Aston Business School

BIRMINGHAM UK

Assessing the impact of peer-evaluation on assessed group work

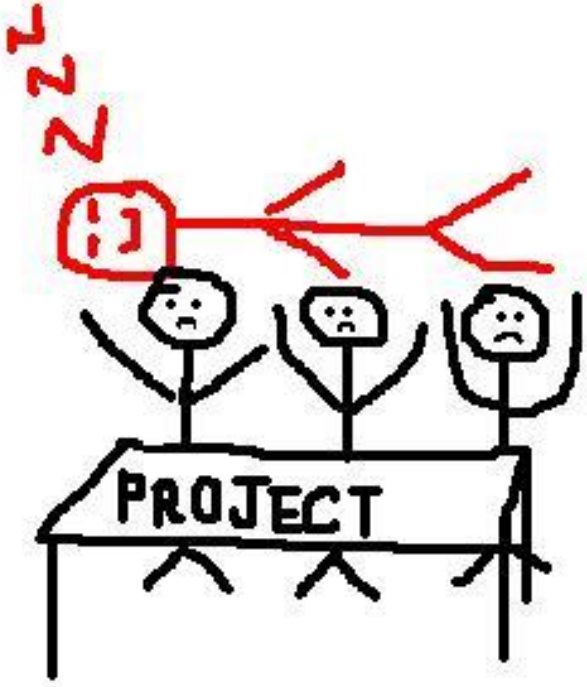
Jon Guest and Robert Riegler

DEE Conference, Heriot-Watt University (05/09/23)

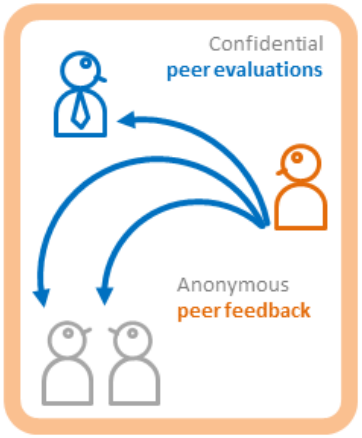
Background

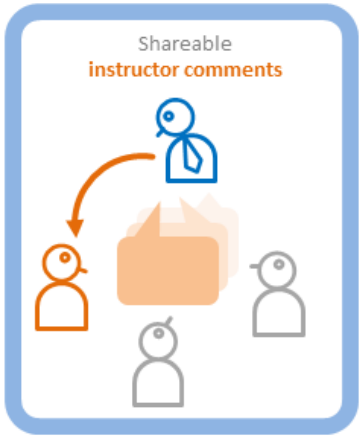
- Large level 4 module – approx. 200 students
- Assessment – Consultancy/presentation
- Conjunctive/additive/complex assessment task suitable for group work
- Randomly assigned groups
- Usual problems!!!

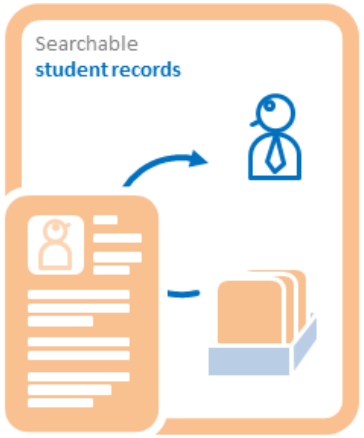
Assessed Group Work - Issues



Peer evaluation







Fixed point system

- Peer evaluation method: Allocate 100 effort points between your peers
- We then create a contribution index

$$CI_i = \frac{Average\ peer\ evaluation\ score\ received\ by\ student\ i}{100/(N-1)}$$

- which feeds into the Individual marks (*IM*). These were then derived from the overall group mark (*GM)* in the following manner:

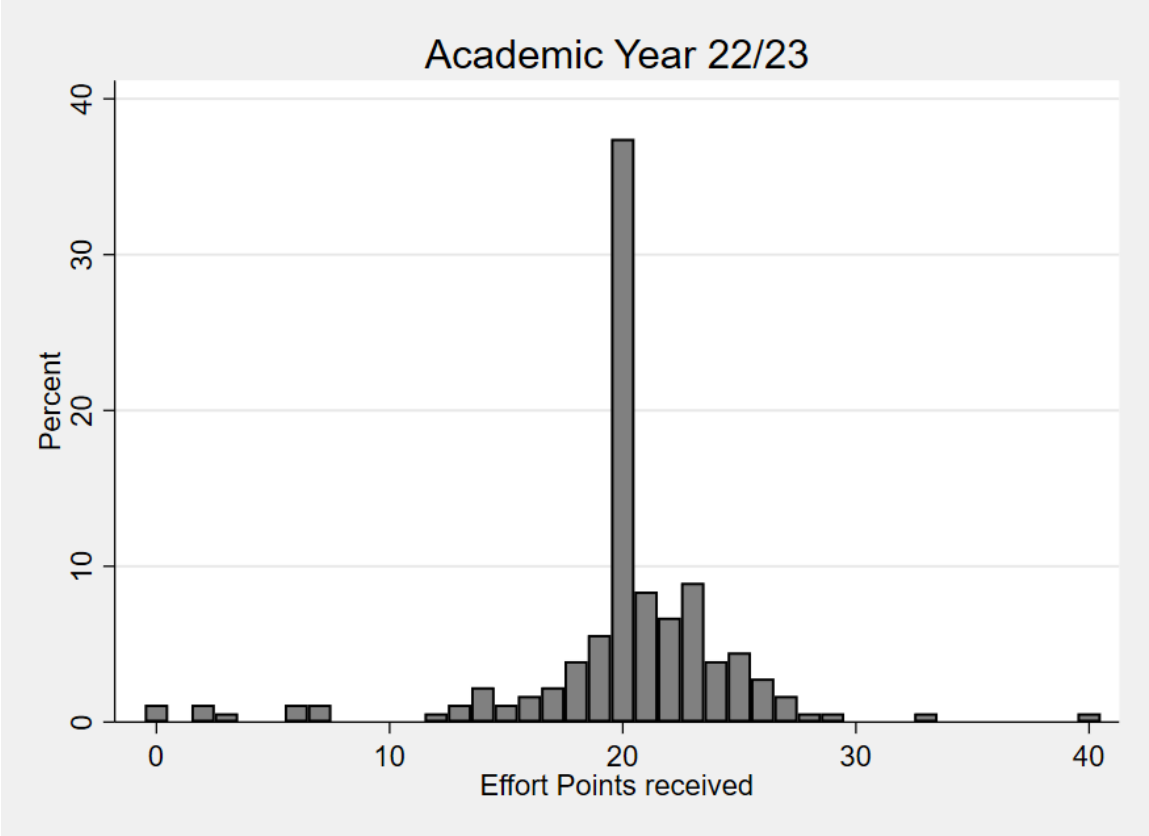
$$IM = \alpha \times GM + (1 - \alpha) \times GM \times CI$$

Descriptive Statistics

	22/23
Number of Students	179
Number of Groups	31
Av. Size	5.8
Groups with equal distr. of effort points	7 (22.6%)
Non-submission of PE	4 (2.2%)

Notes: Excl. drop-out students

Distribution of average peer scores received



Students' perception

"In a group of 5 people and you have 2 people who do absolutely no work, but they're best friends or they made friends during the module, they have the possibility of giving each other full marks and giving people who have actually done the work 0."

Collusion between team members? Example

Rater/Ratee	Karishma	Jon	Matt	Adam	Robert
Karishma		25	25	25	25
Jon	22		22	22	34
Matt	25	25		25	25
Adam	25	25	20		30
Robert	22	34	22	22	

Biases in peer-evaluations

- Common concern about impact of relationship between students within the same group on peer-evaluations (e.g. Brindley & Scoffield, 1998; Helmore & Magin, 1998)
- There is only few empirical evidence about the prevalence and significance of this bias:
  - Montgomery (1986) did not find evidence of reciprocity effects
  - Magin (2010) found that only 1% of the variance in peer scores are explained by reciprocal effects.

Mutual High Scoring

- 4 potential explanations for Mutual High Scoring (MHS)

  1. **Collusion/gaming**
  2. **Less asymmetric information**
  3. **Cognitive bias**
  4. **Actual/accurate scoring that reflect contributions**

- We attempt to disentangle 1-3 from 4

Methods: 2-step approach

- First Step: **Identifying MHS**

1. We calculate the deviation from the mean for each score student *i* gave to group member *j* and vice versa

$$dX_{ij} = X_{ij} - \bar{X} \text{ and } dX_{ji} = X_{ji} - \bar{X}$$

   – Note: Due to fixed point system, $\bar{X} = 20$ for a group of 6, $\bar{X} = 25$ for a group of 5, etc

2. We calculate the product to derive our MHS measure at pair level *p*:

$$MHS_p = dX_{ij} \times dX_{ji}$$

Four types of MHS

Aston Business School
BIRMINGHAM UK

TRIPLE ACCREDITED

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION AMBA ACCREDITED

1. No MHS

a) $MHS_p = 0$, whereby $dX_{ij} = 0$ or $dX_{ji} = 0$, or $dX_{ij} = dX_{ji} = 0$

b) $MHS_p < 0$, whereby $dX_{ij} > 0$ and $dX_{ji} < 0$, or $dX_{ij} < 0$ and $dX_{ji} > 0$

c) $MHS_p > 0$, whereby $dX_{ij} < 0$ and $dX_{ji} < 0$

2. MHS

d) $\boldsymbol{MHS_p > 0}$**, whereby** $\boldsymbol{dX_{ij} > 0}$ **and** $\boldsymbol{dX_{ji} > 0}$

	MHS	Frequency	Precent
None	= 0	287	69.8
	< 0	31	7.5
	> 0 & (dXij & dXji < 0)	24	5.8
MHS	**> 0 & (dXij & dXji > 0)**	**69**	**16.8**
	Total	411	100

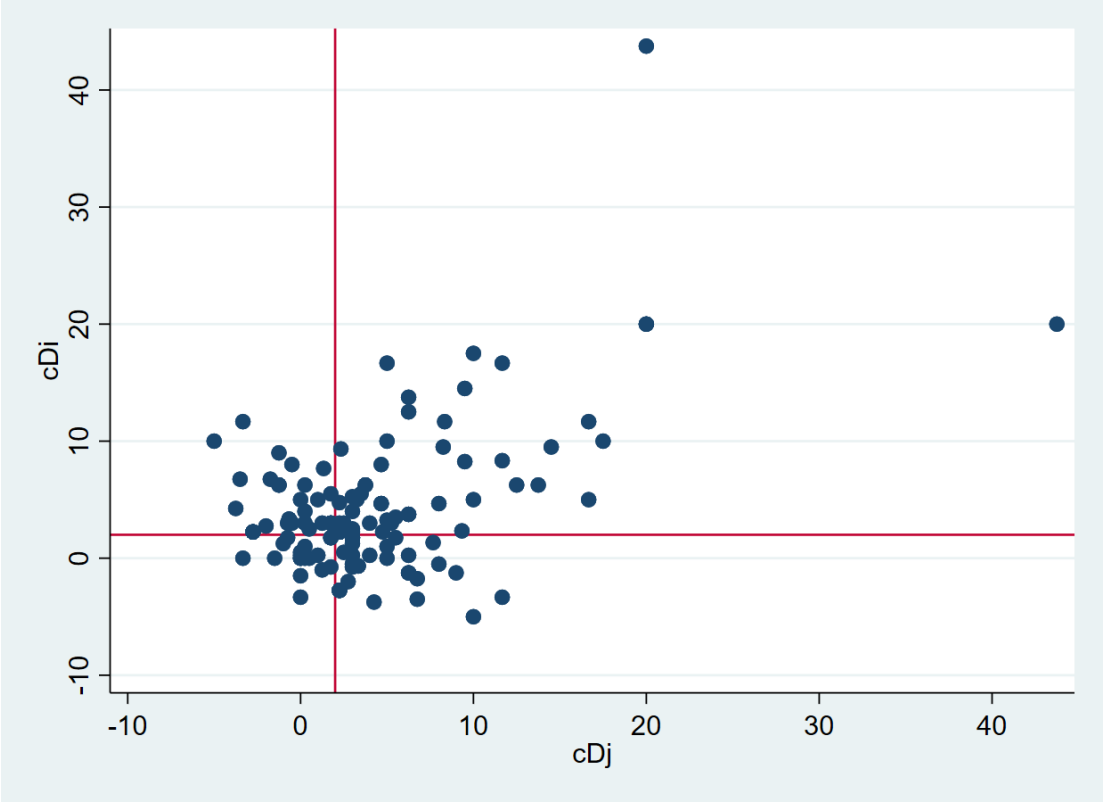
Method: 2-step approach

- Second Step: Identifying cases where MHS is not based on actual contribution
- How many of the 69 cases of MHS are due to gaming, asymmetric information, or behavioural bias?
- Comparative deviations measure:

$$cD_{ij} = \left(X_{ij} - \bar{X}\right) - \frac{\sum_{k\neq i,\, k\neq j}^{n}\left(X_{kj} - \bar{X}\right)}{n-2} = X_{ij} - \frac{\sum_{k\neq i,\, k\neq j}^{n} X_{kj}}{n-2}$$

$$cD_{ji} = \left(X_{ji} - \bar{X}\right) - \frac{\sum_{k\neq j,\, k\neq i}^{n}\left(X_{ki} - \bar{X}\right)}{n-2} = X_{ji} - \frac{\sum_{k\neq j,\, k\neq i}^{n} X_{ki}}{n-2}$$

Results



Results

- For all MHS pairs:
  - $cD_i \,\&\, cD_j \approx 0$: High scores are recognised by other team members.
  - $cD_i < 0 \; or \; cD_j < 0$: Other team members gave on average even higher scores!
  - $cD_i > a \,\&\, cD_j > a$: Both students gave each other '*a*' marks above the average score from other peers.

- Percentage of :

Threshold	MHS pairs (%)	All pairs (%)
a = 1	53.6	9.0
a = 2	**43.5**	**7.3**
a = 3	**24.6**	**4.1**

- We found that only for around 4 - 7% of all student pairs there is potentially an issue of collusion

Conclusions

- While we find some potential cases of excessive MHS, the problem appears to be comparatively small
- Nevertheless, the potential of students evaluating each other on factors that are unrelated to their actual contribution remains a main concern
- Using a measure to identify groups where there is a potential case of excessive MHS could be used to intervene.
- Furthermore, PE is difficult, therefore providing more guidance could help students to evaluate work more accurately

Thank you! Any questions?