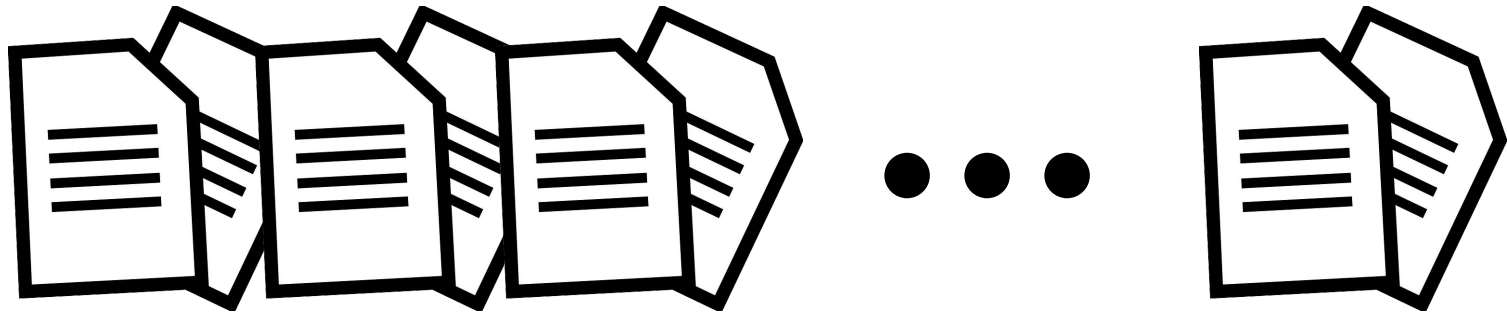
◎ Ordinal

- Easier
- More reliable

◎ Cardinal

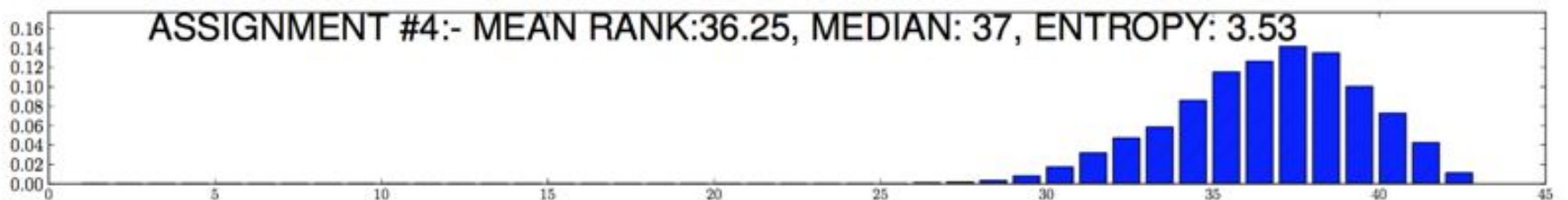
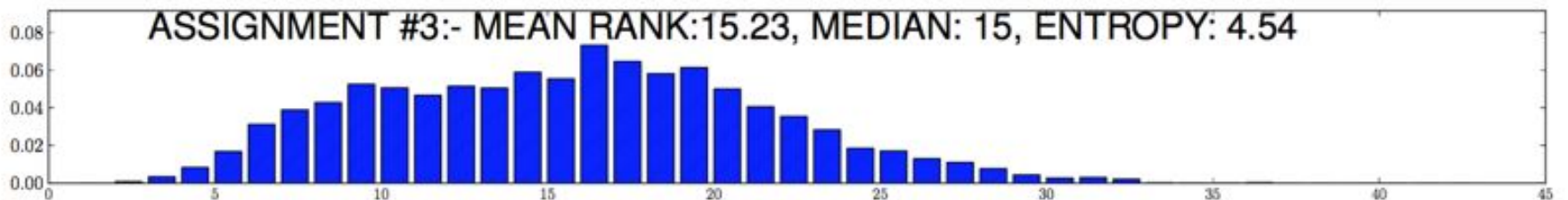
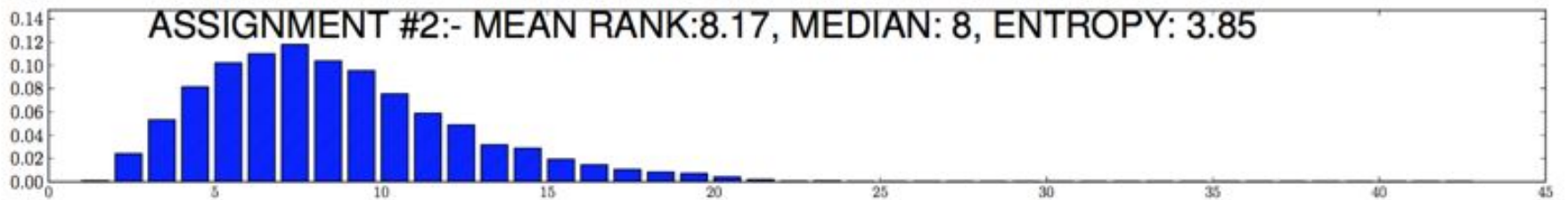
- different scale
- non-linear

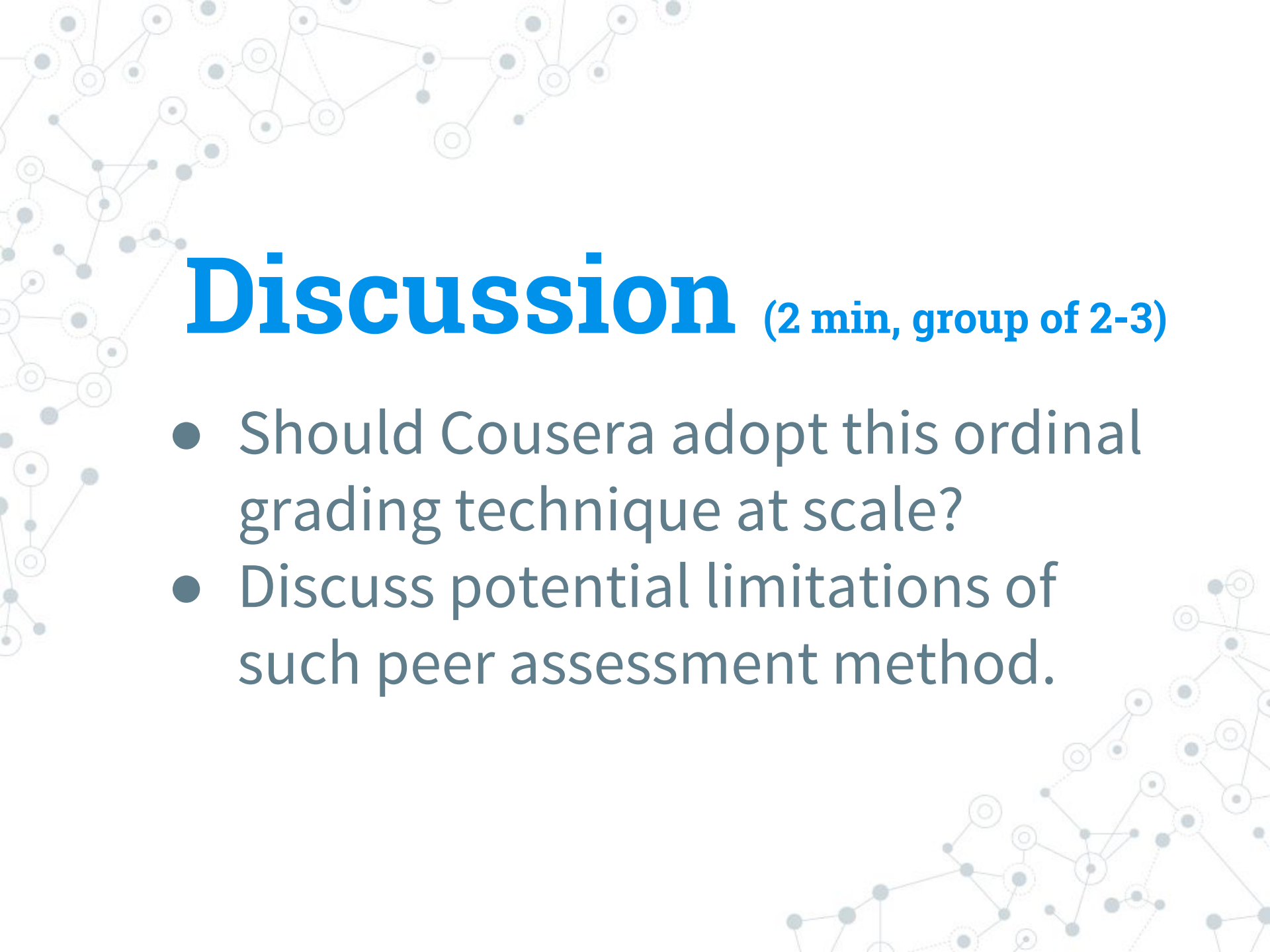
Ordinal Peer Grading (OPG)



Rank Aggregation

- ◎ Limitations for OPG problem
 - top items v.s. full ranking
 - a single ordering may not suffice
- ◎ This paper: Uncertainty





Discussion (2 min, group of 2-3)

- Should Coursera adopt this ordinal grading technique at scale?
- Discuss potential limitations of such peer assessment method.



Thanks!