

Peer assessment

Some principles, a case,
and computer support

Stephen Bostock, Keele University
stephen@cs.keele.ac.uk

Student peer assessment

- Assessment of student work by students
 - Formative (reviewing)
 - Summative (grading)
- Quantitative or qualitative or both
- For assessed coursework (draft & final)

2

'...peer assessment promotes(s) lifelong learning, by helping students to evaluate their own and their peers achievements realistically, not just encouraging them to always to rely on evaluation from on high'

Sally Brown 1996, Assessment,
in DeLiberations
www.lgu.ac.uk/deliberations/

3

Potential benefits to authors

- Extra feedback tutors can't provide
 - More intelligent feedback than MCQs
 - Less expert feedback than from a tutor
- In software development, evaluation by a user or peer is appropriate
- Authors receive multiple views, but using the same criteria

4

Potential benefits to assessors

- Motivating - sense of ownership of the assessment process, integrated with learning
- Encouraging self-assessment, needed to manage own learning
- Encouraging responsibility, autonomy in learning
- Encourages deep learning approach
- Understanding the assessment criteria
- Practising evaluation: a key skill (and a discipline skill)

5

Benefits as a learning activity

- Assessment/evaluation requires subject skills and knowledge so it reinforces subject learning
- Many ICS courses include evaluation as a learning outcome, so an added benefit of practising the skill for its later assessment
- Academic values: apprentice students into the academic community, where anonymous peer review is a key process
- So, a range of benefits ...

6

Win/win/win?

- Feedback is needed for learning, ideally 'detailed, positive and timely' but staff-student ratios restrict it
- Students are a 'free resource' – use it to free staff time to do less, but better, marking and feedback
- Assessors benefit too!
- But
 - dubious quality of feedback or marking, issues about criteria and assessment practice
 - Anonymity needed for reviewers and authors

7

How to organize it?

- Generate multiple assessors per author (4 here) with restrictions on assessor-author pairs
- Student web spaces make work accessible
- Email to point assessors at the work
- Web forms to provide criteria and collect assessments, to email results to authors (if formative) and to store for tutor
- Ideally: Web interface for tutors to create assessment events

8

A Case over 2 years

- The Multimedia and Internet module of the MSc in IT, Keele
- Taught by Stephen Bostock and Dave Collins
- 50% assessment by one piece of coursework, a web site and a short report, against criteria
- Student Web sites accessible within Keele
- Only assessor-anonymity possible as URL includes username – authors are known
- Web sites 'frozen' at the prototype and final submission deadlines

9

Year 1: 1999/00 – 38 students

- Assessor-author pairs created manually and assigned an identifier code
- Emails sent manually to request assessments
- A Web form collected the formative reviews, emailed to authors and tutors (so reviewers know author username)
- A Web form sent summative assessments grades to the tutor
- All grades tutor 'moderated' i.e. re-marked!

10

Results of year 1

- Administration was time-consuming and error-prone
- 35 of 38 students did formative assessments of text and %s
- Only 22 did summative assessments - in 'revision' period - an average of 2.3 per author
- Mean mark 64% (tutor 63%)
- But unreliable: mean range of 11% per author, SD of 7%, correlation tutor-student mark = 0.45

11

Student evaluations

- Of 16 returns from 38, most said:
- Anonymity allowed criticisms to be 'ruthless', and more valuable
 - Text criticisms of prototype were more valued than marks
 - Timing – needed longer to use the criticisms
 - Seeing other students' work was valuable
 - Many anxieties about summative grading, so must be tutor moderated

12

Plans for year 2

- Better assessments by students may be possible if
 - More detailed criteria given or negotiated
 - Practice in assessment given
- Assess prototypes earlier to give more time for improvements
- Automate the administration - PRoMT

13

Web support - PRoMT



Year 2: 2000/01 – 68 students

- Assessments submitted are identified by code number plus assessor username, to allow double-anonymity (except for student URL!) and non-duplication
- Assessments turned into web pages for policing and moderating, as well as emailing to authors and storing
- A practice assessment on last year's work and its 5 criteria
- Further development of 7 criteria

15

Web assignment criteria

1. Robustness
2. Screen design
3. Navigation design
4. Text design and readability
5. Use of graphics, and any appropriate audio or video
6. Content structure
7. Conceptual design

16

Results of year 2

- In initial practice
 - Average marks close to tutors but
 - SD of student marks = 11.3%
- Formative assessments
 - 3.4 per author, SD 8.8%
- Summative assessments
 - 59 authors had 3.5 assessments on average
 - Overall mean 63% (as tutor)
 - SD 6.2% , range 13.5% per author
 - Tutor's quick marking: correlation = 0.59
- Better but still not reliable

17

36 student evaluations

- Was practice marking useful? 89% Yes
- Was criteria discussion useful? 84% Yes
- Reviews done professionally? 60% Yes, 20% No
- Prototype reviews: 66% useful/very useful for improvements
- Happy with *moderated* summative peer assessment? 61% Yes, cautiously
- Should we do it next year? 77% Yes

18

Student views - best aspects

- Constructive criticisms on prototype
- Clarified assessment criteria
- Helped understand design issues
- Seeing other students sites
- Involvement with module

19

Student views - worst aspects

- Lack of author anonymity (and reviewer? – peer pressure)
- Time taken for assessing
- Some poor, unhelpful reviews received
- Own prototype was too incomplete to benefit from review

20

Were the benefits gained?

Yes to a degree but

- Limited by lack of anonymity of authors and (some) assessors
- Unfair pressure on assessors? Is it their job?
- Value to authors limited by expertise of assessors, so multiple assessors needed and that takes more student time. Is 4 enough?
- Average mark as tutors (63%) but unreliable
- Moderation of all summative marking needed: at least for the inconsistent grades

21

Conclusion

- Formative assessments valuable for assessors and authors
 - Constructive criticism of a prototype useful
 - Reviewing others' work a valuable activity
 - Clarifies criteria, demonstrated in action
- Summative assessment - worth the costs?
 - Student anxiety
 - *More* staff time unless software support
- Multiple assessors and double anonymity needed
- Assessment must include penalties for non-compliance

22

Final reflections

- Some students respond well, but others are poor assessors, competitive or unscrupulous. So policing and penalties are necessary, and thus more administration.
- Without computer support no staff time is saved.
- When PRoMT is finished it will be made accessible.

23

Further information

- stephen@cs.keele.ac.uk
- http://www.keele.ac.uk/depts/cs/Stephen_Bostock/
- Student Peer Assessment
http://www.keele.ac.uk/depts/cs/Stephen_Bostock/docs/bostock_peer_assessment.htm
- CAA – experiments in 3 courses
http://www.keele.ac.uk/depts/cs/Stephen_Bostock/docs/caa-ktn.htm

24