Student Reflections on Peer Reviewing Solutions to Model-Eliciting Activities

Heidi A. Diefes-Dux and Matthew A. Verleger
Purdue University, hdiefes@purdue.edu, mverleg1@purdue.edu

Abstract – **A double-blind peer review process is embedded in the implementation of Model-Eliciting Activities - a type of open-ended problem used in a large first-year engineering course. Students conduct the peer review along three dimensions: Mathematical Model, Re-usability & Modifiability, and Audience (Share-ability). Classmates bring to bear both their own solution development experience as well as their own educational and personal backgrounds when providing feedback to their peers. In this paper, we examine the results of a reflection instrument used to investigate how students felt about their ability to provide reviews along the three rubric dimensions across three MEAs implemented in a single semester. Further we exam the results of a second instrument that is used by student teams to provide feedback to their reviewers on the quality of the reviews received.**

Index Terms – Peer Review, Peer-Assessment, Model-Eliciting Activities

INTRODUCTION

Students need to develop teaming and communication skills, proficiency in engineering science and design, as well as an ability to address open-ended problems replete with ambiguity and uncertainty - this is both nationally and globally recognized by engineering educators [1,2]. One instructional approach to developing these competencies is Model-Eliciting Activities (MEAs) – a type of open-ended problem [3-5]. This approach has been taken in a large, required first-year engineering course at Purdue University [3-5]. One of the challenges in adopting this approach, however, is the adequate provision of high-quality and quantity of feedback during students' iterative cycles of improvement to their solution. Time becomes a critical issue when multiple feedback points are desired and the problem solutions being assessed are complex.

To address this challenge, a double-blind peer review process has been adopted in the first-year engineering course. This enables the addition of one feedback point that was not possible when graduate teaching assistants (GTAs) alone were providing feedback. In addition to providing the labor for this additional feedback point, there are numerous learning opportunities for the first-year students. First, students are able to (legitimately) see the work of other students. This provides them with a window on alternative ways to solve the problem. Second, students are asked to critically evaluate the work others – a high-level cognitive activity [6] that they rarely encounter in lower division courses. Third, students are engaged in the engineering community practice of peer review, an obligation of engineers in practice. Consider that IEEE members commit themselves "to the highest ethical and professional conduct and agree…"to seek, accept, and offer honest criticism of technical work [and] acknowledge and correct errors" [7].

These educational benefits are reinforced by Boud (2000), who posits that the focus of assessment as a whole must be rethought to promote lifelong learning skills [8]. Learning to provide and to respond to formative feedback given via both peer and self review are essential skills for success in a continuously working world that doesn't assign an end-of-project grade. Teaching students how to perform peer review and how to utilize constructive criticism for improvement is essential for their future.

Much of the recent research on MEAs has revolved around the development and refinement of an assessment and feedback rubric that is valid and can be reliably used by GTAs in formative and summative assessment [9-12]. It was felt that the rubric used in Fall 2008 was verging on adequately meeting these goals and thus attention could turn towards understanding how students use the rubric within the peer review process. The goal is to improve the impact that peer review has on the student team MEA solution quality while providing the student reviewers with some of the associated learning benefits of the peer review process.

In this paper, we present and discuss the results of a pair of instruments used in the peer review process implemented three times in a single semester in a large required first-year engineering problem solving course at Purdue University. We examine the results of a reflection instrument used to investigate how comfortable, capable and valued students felt, as reviewers, during their engagement in a peer review process that required the use of a three dimensioned rubric. We also examine the results of an instrument used by student teams, whose work was peer reviewed, to provide formative feedback to their peer reviewers on the quality of the reviews received.

BACKGROUND

I. Model-Eliciting Activities (MEAs)

MEAs are open-ended, client driven, team-oriented realistic engineering problems that are solved through iterative solution development phases. These are designed using six principles [13] modified for engineering contexts [14]. The solution of a

MEA requires that student teams use one or more mathematical or engineering concepts that are unspecified by the problem – students must make new sense of their existing knowledge and understandings to formulate a generalizable (share-able, re-usable, modifiable) mathematical model that can be used by the client to solve the given problem and similar problems. The solutions that student teams produce to a MEA are memos to the client that include a rationalized procedure for solving the problem with sample results.

Feedback and assessment on student work is done using the *MEA Feedback and Assessment Rubric* (*MEA Rubric*) [12] which is divided into three dimensions:

- Mathematical Model: Does the mathematical model adequately address the complexity of the problem?
- Re-usability & Modifiability: Can the client use the model on similar types of data and can the client modify the model for use in similar but different situations?
- Audience (Share-ability): Can the client reproduce the results using the test case data provided in the MEA?

Both the GTAs and students use this rubric to assess and provide feedback on student team work. Students are introduced to these dimensions prior to the first MEA that is implemented in the first-year engineering course and have access to it at all implementation stages.

II. MEAs & Peer Review

Much of the literature on peer review (or peer assessment) focuses on two issues: validity of peer assessment and students' perceptions of the process. Validity of peer assessment is outside the scope of this paper; however, Falchikov and Goldfinch (2000) conducted a meta-analysis on 48 quantitative peer assessment studies comparing peer and teacher marks [15].

Research on students' perceptions of the peer review process provides insight that can be grouped into pros (benefits and likes) and cons (drawbacks or dislikes). Hanrahan and Isaacs (2001) identified themes and dimensions underlying students' comments about the benefits and drawbacks students saw in the peer- and self-assessment experience embedded in a research essay assignment in a "third-year tertiary human psychology subject" [16]. Five themes could be classified as pros: "gained better understanding of marking", "productive", "read others' work", "develop empathy", and "motivation". Dimensions within these themes relating to peer assessment that are particularly relevant to the way in which peer review is conducted within the implementation of MEAs are as follows. Students gained a better understanding of the marking (rubric) - it reinforced the marking procedure and it enabled productive self-critique (helped improve one's own work). Students found peer assessment productive in that it provided more feedback, improved their own assignment prior to submission, and helped develop critical thinking. In reading other's work, students commented that they learned what others are doing and they benefited from seeing good and bad work. The peer assessment experience provided empathy for lecturers / tutors in regards to the marking task, and students felt motivated to impress peers. Similarly, Sitthiworachart and Joy (2003) found that 69% of first-year undergraduate students in computer science reported that they discovered mistakes in their own code while reviewing code written by their peers, and 80% of the students felt that seeing other students' work, both high- and low-quality work, was helpful for their learning [17]. Ballantyne, Hughs, and Mylonas (2002) reported that students "agreed that peer assessment was an awareness-raising exercise because it made them consider their own work more closely, highlighted what they needed to know in the subject, helped them make a realistic assessment of their own abilities, and provided them with skills that would be valuable in the future." [18].

Three of the themes identified by Hanrahan and Isaacs (2001) could be classified as cons: "difficult", "discomfort", and "problems with implementation" [16]. On the drawbacks side, students found it difficult to be objective and had no experience with marking and were not sure of standards. Also, students experienced discomfort having peers read their papers and critiquing the work of others, and they found that peers can be too critical. Further, the students found peer assessment to be time consuming, not taken seriously / doesn't count for marks, and too late to be useful. Students also commented that there should be feedback on their peer assessments. Ballantyne, Hughs, and Mylonas (2002) found that 40% of students felt that their peers would be unable to assess them fairly and did not feel confident in their own abilities to correctly evaluate their peers [18]. Moreira and Silva (2003) also reported concerns about fairness [19]. In a study of 48 computer science undergraduate and graduate students, 10% expressed concerns about having clear judging criteria, specifically with the fact that the individuals evaluating them may not all be applying the criteria in the same way. Liu, et el. (2001) showed similar results with 9% of the 143 third-year computer science students surveyed questioning the fairness of peer evaluations and having fellow students be in control of a significant portion of their overall grade [20].

Recommendations for successful implementation of peer assessment were often made in these earlier works. Many of these recommendations have been adopted in the MEA implementation of peer review. For instance, training on peer assessment is provided through the calibration exercises [16] and through instruction with the rubric [18]; the work under review is of sufficient complexity to "require critical analysis by student assessors" [18]; and anonymity of peer reviewers and reviewees is maintained to alleviate concerns about bias and unfairness in grading [18].

This study differs from earlier works in that it does not look at students' perceptions of the process in general but rather investigates how students perceive their ability to use the dimensions of an established rubric and how teams perceive the quality of the peer reviews they are receiving. This study is part of a larger project that is looking at the impact of pedagogical approaches on the quality of instructor and peer feedback and students' responses to this feedback. The research questions that guide this study are:

1. Are there differences in students' self-reported ability to apply the three dimensions of the *MEA Rubric* across the three MEAs implemented in a single semester?
2. Are there differences in teams' evaluations of the quality of students' peer reviews along the three dimensions of the *MEA Rubric* across the three MEAs implemented in a single semester?
3. How do students' perceptions of their ability to provide a peer review compare to the teams' perceptions of the quality of peer reviews that they are receiving?

THE STUDY

I. Setting

In Fall 2008, three MEAs, *Paper Plane Challenge*, *Just-In-Time Manufacturing*, *and Student Travel Modes*, were implemented in a required first-year engineering problem solving and computer tools course with an enrollment of approximately 1200 students. A brief description of each of these MEAs is provided by Zawojewski, Diefes-Dux, and Bowman [5].

The implementation of each MEA was conducted as follows over a three to four week period. Each MEA was launched in the laboratory setting which was facilitated by a team of four GTAs supported by undergraduate assistants. Student teams of 3-4 students developed the first draft of their memo. Following class, each GTA individually assessed the work of 14-15 student teams using the *MEA Rubric*. The remaining stages of the implementation of the MEA were completed in a series of homework assignments. Student teams used their GTA's feedback to revise their solutions, creating a second draft of their memo. This second draft entered a double-blind peer review process. In preparation of the peer review, students participated in a calibration exercise in which they practiced giving feedback on one prototypical piece of student work using a modified *MEA Rubric*, were provided an expert's review of that student work, and reflected on what they needed to do differently to improve their ability to give a peer review. For the peer review process, each student provided a peer review on one other team's solution to the MEA. Each student team received 1-5 (with 92% receiving 2-4) peer reviews which they used to revise their solutions, creating a final team solution. The final team solution was graded by the student team's GTA. Feedback provided at this stage was intended to help students perform better on the next MEA.

For the calibration exercise and the peer review, the students used a modified version of the *MEA Rubric*. This version of the rubric differs from the *MEA Rubric* that the GTAs use in that it requires the students to show that they have attempted to not only understand the team's procedure but also applied it to a specified data set. So, in addition to providing recommendations for improvement across the three dimensions, the students are asked to summarize in their own words the mathematics used, assumptions made, and limitations listed in the team's procedure. They were also asked to present results from applying the procedure to the specified data, or in lieu of that describe problems they had applying the procedure to generate results.

II. Participants

All students enrolled in the first-year engineering course in Fall 2008 were considered eligible for this study (N ~ 1200). Table 1 summarizes the number of students who completed a peer review (PR) and the two instruments used in this study.

TABLE 1
NUMBER OF STUDENTS COMPLETING OR RECEIVING PEER REVIEW ITEMS
(N ~ 1200)

Item(s) Completed or Received by Student[a]									
PRC	**PR**	**PCR**	**PCR - gfe**	**TCMCP**	**TCMCP - gfe**	**MEA 1**	**MEA 2**	**MEA 3**	**MEA1, 2& 3**
X	X					907	749	843	532
?	?	X				913	941	856	660
X	X	X				797	662	707	413
X	X		X			729	568	556	288
X	X			X		880	747	841	511
X	X				X	813	669	706	359
X	X		X		X	649	506	459	186

[a] PRC = peer review calibration, PR = peer review, PCR = *MEA Peer Critique Reflection*, TCMPC = *Team Critique of MEA Critique by Peers*, gfe = "good faith effort"

III. Methods

Team Critique of MEA Critique by Peers (TCMCP). Following the submission of the final team solution to a given MEA, the team provided each peer reviewer with an assessment of their review. This was done using an 8 item instrument called the *Team Critique of MEA Critique by Peers* (*TCMCP*). The first two items were:

1. This peer critique included an honest attempt to use our procedure to produce a solution to the test case. (Honest Attempt)
2. This peer critique was clearly written. (Clearly Written)

The remaining 6 items consisted of two items per dimension of the *MEA Rubric*.

3. The ___ portion of this peer critique was a fair and accurate assessment of our team's procedure. (Fair & Accurate)
4. The ___ portion of this peer critique resulted in our making substantive changes to our MEA solution. (Resulted in Changes)

The blanks were filled with the words Mathematical Model, Re-Usability/ Modifiability, or Audience (Share-ability), as appropriate. These items were assessed using a 5-point Likert scale where 1 = Strongly Disagree, 3 = Neither Agree not Disagree, and 5 = Strong Agree.

MEA Peer Critique Reflection. Immediately following the submission of the TCMCPs for a given MEA, each student was to complete the *MEA Peer Critique Reflection*. This instrument consisted of 15 items, five for each of the three dimensions of the *MEA Rubric*. These questions were:

1. I was comfortable assigning marks in the ___ category of the rubric. (Comfortable)
2. I was capable of proving a good critique in the ___ category of the rubric. (Capable)
3. I could identify problems related to the ___ category of the rubric. (Identify Problems)
4. I could provide written feedback related to the ___ category of the rubric. (Provide Written Feedback)
5. I thought the team I was reviewing would value the feedback I was providing in the ___ category of the rubric. (Valued By Reviewed Team)

The blanks were filled with the words Mathematical Model, Re-Usability/ Modifiability, or Audience (Share-ability), as appropriate. These items were assessed using the same 5-point Likert scale as described for the TCMCP.

This instrument also included 6 items that asked students to assess the impact of engagement in the peer review process on their ability to identify ways to improve their team's procedure along each rubric dimension and their understanding of what was being asked for by each rubric dimension. These items are not analyzed in this paper as the focus here is on students' comfort with and ability to apply the *MEA Rubric*. However, all 21 items are used to detect "good faith effort" responses as indicated below.

Data Collection. MEAs are conducted via a web-based interface connected to a database system. This overall system manages the organization and sequencing of the various stages of a MEA implementation; it also facilitates the interactions between individual students, their team, their GTA, and their peers. All student, GTA, and peer responses and interactions during an MEA are stored in the database for later review by the MEA research team. For this study, the data consisted of students' responses to the *MEA Peer Critique Reflection* and the team's responses to the *TCMCP*s for each of the three MEAs implemented in Fall 2008.

Data Analysis. An amount of cleaning of the data needed to be done. The first issue was that a number of students completed the *MEA Peer Critique Reflection* who had never completed a peer reflection (Table 1). Second there were issues of valid responses on the two instruments. Only those students who made a good-faith effort to complete *MEA Peer Critique Reflection* and received a good-faith effort TCMCP were kept in the study (designated gfe in Table 1). "Good-faith effort" here refers to not responding to all items on instruments in exactly the same way (e.g. all items are marked "agree"). Such responses were seen to be indicative of students rushing to complete the instrument simply to gain credit for having completed the assignment rather than adequate reflection. Only 186 students are included in this study (study population) because these were the only students who completed the peer calibration exercise, the peer review, and a good-faith effort MEA Peer Critique Reflection for all three MEAs implemented *and* received good-faith effort *TCMCP*s on their peer reviews for all three MEAs. Simple comparisons across MEAs were made for each item on the *TCMCP* and the *MEA Peer Critique Reflection* using t-tests and F-tests. To investigate whether this sub-population was representative of the larger course population, an item by item, MEA by MEA t-test comparison was made between this small sub-population and the one that had completed the peer calibration exercise, the peer review, and a good-faith effort MEA Peer Critique Reflection (Table 1, row 4, called course sample population type 1), and the one that completed the peer calibration exercise, the peer review, and a received a good-faith effort TCMCP (Table 1, row 6, called course sample population type 2).

RESULTS

Study Population vs. Course Sample Populations. Considering the *MEA Peer Critique Reflection* for each MEA (course sample population type 1) and the *TCMCP* for each MEA (course sample population type 2) independently, t-tests were performed on responses for each item between the study population and these course sample populations to investigate whether the study population was significantly different than the larger course population. There was no significant difference in the item means on the *MEA Peer Critique Reflection*. The means of only 2 *TCMCP* items (MEA 1 Audience (Share-ability) "Resulted in Changes" and MEA 3 "Honest Attempt") showed statistically significant differences at the $p < 0.05$ level. So, we are fairly confident that the study population is representative of the larger course populations.

Study Population. Item response means and standard deviations for both the *TCMCP* and *MEA Peer Critique Reflection* are shown in Table 2 for each MEA for the study population. P-value results of paired t-tests between item means on different MEAs are also shown in Table 2. Differences in variance between item responses to different MEAs are indicated (Table 2). There were no significant differences found in the variance between items on different MEAs for items on the *TCMCP*.

DISCUSSION

For the *MEA Peer Critique Reflection*, the means of the Mathematical Model items all peaked on MEA 2 and dropped back off for MEA 3. We speculate that this is a function of the nature of the problems being presented in the respective MEAs. MEA 2 often results in only a small set of unique approaches to the problem, potentially resulting in reviewers being more likely to feel able to provide feedback, as the solutions being reviewed are likely to be similar to the reviewers' own solution. Likewise, MEA 3 often results in a wide variety of unique solutions, potentially causing reviewers more difficulty in assimilating the solution they are reviewing

TABLE 2
TCMCP AND *MEA PEER CRITIQUE REFLECTION* ITEM RESULTS (N = 186)

		Mean (Standard Deviation)			Significance Level of Difference in Means (p-values)		
		MEA1	**MEA2**	**MEA3**	**MEA 1 to MEA 2**	**MEA 2 to MEA 3**	**MEA 1 to MEA 3**
TCMCP	**Honest Attempt**	3.91 (1.08)	3.96 (0.96)	3.75 (0.99)	0.637	0.037	0.100
TCMCP	**Clearly Written**	3.89 (0.88)	3.98 (0.83)	3.87 (0.89)	0.286	0.248	0.823
Mathematical Model							
Reflection	**Comfortable**	3.91 (0.71)	4.15 (0.48)	4.11 (0.62)	0.000^{**}	0.373^{**}	0.003
Reflection	**Capable**	3.92 (0.65)	4.00 (0.54)	3.98 (0.70)	0.112^{*}	0.716^{**}	0.372
Reflection	**Identify Problems**	3.99 (0.61)	4.14 (0.59)	4.06 (0.58)	0.003	0.187	0.158
Reflection	**Provide Written Feedback**	3.91 (0.58)	4.02 (0.55)	4.01 (0.60)	0.053	0.832	0.119
Reflection	**Valued By Reviewed Team**	3.79 (0.72)	3.93 (0.63)	3.92 (0.77)	0.020	0.863^{**}	0.074
TCMCP	**Fair & Accurate**	3.66 (0.92)	3.68 (1.00)	3.58 (1.04)	0.784	0.324	0.444
TCMCP	**Resulted in Changes**	3.13 (1.24)	2.99 (1.14)	2.97 (1.20)	0.202	0.865	0.196
Re-Usability / Modifiability							
Reflection	**Comfortable**	3.72 (0.72)	3.94 (0.64)	4.02 (0.61)	0.001	0.145	0.000^{*}
Reflection	**Capable**	3.71 (0.65)	3.80 (0.62)	4.00 (0.63)	0.113	0.001	0.000
Reflection	**Identify Problems**	3.80 (0.67)	3.84 (0.63)	4.04 (0.61)	0.407	0.001	0.000
Reflection	**Provide Written Feedback**	3.78 (0.62)	3.82 (0.57)	4.01 (0.59)	0.500	0.000	0.000
Reflection	**Valued By Reviewed Team**	3.69 (0.71)	3.83 (0.67)	3.86 (0.73)	0.022	0.666	0.014
TCMCP	**Fair & Accurate**	3.53 (0.95)	3.66 (0.96)	3.63 (1.01)	0.183	0.726	0.320
TCMCP	**Resulted in Changes**	2.98 (1.12)	2.95 (1.11)	2.96 (1.08)	0.828	0.948	0.874
Audience (Share-ability)							
Reflection	**Comfortable**	3.83 (0.69)	3.96 (0.62)	4.01 (0.65)	0.026	0.367	0.004
Reflection	**Capable**	3.82 (0.61)	3.91 (0.62)	4.00 (0.65)	0.078	0.123	0.003
Reflection	**Identify Problems**	3.88 (0.59)	3.88 (0.62)	4.05 (0.66)	0.922	0.002	0.005
Reflection	**Provide Written Feedback**	3.88 (0.57)	3.85 (0.62)	4.05 (0.63)	0.613	0.000	0.002
Reflection	**Valued By Reviewed Team**	3.73 (0.67)	3.81 (0.64)	3.82 (0.75)	0.148	0.930^{*}	0.195
TCMCP	**Fair & Accurate**	3.67 (0.95)	3.60 (1.06)	3.62 (0.95)	0.463	0.845	0.584
TCMCP	**Resulted in Changes**	3.24 (1.08)	2.90 (1.12)	3.06 (1.12)	0.002	0.148	0.102

* Differences in variance significant at $p < 0.05$ level
** Differences in variance significant at $p < 0.01$ level
Underlined values are significant at the $p < 0.05$ level

into their own mental model of what looks like a good solution.

While the Mathematical Model items peaked on MEA 2, the majority of the Re-Usability / Modifiability and Audience (Share-ability) item means increased across the three MEAs. We speculate that this is because, while the Mathematical Model items are highly problem dependent, the general properties the reviewers are looking for along the Re-Usability / Modifiability and Audience (Share-ability) dimensions do not change even though the mathematics applied and the problem context is changing. This may be translated to an increased feeling of comfort and ability to provide feedback along these two dimensions.

There were few significant differences in the TCMCP items means across MEAs. There was a significant decrease in the "Honest Attempt" item between MEA 2 and MEA 3; again, perhaps due to the greater number of unique solutions possible in MEA making it more difficult for students to provide a review. There was also a significant decrease in Audience (Share-ability) "Resulted in Change" between MEAs 1 and 2. This may be due to students not finding enough to provide feedback on in MEA 2 as there were fewer unique solutions.

The two item means that increased the most were Re-Usability / Modifiability "Comfortable" and "Capable". It is speculated that this is most likely because the ideas of Re-Usability / Modifiability are the most foreign to students at the beginning of the semester, specifically the idea of justifying their procedural steps and the importance of explaining why they included the procedural steps as they do. As students better understand these ideas, they are more comfortable and capable of providing feedback along this dimension.

Despite the generally high response means for the *MEA Peer Critique Reflection* items, scores on the *TCMCP* were typically lower. Perhaps indicating that students often felt the feedback they were receiving from their peers was not as helpful as the feedback they were providing (though this is difficult to assess as there is not strict one-to-one correspondence between then instrument items). Similarly, even though students thought that they were capable of identifying problems and providing written feedback, *TCMCP* results indicate a nearly neutral (neither agree nor disagree) response about whether feedback resulted in substantive changes to MEA solutions.

One issue that is not explicitly identified through the results in Table 1 is the idea that teams may be responding to the *TCMCP* items based on what they are expecting from the feedback and not necessarily the feedback they received. For example, one of the comments made in one *TCMCP* stated, "We thought this peer's comments didn't make much sense because he was critiquing our actual procedure rather than how it was written and whether or not it was rationalized, etc." While the critique did focus on the issues associated with the Mathematical Model the reviewer explicitly stated, "You did a good job rationalizing all of your steps." This is a situation where the team did not accept the feedback on the Mathematical Model because they were not expecting feedback on the Mathematical Model.

CONCLUSIONS

In this study, we found that there were differences in students' self-reported ability to apply the three dimensions of the *MEA Rubric* across the three MEAs implemented in a single semester. And there were a few differences in teams' evaluations of the quality of students' peer reviews along the three dimensions of the *MEA Rubric* across the three MEAs. We also found some indication that students' perceptions of their ability to provide a peer review exceeds the teams' perceptions of the quality of peer reviews they received. We speculated that the reasons for these differences are related to the MEA problems themselves and students' familiarity with certain dimensions of the *MEA Rubric*, and expectations for feedback, but these reasons are not definitive. The reasons most likely lie in the actual written feedback that students provided in the peer reviews, in the open-ended question asked at the end of the *MEA Peer Critique Reflection* ("What kinds of challenges did you encounter while providing feedback?), and in the written feedback teams provide to peers in the *TCMCP* (e.g. about what was helpful and challenging about the feedback received by a given peer and recommendations for how this peer could improve). Investigations into this data are underway.

ACKNOWLEDGMENT

This work was made possible by a grant from the National Science Foundation (EEC 0835873). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] National Academy of Engineering. *The engineering of 2020: Visions of engineering in the new century,* 2004, Washington, DC: The National Academies Press.

[2] ABET Engineering Accreditation Commission. "Criteria for Accrediting Programs in Engineering". 2008, ABET Inc.: Baltimore, MD, Retrieved March 22, 2009 from http://www.abet.org/forms.shtml#For_Engineering_Programs_Only

[3] Diefes-Dux, H.A., Hjalmarson, M., Zawojewski, J., & Bowman, K. "Quantifying aluminum crystal size part 1: the model-eliciting activity," *Journal of STEM Education: Innovations and Research*, vol. 7, No. 1&2, 2006, pp. 51-63.

[4] Diefes-Dux, H.A., Moore, T., Zawojewski, J., Imbrie, P.K., & Follman, D. "A framework for posing open-ended engineering problems: model eliciting activities", *Paper presented at the 34th ASEE/IEEE Frontiers in Education Conference*, October 2004, Savannah, GA.

[5] Zawojewski, J. S., Diefes-Dux, H., & Bowman, K. (Eds.) *Models and modeling in engineering education: designing experiences for all students,* 2008, Rotterdam, the Netherlands: Sense Publishers.

[6] Bloom, B. *Taxonomy of educational objectives: the classification of educational goals, by a committee of college and university examiners,* 1956, New York: D. McKay.

[7] Institute of Electrical and Electronics Engineers. "IEEE Policies, Section 7 - Professional Activities (Part A - IEEE Policies), 7.8 Code of Ethics", Retrieved March 22, 2009, from http://www.ieee.org/web/aboutus/whatis/policies/p7-8.html

[8] Boud, D. “Sustainable Assessment: Rethinking Assessment for the Learning Society”, *Studies in Continuing Education,* Vol. 22, No. 2, 2000, pp. 151-167.
[9] Cardella, M.E., Diefes-Dux, H.A., Verleger, M.A., & Oliver, A. “Insights into the process of providing feedback to students on open-ended problems,” *Proceedings of the American Society for Engineering Education (ASEE) Annual Conference*, June 2009, Austin, TX.
[10] Diefes-Dux, H.A. & Imbrie, P.K. "Chapter 4: Modeling Activities in a First-Year Engineering Course" In Zawojewski, J. S., Diefes-Dux, H., & Bowman, K. (Eds.) *Models and modeling in engineering education: designing experiences for all students,* 2008, Rotterdam, the Netherlands: Sense Publishers, pp. 37-92.
[11] Verleger, M. & Diefes-Dux, H. “Impact of feedback and revision on student team solutions to model-eliciting activities”, *Proceedings of the American Society for Engineering Education (ASEE) Annual Conference*, June 2008, Pittsburgh, PA.
[12] Verleger, M. & Diefes-Dux, H. “Multi-dimensional tool for assessing student team solutions to model-eliciting activities”, *Proceedings of the American Society for Engineering Education (ASEE) Annual Conference*, June 2009, Austin, TX, 2009.
[13] Lesh, R., Hoover, M., Hole, B., Kelly, A., & Post, T., “Principles for developing thought-revealing activities for students and teachers”, *Handbook of Research Design in Mathematics and Science Education*, 2000, Mahwah, NJ: Lawrence Erlbaum, pp. 591-645.
[14] Diefes-Dux. H.A, Hjalmarson, M., Miller, T., & Lesh, R. "Chapter 2: Model-Eliciting Activities for Engineering Education," In Zawojewski, J. S., Diefes-Dux, H., & Bowman, K. (Eds.) *Models and modeling in engineering education: designing experiences for all students,* 2008, Rotterdam, the Netherlands: Sense Publishers, pp. 17-35.
[15] Falchikov, N. & Goldfinch, J. “Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks”, *Review of Educational* Research, Vo. 70, No. 3, 2000, pp. 287-322.
[16] Hanrahan, S.J. & Isaacs, G. “Assessing self-and peer-assessment: the students’ views”, *Higher Education Research &* Development, Vol. 20, No. 1, 2001, pp. 53-70.
[17] Sitthiworachart, J., & Joy, M. “Deepening computer programming skills by using web-based peer assessment”, *Paper presented at the 4th Annual Conference of the LTSN Centre for Information and Computer Sciences*, 2003, NUI Galway, Ireland.
[18] Ballantyne, R., Hughes, K., & Mylonas, A. “Developing procedures for implementing peer assessment in large classes using an action research process”, *Assessment and Evaluation in Higher Education,* Vol. 27, No. 5, 2002, pp. 427-441.
[19] Moreira, D. A., & Silva, E. Q. “A method to increase student interaction using student groups and peer review over the internet”, *Education and Information Technologies,* Vol. 8, No. 1, 2003, pp. 47-54.
[20] Liu, E. Z., Lin, S. S. J., Chiu, C., & Yuan, S. “Web-based peer review: the learner as both adapter and reviewer”, *IEEE Transactions on Education,* Vol. 44, No. 3, 2001, pp. 246-251.

AUTHOR INFORMATION

Heidi A. Diefes-Dux, Associate Professor of Engineering Education, School of Engineering Education, Purdue University, hdiefes@purdue.edu.

Matthew A. Verleger, Doctoral Candidate, School of Engineering Education, Purdue University, mverleg1@purdue.edu.