Peer Assessment in DBL

Mahender Mandala, MS
Mary Goldberg, PhD
Jonathan Pearlman, PhD
Peer Assessment
Methods

- MEMS 0024 – Introduction to Mechanical Engg. Design (Fall 2014)
- Students worked in teams of 3-6 and submitted design log books for review
- Students assigned 1 project to review individually and 1 project to review as a team (their own project team)
- Two assignments – mid-point and final
- Data collected Pre-course and Post final deliverables.
SWORD

- Scaffolded Writing and Re-writing in Discipline
  - Developed by Schunn et al
  - Used in writing domain
  - Allows for summative and formative peer assessment
  - Grades for document submitted, accuracy of reviews, and back-evaluation scores.

- In MEMS0024, SWORD used primarily for feedback; grades are calculated by TA and Teacher by reviewing projects and group evaluations in the end.
Class Composition

**Major**

- Bioengineering: 45
- Electrical & Computer Engineering: 2
- Mechanical Engineering & Material Science: 121
- Other: 2
- Undecided: 1

**Year in college**

- 5th Year Senior: 7
- Freshman: 1
- Junior: 26
- Senior: 20
- Sophomore: 117

Data labels are student ‘N’
Pre-college K-12 Information

Type of K-12 School

- Public district school: 139
- Private school: 24
- Public magnet school: 5
- Home school: 1
- Prefer not to answer/Don't know: 2

Have you taken pre-engineering or specialized STEM related course?

- Yes: 76
- No: 93
- Prefer not to answer: 2

Have you participated in a special STEM-related team or similar extracurricular activity?

- Yes: 43
- No: 125
- Prefer not to answer: 3

Data labels are student ‘N’
Opinion about critiquing projects

Key features
✓ Gave me a chance to see how others approach problems and get new ideas
✓ Allowed me to think of my own design flaws when reviewing others

It's easier to be critical of other people's projects, so you find yourself making comments for improvements in peer projects for issues that your own project has. When you return to your own project after the review, you look at it with a different perspective and can see those issues.

It allowed me to step back and see other's work. In doing this I noticed many similarities and differences between different groups and used this information to enhance my own group project.

Waste of my time

Was critiquing peer projects beneficial to you?

Data labels are student ‘N’
Opinion about feedback received

Key features

- Received constructive feedback, but were hesitant receiving peer feedback
- Comments were vague/less critical

The feedback we received was okay but nothing stellar since the reviewers didn't fully understand our project.

We mostly paid more attention to the feedback we got from our professor. Oftentimes, peers would misunderstand a concept because they hadn't read something earlier.

I liked getting feedback, especially positive, about my work. It is assuring and increases my confidence in my work and allows me to see that my time and effort paid off in the eyes of others.

Was feedback you received helpful in improving your design/project?

Data labels are student ‘N’
Reviewing as a team vs as an individual

What was beneficial to you or your team?

- Group Review: 64
- Individual Review: 77

Individual Learning Gains

- Group Review: 49
- Individual Review: 91

Ease of Conducting

- Group Review: 50
- Individual Review: 91

Quality of Review Generated

- Group Review: 73
- Individual Review: 67

Data labels are student ‘N’
Reviewing as a team vs as an individual

Each member of the group had their own interpretation of the project we had to review and it made me, personally, more open-minded.

... with the group you have many different input critic at the same time. it allows you to see if others see the same thing as flaw as you do and you get the chance to actually debate it. It is more fun too compared to doing it alone.

The individual reviews were much more constructive and honest.

It was a little bit annoying to meet up with the group just so we could review someone else's assignment. I think we would all rather have done it individually.
Higher is better, except for anxiety
Confidence significantly increased, $t(122)=10.273, p<0.001$

Motivation significantly increased, $t(122)=2.477, p=0.015$

Success significantly increased, $t(122)=7.484, p<0.001$

Anxiousness significantly decreased, $t(122)=-3.223, p=0.002$

Higher is better, except for anxiety
Limitations

- Self-reported outcomes for experience with peer-assessment
- Difficult to determine impact of SWORD specifically
Future Work

- Develop topology of peer feedback in design domain
- Analyze revision of projects and relations to feedback received
- Compare student comments to expert comments
- Empirical research on feedback received and student project performance
**1. Preliminary Design Review, 1 draft (10/21/2014)**

**Name**

Preliminary Design Review

**Description**

In this assignment, you will submit your Design Log Book containing preliminary designs for your team, then review Design Log Books from three other groups.

Note that files cannot exceed 4Mb in size, so it is best to submit a .pdf file. If needed, paste images into your document at a low resolution.

Peer review will open one day after the submission due date. When doing a review, please answer all the questions and give useful feedback to your classmates. Your efforts in providing feedback will be evaluated and graded. Please refrain from posting comments that are offensive or that do not contain constructive feedback.

**Drafts**

**Draft 1:** Edit Deadlines and Settings
- Submit by 10/21/2014 + 1 day grace
- Review by 10/31/2014 + 1 day grace
- Backevaluate by 11/16/2014
- Revision planning tool is disabled.
- Thesis detection tool is disabled.

Add a Draft
### Reviewing Dimensions

**Settings**

- **Reviewers:**
  - Must do 3 reviews
  - Manually assign reviewers for each draft
  - Backevaluations required

- **Example Document:** None

- **Grading:**
  - Curved to mean of 85.0%, Stdev 10.0%
  - Weights: Task 20%, Reviewing 40%, Writing 40%
  - Teacher's grade overrides student's grade

### Hypothesis

- Did the group correctly identify all design requirements, constraints, limitations and features of the chosen project?
- Based on the client statement and the requirements, are the hypotheses acceptable?
- Is there anything that has been overlooked and not taken into consideration?
- Do you think the group understood the client statement and requirements?

(1-5 comments)

**Hypothesis Rating.** How well did the team understand the project requirements and form hypotheses?

Levels

### User Discovery

- Based on the hypotheses and client statement, did the group ask appropriate questions to the potential users?
- Are the questions well thought and clear for the interviewees to understand?
- Did the group identify the appropriate potential users?
- Did you identify anyone in the list of people interviewed who are most likely not potential users?

(1-5 comments)

**User Discovery Rating.** How well do you think the team did in initial user discovery?

Levels

### Idea Generation

...
## Idea Generation

Did the group re-write the client statement based on the interviews? Are the changes appropriate?

Did the group demonstrate an appropriate level of maturity in interpreting and understanding user feedback?

When generating ideas, did the team sufficiently discuss and take into account all of their user feedback when coming up with ideas?

Were multiple methods used to generate ideas?

Did the team seek additional solutions beyond the first potential concepts?

(1-5 comments)

**Idea Generation Comments.** How well do you think the team used brainstorming and other techniques to generate a rich set of possible design solutions and features?

<table>
<thead>
<tr>
<th>Levels</th>
</tr>
</thead>
</table>

## Preliminary Designs

Were the candidate solutions reasonable attempts at meeting the problem statement?

Were they distinct enough from each other to represent fundamentally different ideas? Did they demonstrate an appropriate level of technical maturity?

Could you understand the design intent? Were they adequately evaluated against each other?

(1-5 comments)

**Preliminary Designs Rating.** Do the designs presented represent different and meaningful attempts at solving the problem?

<table>
<thead>
<tr>
<th>Levels</th>
</tr>
</thead>
</table>

## Presentation of Designs

Did the level of detail and images used in presenting the chosen design adequately convey the design intent?

Did the models emphasize the important parts of the design?

Do you feel that another engineer could pick up the concept based on this presentation?

(1-5 comments)

**Presentation of Designs Rating.** How well did the sketches convey the design intent?

<table>
<thead>
<tr>
<th>Levels</th>
</tr>
</thead>
</table>
Reliability of the mean peer judgments for each dimension (based on an average of 8 ratings per dimension)

The stability of the mean rating on each dimension; it is influenced by 1) the consistency of peer ratings with each other and 2) the number of peer ratings.

Rating Dimension

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Overall rating reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Discovery</td>
<td>0.95</td>
</tr>
<tr>
<td>Initial Prototype</td>
<td>0.94</td>
</tr>
<tr>
<td>Evaluation of Design</td>
<td>0.55</td>
</tr>
<tr>
<td>Presentation of Final</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Tip: Go to high disagreement cases for the problematic dimensions to see what possible areas of disagreement might be.

Average peer rating for each dimension

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Average rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Discovery</td>
<td>5.9</td>
</tr>
<tr>
<td>Initial Prototype</td>
<td>5.8</td>
</tr>
<tr>
<td>Evaluation of Design</td>
<td>6.1</td>
</tr>
<tr>
<td>Presentation of Final</td>
<td>5.8</td>
</tr>
</tbody>
</table>
Mary Goldberg, PhD  |  mrh35@pitt.edu
Assistant Professor

Mahender Mandala, MS  |  m.mandala@pitt.edu
Graduate Student Researcher