Student Attitudes towards Partially Automated Peer Grading in Mechanical TA

Jessica Q. Dawson
James R. Wright
Kevin Leyton-Brown
BACKGROUND

**CPSC 430 – Computers and Society**

- 4\textsuperscript{th} year undergraduate course
  - focus on critical reasoning about social implications of computational advances

- focus on frequent short, frequent writing assignments
  - effective way to teach writing skills [Seabrook et al 2005]
  - provides many opportunities to practice

- students complete a 300-word essay every week
  - Problem: inefficient and expensive for manual TA marking
  - Solution: peer grading
BACKGROUND

Peer Grading and Mechanical TA

• in peer grading: students grade each others’ assignments

• peer grading often **negatively perceived** by students
  – tend to believe **lower quality/less fair** than TA grading
    [Kaufman & Schunn 2011]

• a solution: **Mechanical TA** (see companion poster for details.)
  – software system for partially automated peer grading,
    developed by CPSC 430 course staff
  – TAs remain in the loop:
    • mark essays/reviews before students graduate to ‘independent’
    • for ‘independent’ students: manage appeals and spot-checks
  – results over 3 offerings found evidence that MTA helped
    improve student learning and grading ability
    [Wright et al. 2015]
Measuring Perceptions of MTA

Research Questions

• what are the students’ perceptions of the peer grading?
  – how do students perceive the **quality, appropriateness, fairness helpfulness and accuracy** of:
    1. reviews they gave their peers on their writing,
    2. reviews they received from peers
      …and how did these compare to TA reviews?
  – how **helpful** was the calibration (built in practice reviewing) and peer grading and in MTA for learning?

Data Collection

• End-of-term survey conducted in CPSC 430 (2014 W1)
• n = 76 (response rate 83%)
The majority of students rated the reviews they wrote favorably with respect to each factor.
The majority of students the reviews they wrote were *about the same* as how a TA would have reviewed the same paper.
The students’ perceptions of the reviews written by peers were more *mixed*. 

Q: The reviews my peers gave me on my writing . . .

- **were of high quality.**
  - Strongly disagree: 11%
  - Disagree: 28%
  - Neither agree nor disagree: 38%
  - Agree: 16%
  - Strongly Agree: 7%

- **were appropriate in tone.**
  - Strongly disagree: 18%
  - Disagree: 36%
  - Neither agree nor disagree: 36%
  - Agree: 36%
  - Strongly Agree: 8%

- **were fair.**
  - Strongly disagree: 7%
  - Disagree: 24%
  - Neither agree nor disagree: 43%
  - Agree: 18%
  - Strongly Agree: 8%

- **were helpful when I wrote future essays**
  - Strongly disagree: 12%
  - Disagree: 28%
  - Neither agree nor disagree: 26%
  - Agree: 28%
  - Strongly Agree: 5%

- **were accurate with respect to the grading**
  - Strongly disagree: 7%
  - Disagree: 31%
  - Neither agree nor disagree: 35%
  - Agree: 22%
  - Strongly Agree: 5%
**RESULTS**

**Reviews received, compared to TAs**

Q: Compared to a TA review, the __________ of the reviews I received were

For most factors, majority felt their peers’ reviews were **worse** than how a TA would have graded the same paper.
Perceptions of helpfulness were *mixed* for activities asked about. For learning content and concepts, a small majority found peer review somewhat or very helpful (57%).
**Results**

**Helpfulness of calibration for learning**

Q. How helpful was calibration [built in practice reviews] for . . . .

- Improving the quality of your essays:
  - Very unhelpful: 5%
  - Somewhat unhelpful: 32%
  - Neither helpful or unhelpful: 48%
  - Somewhat helpful: 12%

- Improving the quality of your reviews of peers' essays:
  - Very unhelpful: 5%
  - Somewhat unhelpful: 25%
  - Neither helpful or unhelpful: 48%
  - Somewhat helpful: 19%

- Becoming an independent reviewer more quickly:
  - Very unhelpful: 4%
  - Somewhat unhelpful: 16%
  - Neither helpful or unhelpful: 43%
  - Somewhat helpful: 35%

- Obtaining a lot of practice reviewing:
  - Very unhelpful: 5%
  - Somewhat unhelpful: 18%
  - Neither helpful or unhelpful: 55%
  - Somewhat helpful: 20%

 Majority of students found calibration *helpful* for all activities asked about – considered most helpful for activities specifically tied to reviewing.
Conclusion

and possible next steps

• Students feel positively about their own reviewing ability, perceived to be similar to TAs.
  – calibration helpful as expected for learning how to review.

• But many still doubt peers’ abilities.
  – even though course staff also satisfied with reviewing ability.
  ➔ how can we bridge this gap?

• Perceived helpfulness of peer reviewing for learning and improving writing was mixed.
  – could adjust types of feedback reviewers expected to provide e.g., more qualitative and focused on writing skills.
References

