

Peer and self assessment of oral presentations: effects of learner attributes



A. Mark Langan¹ & Dave M. Shuker²

Manchester Metropolitan University¹

Edinburgh University²



PHIL WHEATER
(Manchester Met Uni)

DAVE SHUKER
(Edinburgh Uni)

RICHARD PREZIOSI
(Manchester Uni)



Self and peer assessment



Catalyst for self- and peer assessment

Enhance reflection
Save marking!!



Personality types

Why we don't we do it more? Maybe perceived problems of accuracy, reliability, dishonesty?



Are these problems real? Are there predictable biases to self- and peer assessment? e.g. gender, background, assessment 'structure', (cognitive ability)

Self- and peer assessment: example

Calahonda field course (2002 & 2003)



Calahonda field course



Calahonda field course



Calahonda field course



Calahonda field course



Aim

To explore effects of learner attributes on self and peer assessment grades

Why use oral presentations on a field course?

Problems with anonymity, full participation, knowledge of student group, rapid, multiple tutors, popular form of assessment, relevant to employment, requires concise synthesis, difficult to plagiarise, personal experience

Method

- Talks 5 minutes long, in “thematic” sessions, with student chair, and designated “questioners”
- Talks assessed via student-driven criteria (‘participants’; n = 12)
- Marking: 40% “content”, 40% “presentation”, 20% “structure” (threshold descriptors provided)

Method

Each talk assessed by the tutors ($n = 11$), a subset of peers...

...and at end of the day self assessed by the student who gave the talk.

Data from 2002-3, $n = 60$ students

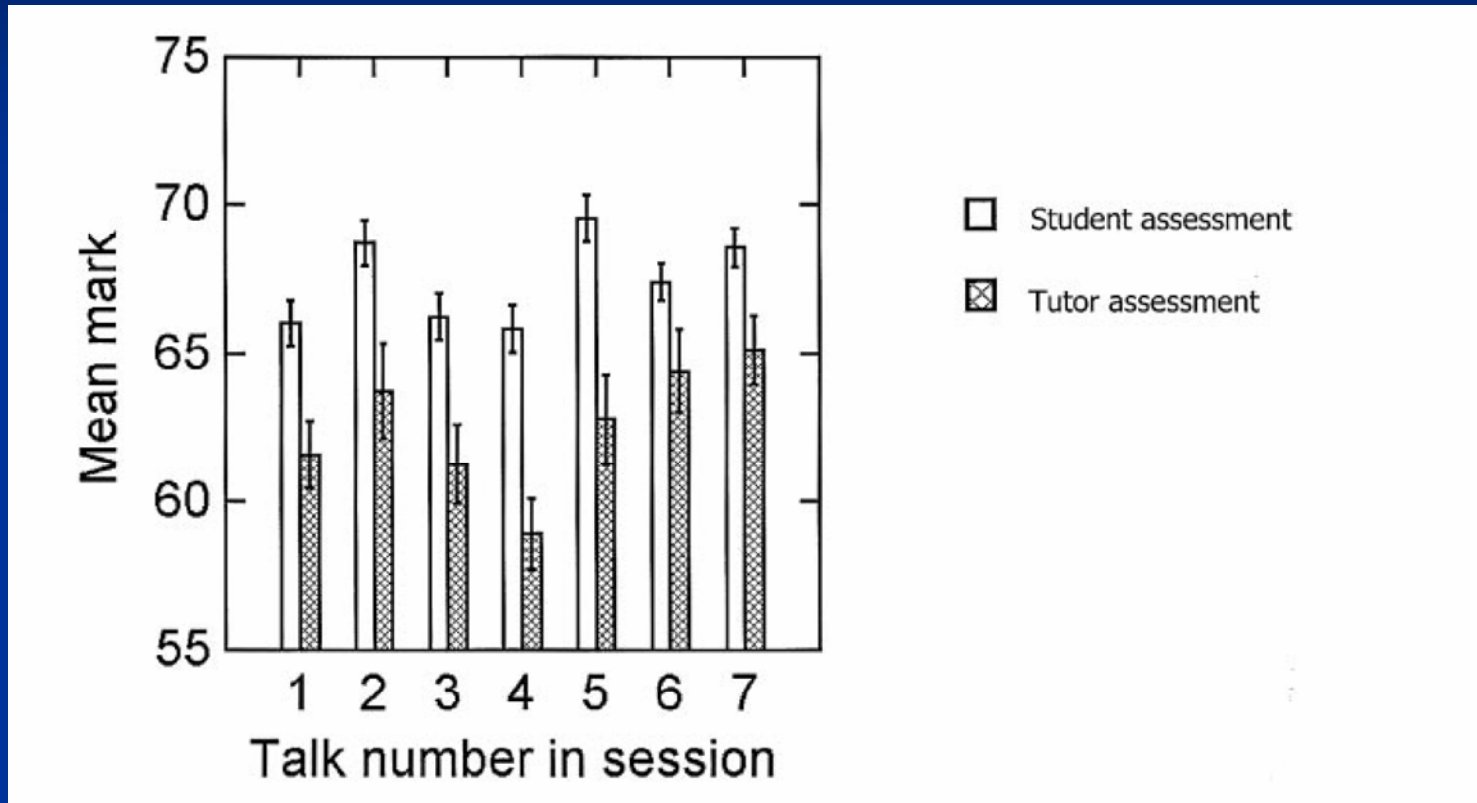
Langan et al. (2005), Langan, Shuker et al. (in prep.)

Usual assumption.....



....that... tutor marks measure “actual” attainment of students, and are free from bias (11 tutors provided assessments).

Marks dip during sessions (2002)

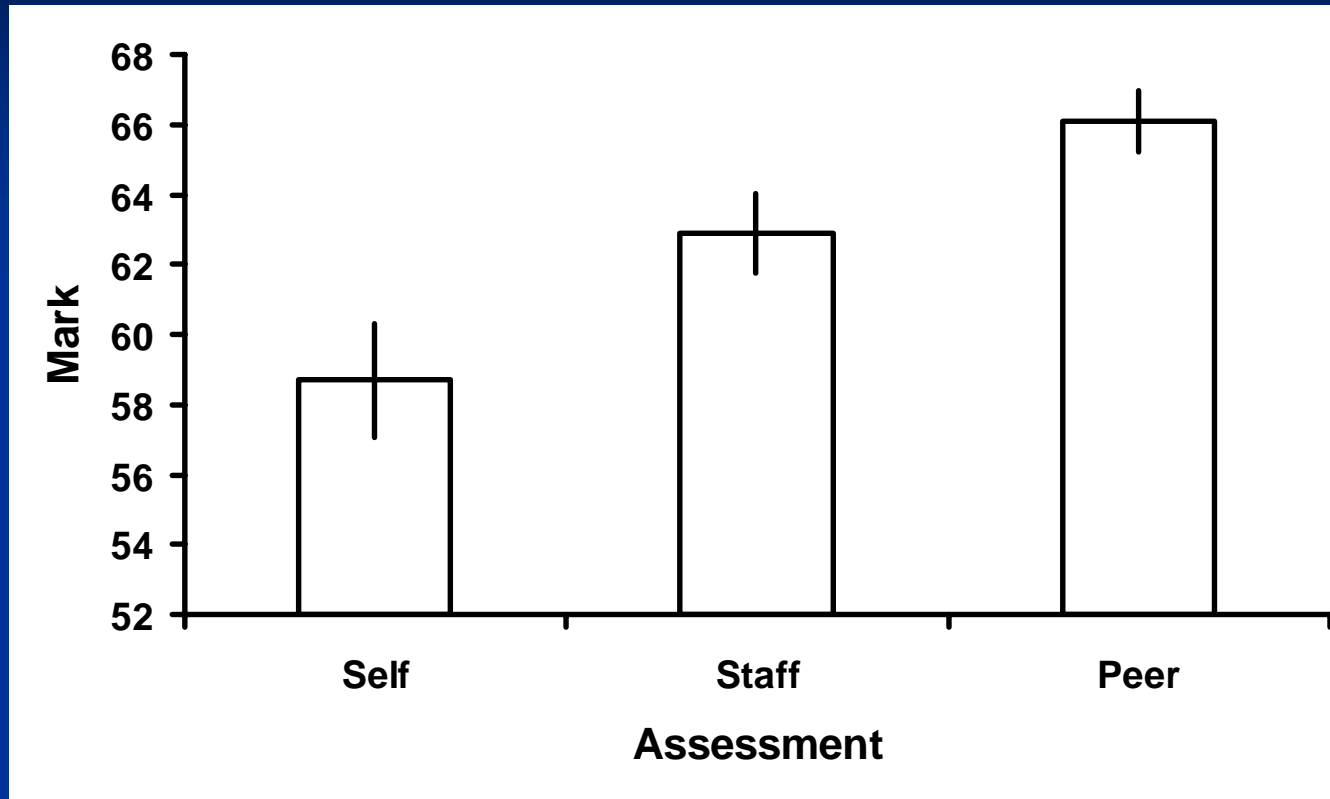


Effect was **WORSE** for tutors! ($P < 0.001$).

Hard work for tutors...

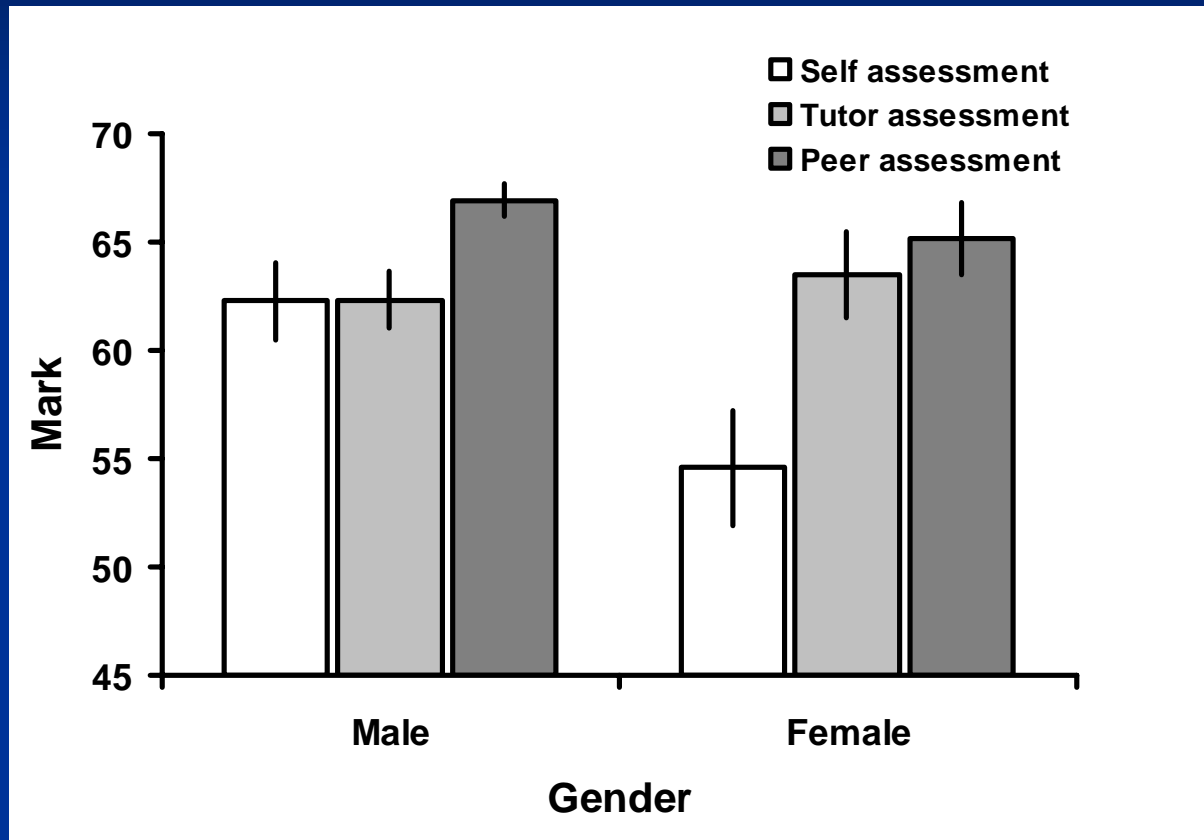


Self, peer and tutor assessment



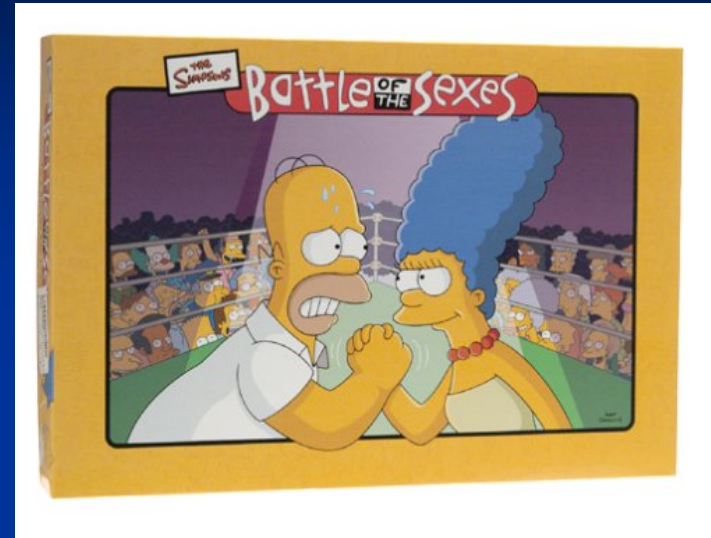
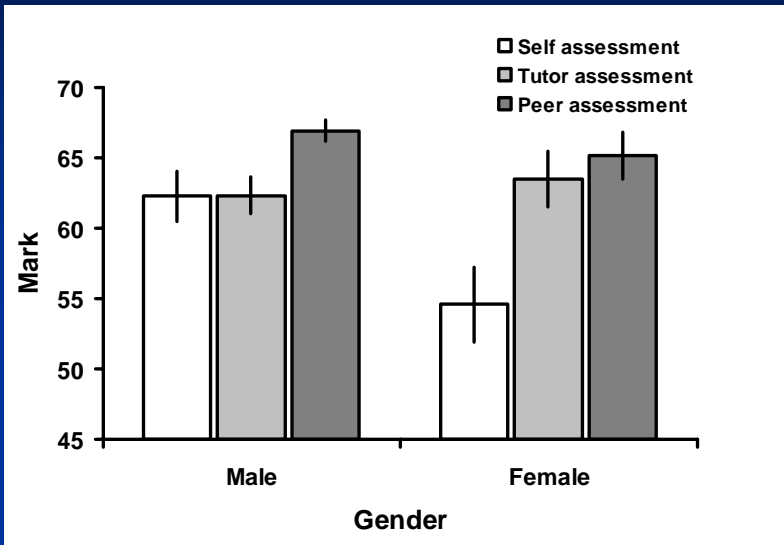
Compared to tutor grades, students **over-marked** each other, and **under-marked** themselves ($P < 0.001$)

The role of gender?



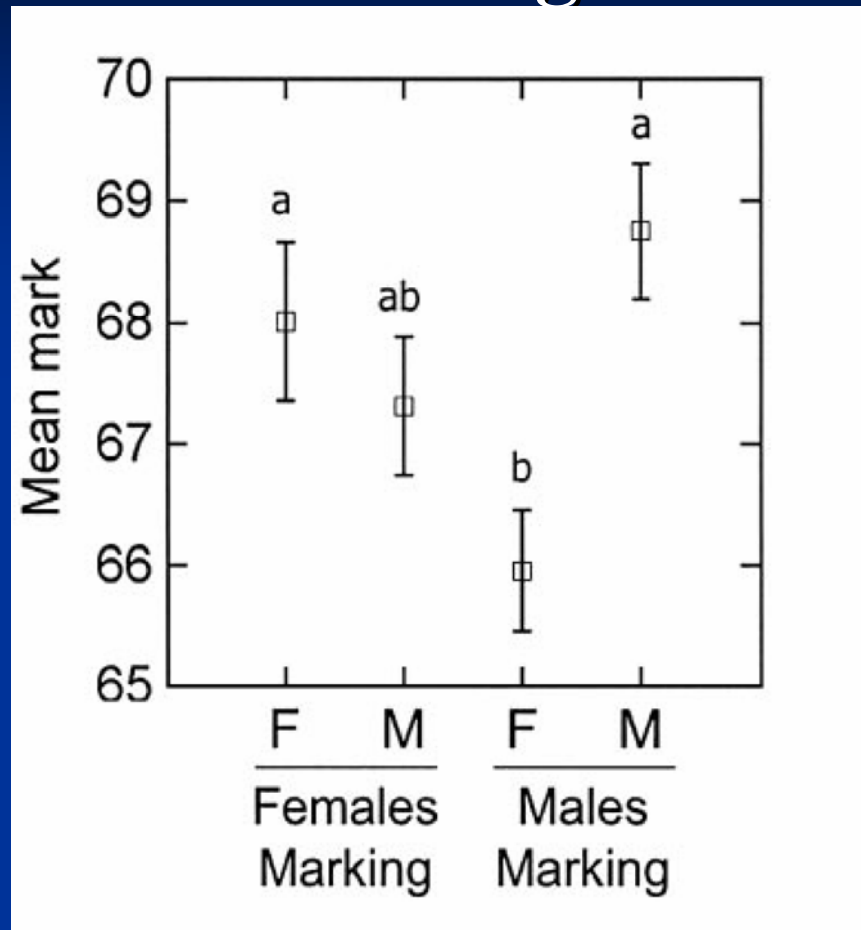
Female students **under-marked** themselves ($P=0.01$)

The role of gender



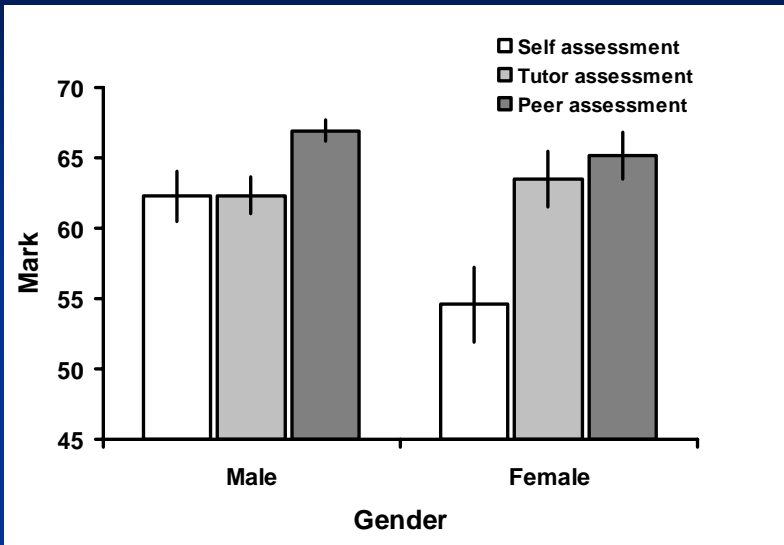
- Female students **under-mark** themselves
- Males receive **higher** or **lower** marks from peers, dependent on their University ($P=0.04$)
- Sexes award **higher** marks to own sex ($P<0.001$; due to males)
 - **No sex bias** from tutors ($P=0.67$)

The role of gender



- Sexes award **higher** marks to own sex ($P < 0.001$; due to males)

The role of gender



- Female students **under-mark** themselves
- Males receive **higher** or **lower** marks from peers, dependent on their University ($P=0.04$)
- Sexes award **higher** marks to own sex ($P<0.001$; due to males)
 - **No sex bias** from tutors ($P=0.67$)

The role of gender



What drives these sex differences?

Other factors:

■ Tutor assessment

No difference across years, small effect of University ($P=0.04$)

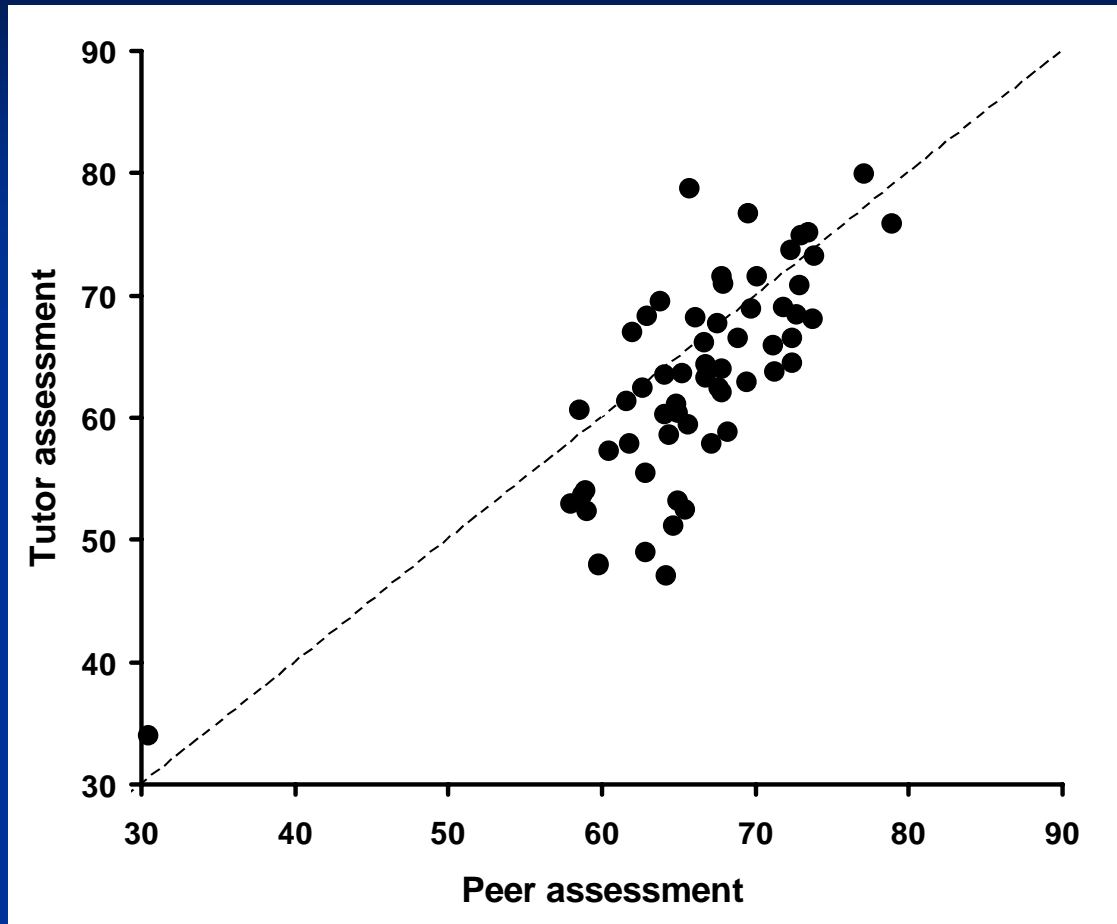
■ Peer assessment

Differences across years ($P=0.04$), and effect of University ($P=0.01$; and interaction with gender, $P=0.04$), with positive discrimination towards “own” University ($P<0.001$)

■ Self assessment

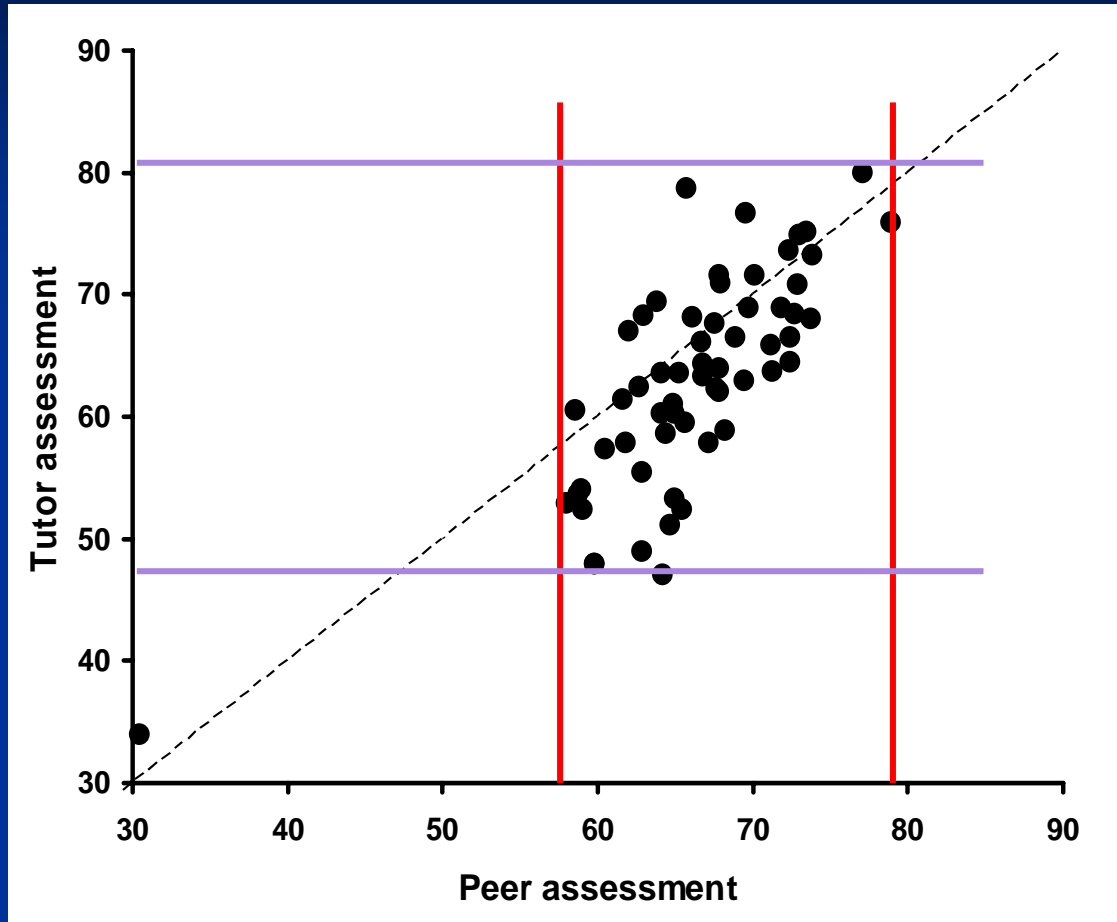
No difference across years ($P=0.80$), small effect of University ($P=0.04$)

Peer assessment



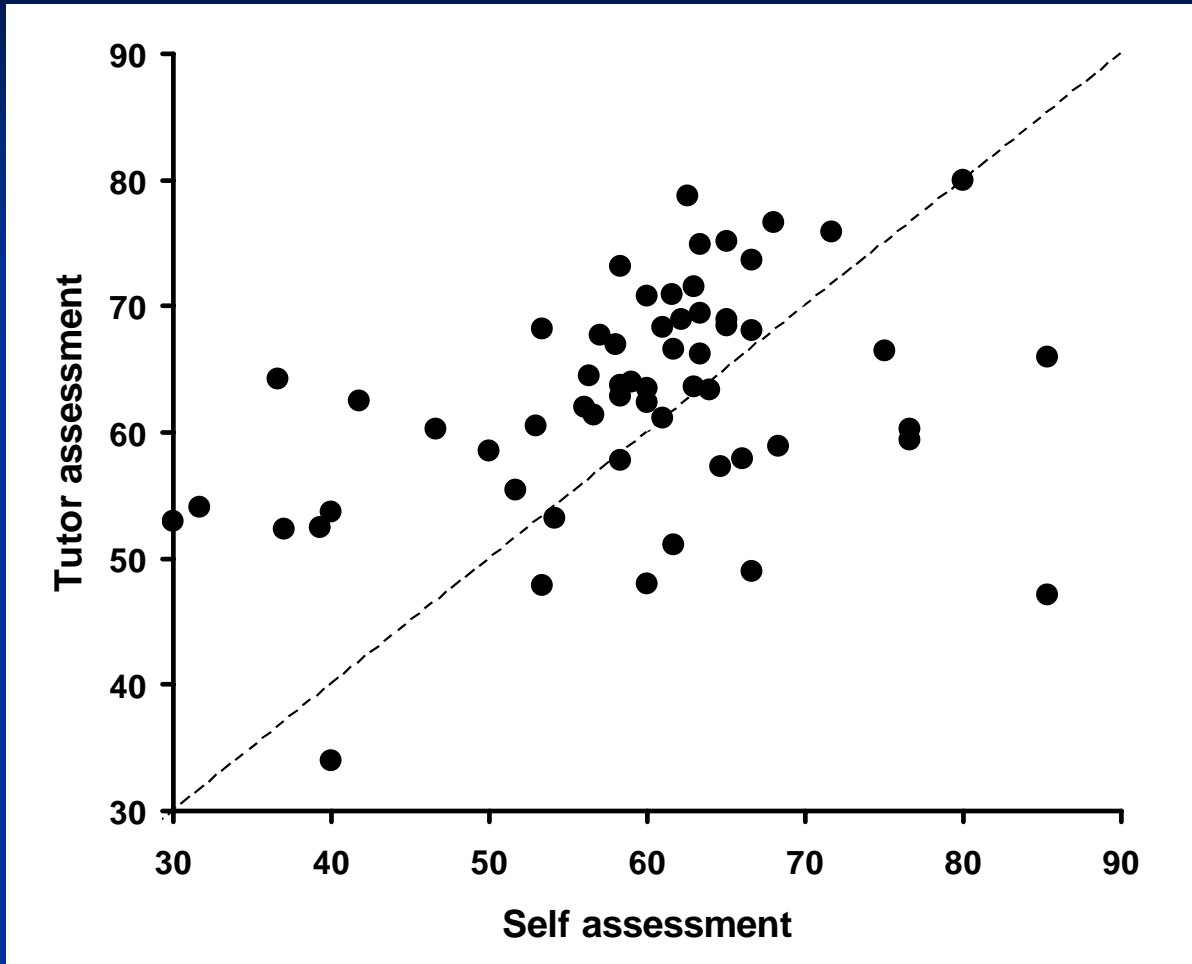
Peer and tutor marks correlated ($P < 0.001$, $R^2 = 0.59$)

Peer assessment



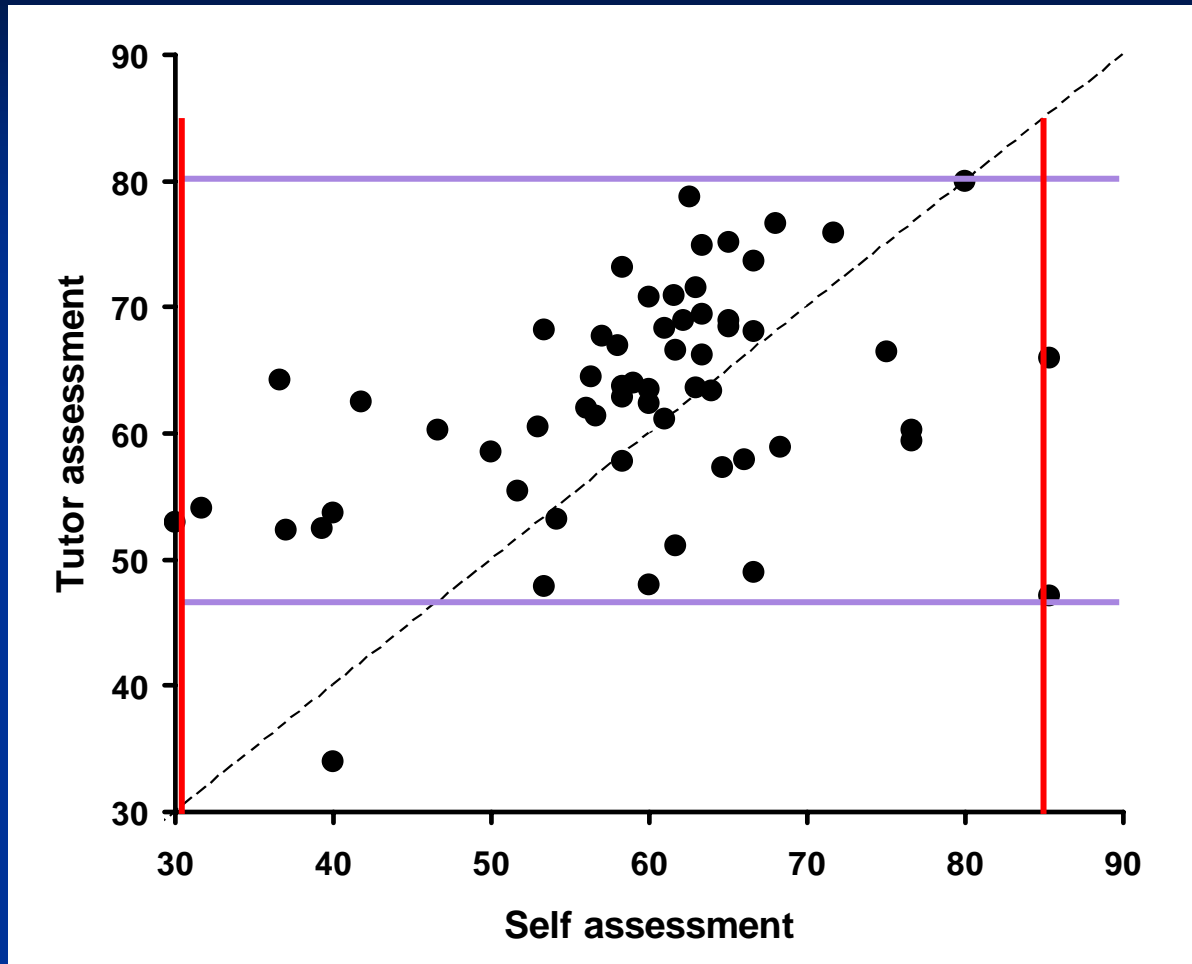
NARROWER range of marks than tutors

Self assessment



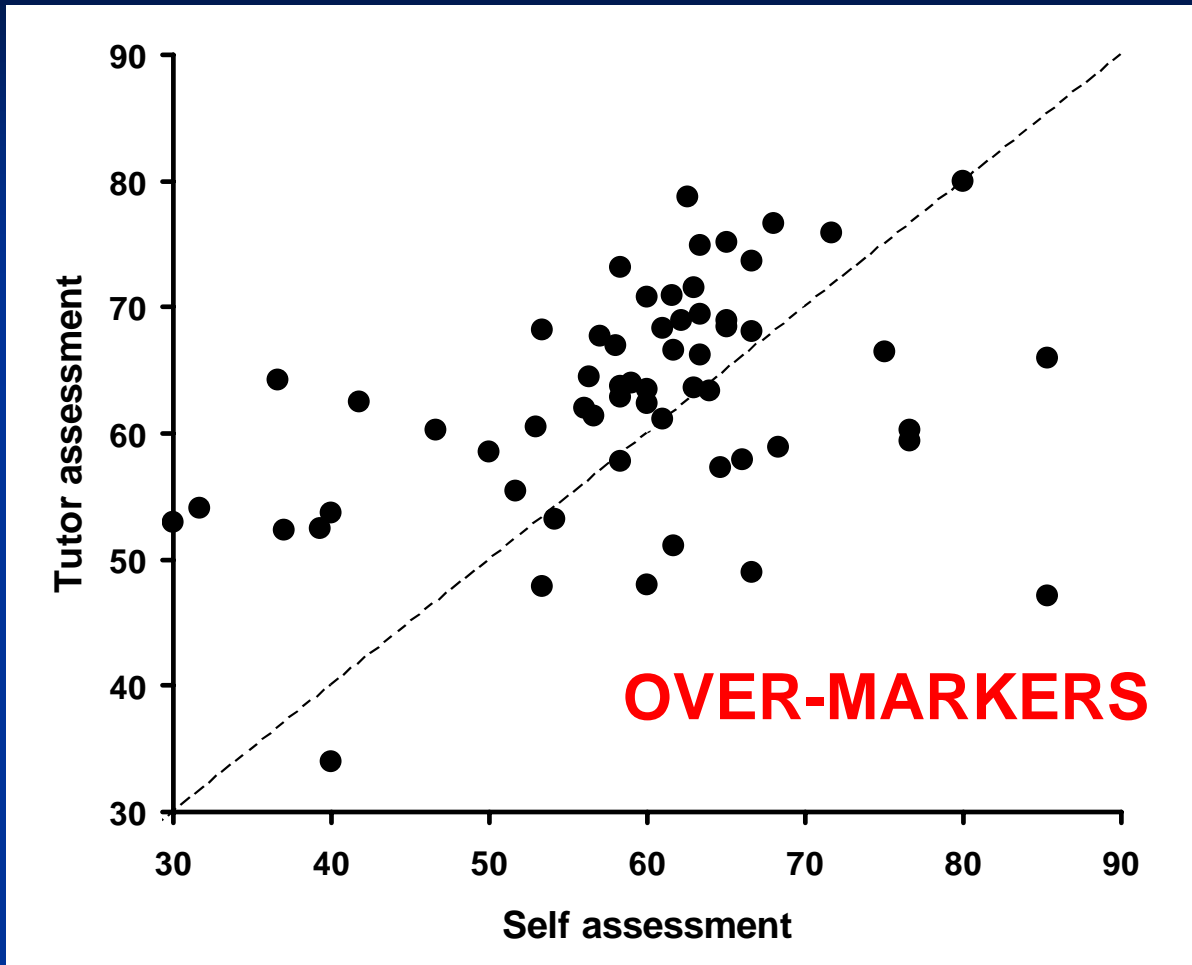
Self and tutor marks less strongly correlated ($P < 0.01$, $R^2 = 0.10$)

Self assessment



BROADER range of marks than tutors

Self assessment



LOW achievers over-marked themselves

Involvement with creating assessment criteria

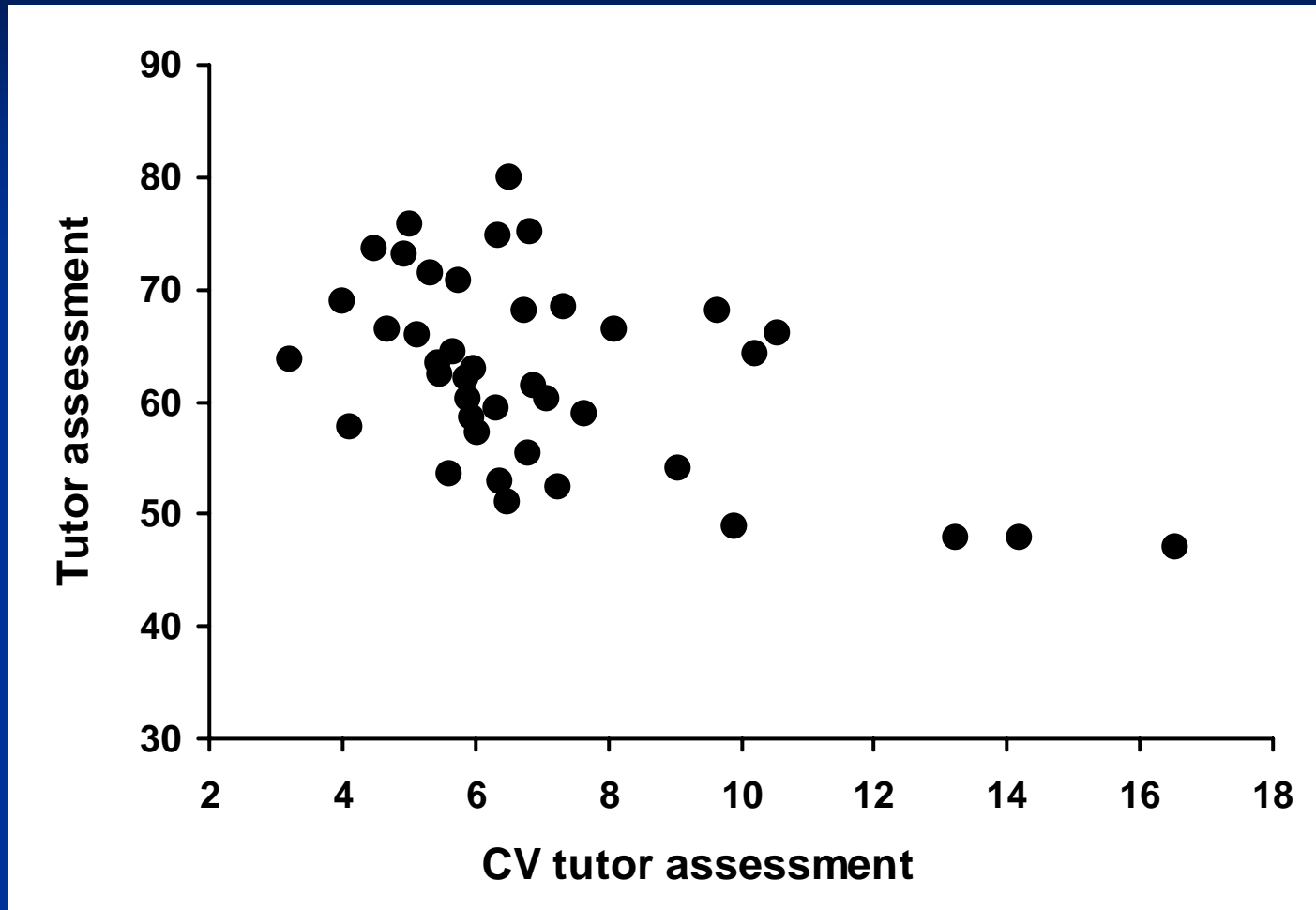
■ Peer assessment

Students involved in the criteria gave peers **lower** marks ($P=0.002$), received **similar** marks from peers ($P=0.14$), and received **lower** marks from tutors ($P=0.05$)

■ Self assessment

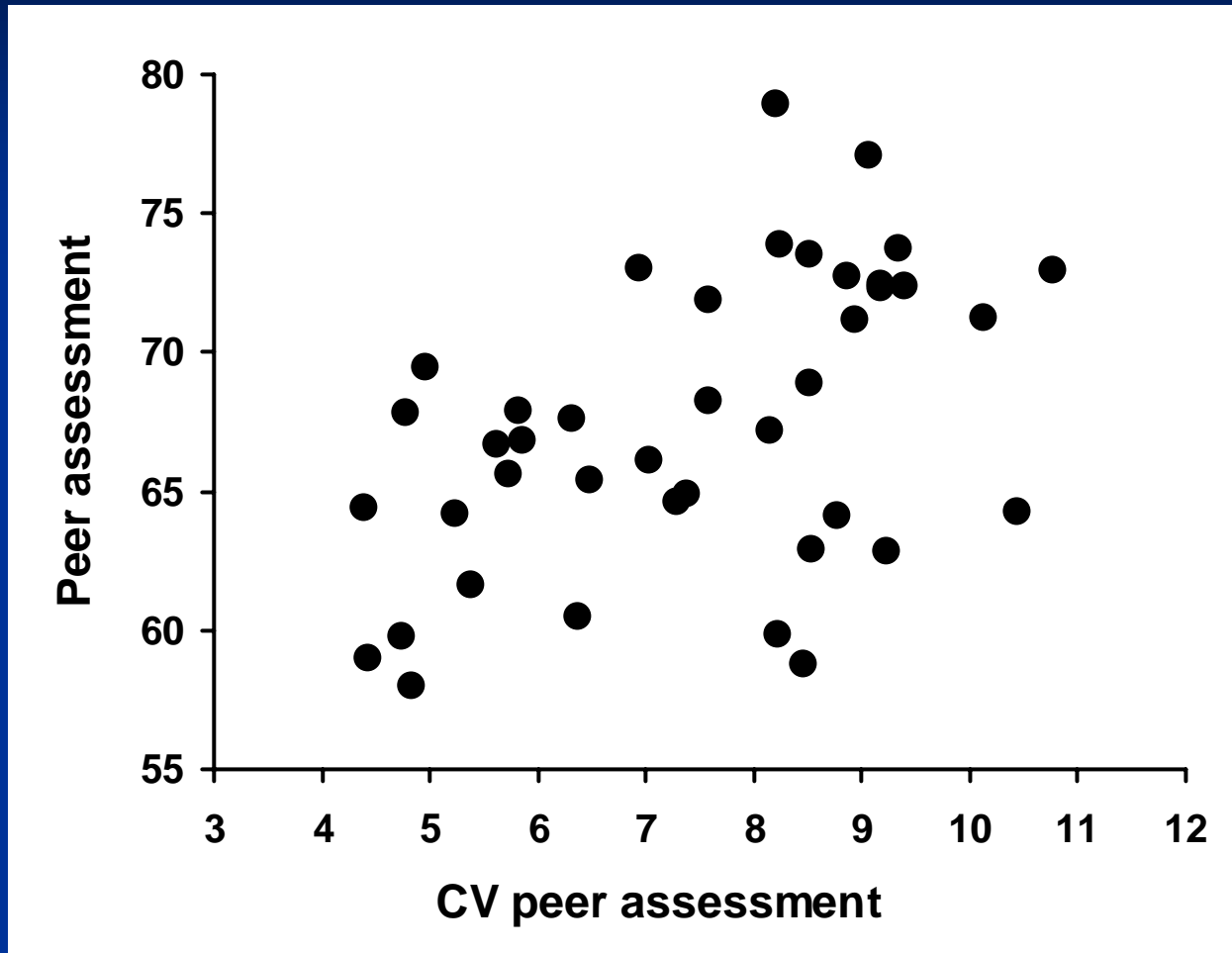
Student involvement in criteria **did not influence** self-assessment mark ($P=0.10$; weak interaction with gender and university affiliation, $P=0.04$)

Agreement within assessor class



Tutors disagreed more about low achievers ($P < 0.001$)

Agreement within assessor class



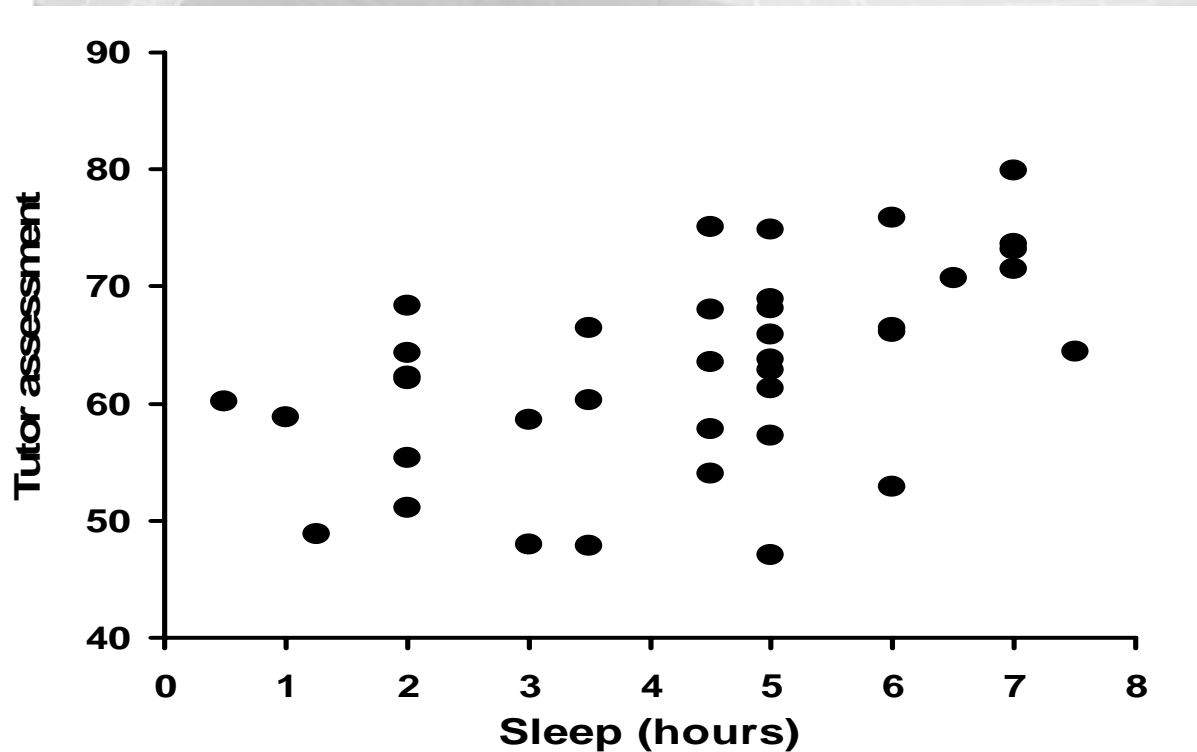
Peers disagreed more about **high** achievers ($P < 0.01$)

Sleep



Sleep

- **Self assessment:** not correlated with (reported) sleep
- **Peer and tutor assessment:** positively correlated with sleep ($P=0.02$ and <0.001)



Student feedback



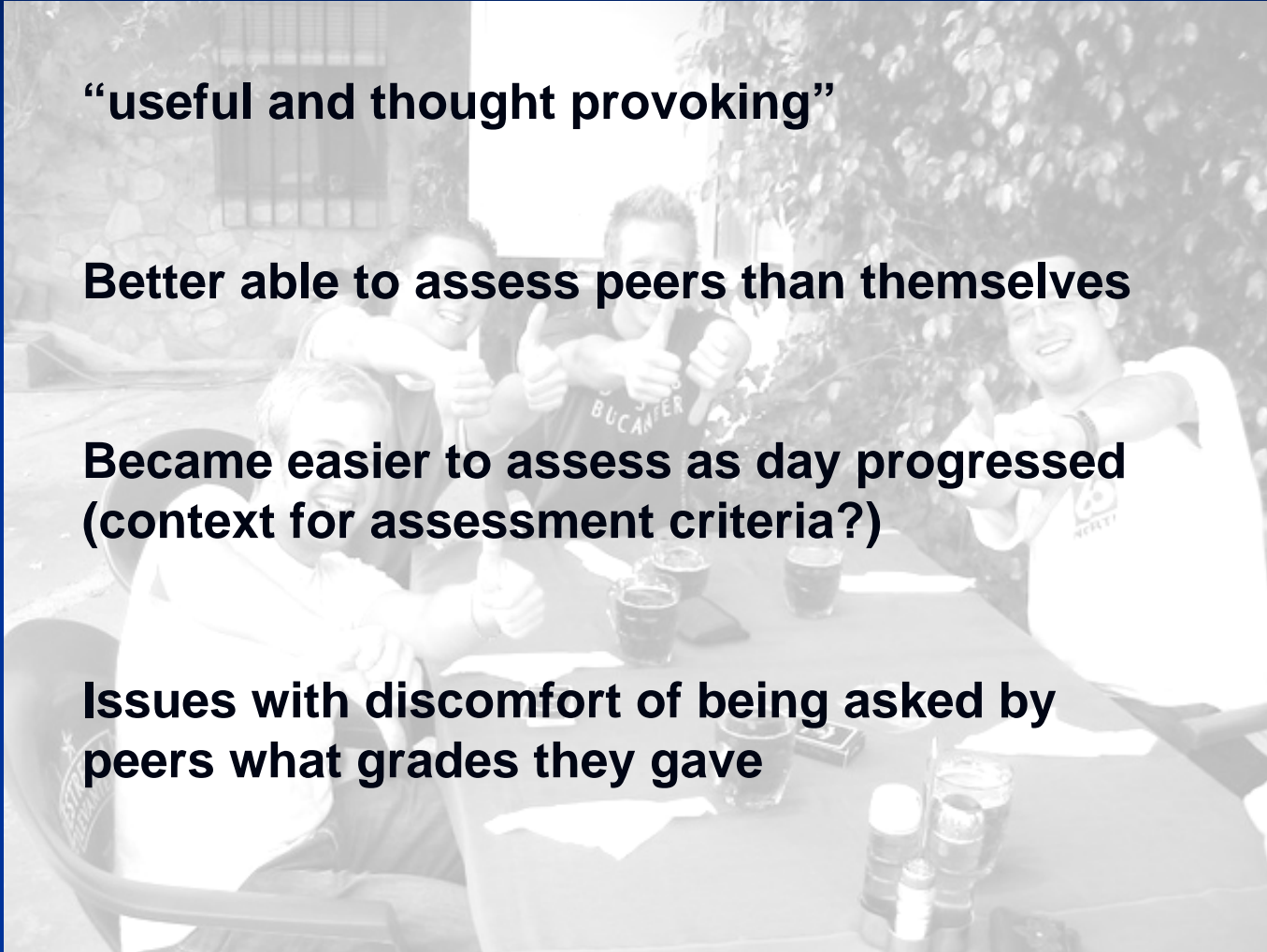
Student feedback

“useful and thought provoking”

Better able to assess peers than themselves

**Became easier to assess as day progressed
(context for assessment criteria?)**

**Issues with discomfort of being asked by
peers what grades they gave**



Implications

- Oral presentation assessment risk gender bias if assessed via self and/or peers
- Raises the question ‘Are the peer awarded marks unsuitable for summative assessment?’ Positive feedback in terms of formative assessment
- Are presentations too difficult to assess? Can students be trained in advance?
- Follow-up study – do students learn from their experiences?

Review

- Specific situation: oral presentations on a field course (how far can findings be generalised?)
- Minor biases detected. Can biases be ameliorated (e.g. gender differences, narrow range of peer awarded marks)?
- Are the students learning? Follow-up studies to demonstrate improvement in PA/SA
- Tutor experience ranged from 1 - 30 years

Why do we view ourselves and others in different ways?





Mark Langan: m.langan@mmu.ac.uk

		Self awarded marks (by speakers)	Peer awarded marks (to speakers)	Tutor awarded marks
Learner attributes	Gender	Males closer to tutor grades. Females greatly differed between Universities	Males marked males higher	NS
	University	Minor institutional differences	Minor institutional differences	Minor institutional differences
	Sleep	NS	Speakers with less sleep received lower grades	Speakers with less sleep received lower grades
	Participant	NS. However, effects of participating in assessment criteria generation interact with gender and university of speaker	Participants received slightly lower marks	Participants received slightly lower marks
Structure	Session number	NS	Differences between quality of sessions detected	Differences between quality of sessions detected
	Position of talk in session	NS	Dip in marks observed in 2002 with seven talks per session did not occur with only four talks per session.	Dip in marks observed in 2002 with seven talks per session did not occur with only four talks per session
	Year	NS	NS	NS
	Timing accuracy	Shorter talks led to lower self awarded marks	Shorter talks received lower marks	Shorter talks received lower marks

Table 1. Summary of the effects of learner attributes and ‘structural’ components of oral presentations on self, peer and tutor marks. Conclusions are drawn from a range of analyses, where significance was set at $P < 0.05$ in all cases. NS = no significant effects. For completeness, findings from a previous study are included (Langan et al. 2005).