2008

Impact Of Feedback And Revision On Student Team Solutions To Model Eliciting Activities

Matthew Verleger
Purdue University

Heidi Diefes-Dux
Purdue University - Main Campus, hdiefes@purdue.edu

AC 2008-1126: IMPACT OF FEEDBACK AND REVISION ON STUDENT TEAM SOLUTIONS TO MODEL-ELICITING ACTIVITIES

Matthew Verleger, Purdue University

Heidi Diefes-Dux, Purdue University

Impact of Feedback and Revision on Student Team Solutions to Model-Eliciting Activities

Abstract

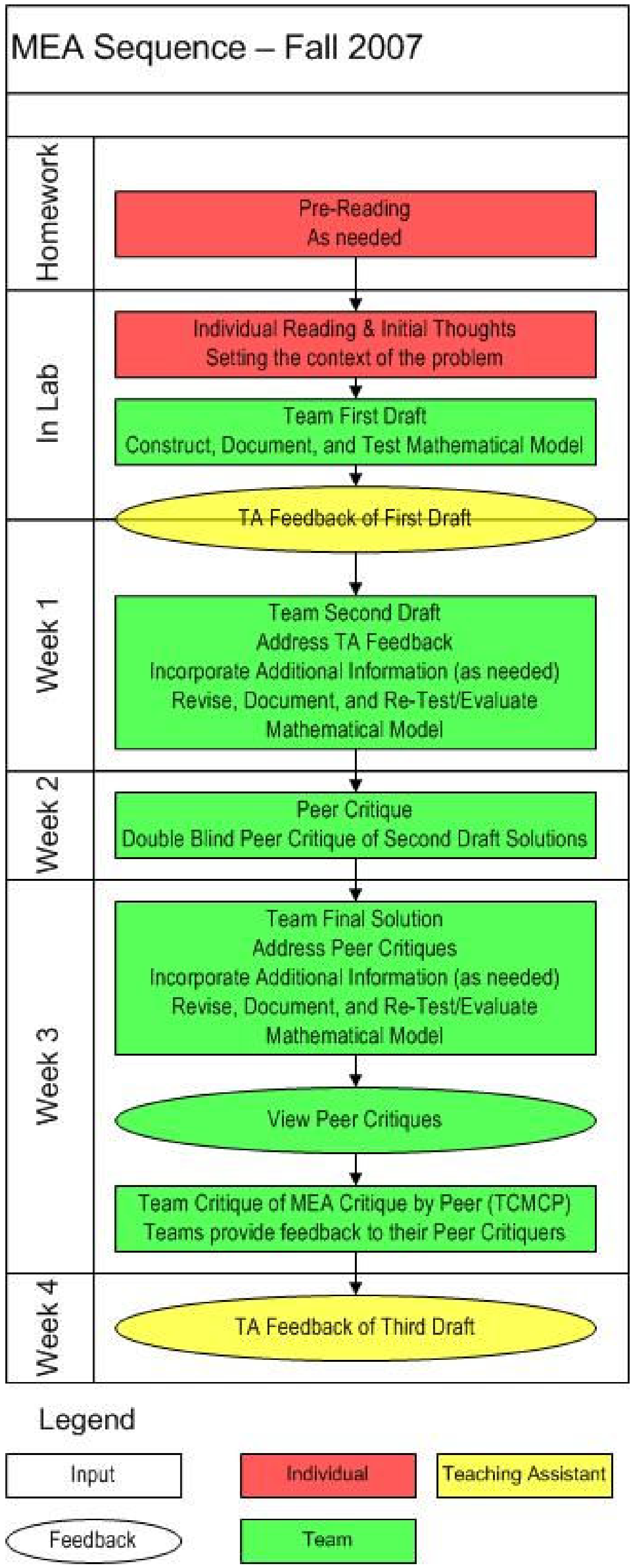
Helping first-year engineering students to embrace the iterative and open-ended nature of engineering problem solving is a challenge when their prior learning experiences have focused heavily on achieving a correct answer in a single attempt. In this paper, the authors will present a case study of student work from the Fall 2007 implementation of Model-Eliciting Activities (MEAs) to demonstrate the impact of the iterative process of feedback and revision on the quality of student products. They will also discuss some of the future research questions resulting from the iterative process used with MEAs.

Introduction

Model Eliciting Activities (MEAs) are realistic, open-ended, client driven problems designed to foster students' mathematical modeling abilities. Built around the models and modeling perspective established by Lesh and Doerr [1], MEAs are carefully developed around six guiding principles. The development process is described in greater detail by Moore and Diefes-Dux [2]. The product students generate from an MEA is a memo directed to the client describing a process (procedure) for solving the client's problem that is sharable, repeatable, and generalizable. Sharable solutions are ones with clearly articulated steps that the client can easily understand. Repeatable solutions are those where the output of the procedure is the same regardless of the individual implementing the procedure. Generalizable solutions are applicable to other similar situations.

Engineering-based MEAs were introduced into Purdue's First-Year Engineering (FYE) course, ENGR 126, Engineering Problem Solving and Computer Tools, as part of a NSF-HRD Gender Equity in STEM grant titled "Small Group Mathematical Modeling (SGMM) Approaches to Improved Gender Equity in Engineering" (NSF HRD 0120794). The use of MEAs in this required first-year engineering course was investigated as a means of keeping underrepresented students, especially females, as interested and persistent in engineering as their counterparts [3, 4].

Figure 1 – Fall 2007 MEA Sequence



From 2002 to 2004, four MEAs were implemented each semester. Each was completed in the span of only one week and each was used to launch assignments around more traditional engineering content. During a single 110 minute lab section, the MEA was introduced. During that lab period, students individually read the MEA problem statement and entered into a discussion with their teammates using an online discussion board (in 2002) or worked with their team to generated a memo in which they began to articulate their ideas about the problem context (in 2003-2004). A complete history of MEAs and their implementation in the Purdue's FYE program can be found in Diefes-Dux and Imbrie [5].

In Fall 2007, the MEA implementation was expanded to a multi-stage sequence spanning 4 weeks with multiple iterations of feedback and revision. A flowchart of the process can be seen in Figure 1. An MEA may start with a pre-reading exercise designed to introduce any background information (e.g. technical terminology) the students will need to understand the context of the MEA. This pre-reading is assigned as homework in the week prior to the lab containing the MEA. In lab, students work through the sequence to produce a first draft of their procedure. First, they are given an individual warm-up activity designed to introduce them to the problem context. This consists of an advanced organizer detailing the client and their problem followed by a set of free-response questions about who the client is, what the client needs, and issues to be considered when producing a solution. After all team members have responded to the individual questions, the team comes together to develop a solution to the client's problem. The deliverable at the end of the lab period is a first draft of a memo to the client detailing the solution to the problem.

Following the lab, the teaching assistant provides the students with feedback. Feedback is organized along three dimensions: mathematical model, re-usability/share-ability, and audience. The mathematical model dimension is focused on the degree to which the student team has addressed the complexity of the problem, the utilization of the sample data or test cases provided by the client, and the presence of rationales for the steps in the procedure. The re-usability/share-ability dimension is focused on the degree to which the procedure is adaptable to scenarios not explicitly given in the problem statement. The audience dimension is focused the delivery of results using the procedure and the degree to which the client can easily use the procedure and repeat the results. These three dimensions are presented in a rubric and used throughout the entire sequence for feedback and assessment. This rubric is provided to the students before they begin writing the first draft. This rubric is currently being examined for reliability and validity.

After students receive feedback on their first draft from the teaching assistant, they make revisions to their procedure and submit a second draft that enters a calibrated double-blind peer review. Each team receives three or four critiques. Teams then utilize these critiques to finalize their procedure which is submitted for grading to the teaching assistant.

In the five years since MEAs were first implemented in the first-year engineering course, there have been numerous changes to nearly every facet of the of the MEA implementation, all designed to help students produce higher quality mathematical models. The greatest change occurred in Fall 2005 with the addition of both multiple iterations of revision and the double-blind peer review. This change provided students with an opportunity to iterate on the development of their procedure, modifying their procedure based on feedback from multiple sources. This paper will present one first-year engineering student team's work from the Fall 2007 implementation of MEAs. This case will be used to demonstrate the impact of the iterative process of feedback and revision on the quality of student products. The paper will then discuss lessons learned and future research directions found as a result of preparing this case study.

Methods

Setting and Participants

ENGR 126, Problem Solving and Computer Tools, is an introductory service course covering a wide array of topics, including general problem solving strategies and MATLAB programming. During a fall semester, the course contains primarily first-semester engineering students interested in all of the engineering disciplines offered at Purdue. It is broken into four large lectures taught by faculty members each containing approximately 400 students that meet for 50 minutes twice per week. Students also attend one 110-minute lab section once per week taught by one of the 19 graduate teaching assistants. Each lab division contains a maximum of 32 students. During the Fall 2007 semester, 1512 students were placed onto 402 teams.

Early in the third week of the semester, students are introduced to MEAs in lecture and work through a simple problem in ad-hoc groups. Issues of open-endedness and meeting the client's needs are stressed to students. During lab in the third week, students undertake their first MEA. The second MEA is typically done during week six, and the third during week 11.

Teaching assistants are trained throughout the year in how to properly guide students through the MEA sequence. At the beginning of the year, they are introduced to MEAs and the theory driving their use. They are also given the opportunity to work through the first MEA the students will be completing. Before each of the three MEAs, the teaching assistants are also given a training session. This session is focused on grading and providing adequate and appropriate feedback to the students. As part of this training, teaching assistants grade five samples of student work. These grades are reviewed by course administrators and feedback is given to the teaching assistant about their grading.

Model Eliciting Activity

The model-eliciting activity that will be discussed throughout this paper is called Nano Roughness. It should be noted that this was the third MEA of the semester and therefore the expectations had been clearly established. An abbreviated version of the first draft instructions of the Nano Roughness MEA is shown in Table 2. The complete version can be found in (Zawojeski, Diefes-Dux, and Bowman, in review). Prior to the lab, students were given a pre-reading activity about Atomic Force Microscopes (AFM) and the images they produce. In the lab setting, students were given AFM images of gold samples (Sample B is shown below in Figure 2) to create and test their procedures for quantifying roughness.

Table 2 – Nano Roughness MEA

Abbreviated Problem Statement

Interoffice Memo:	Liguore Labs
To:	Nanosurface Engineering Team
From:	Kerry Prior, Vice President of Research
RE:	Surface Roughness

Liguore Labs is very interested in the innovations of biomedical science. Recently a physicist from University of Alabama, Birmingham accidentally produced smooth diamond. The array of diamond created was smooth and adhered very easily to metal. Because diamond is durable, it makes a very good candidate for coating artificial hip replacements. The current coatings wear down or loosen from constant use after about 10 years, which could mean more surgery for the recipient. The diamond coating is projected to last around 40 years which would improve the comfort and health of the patient.

Liguore Laboratories would like to expand our product line to include diamond coatings for hip joints. The research laboratory is working on replicating the smooth diamonds. In order for the scientists to know if their process is working, they will need a procedure that will measure the roughness of the diamond in nanoscale.

Since we have experience with gold coatings and have many images available, we can use these images to develop our procedure. Attached are three atomic force microscope (AFM) pictures of the gold we have been using in our artery stent research. Your team needs to create a procedure using these images to measure (or quantify) the roughness of the gold at the nanoscale and generate a description of how the process would work by applying the procedure to the three AFM pictures of gold. With this procedure in place, our research team will be able to measure the roughness of the diamond samples as they are produced.

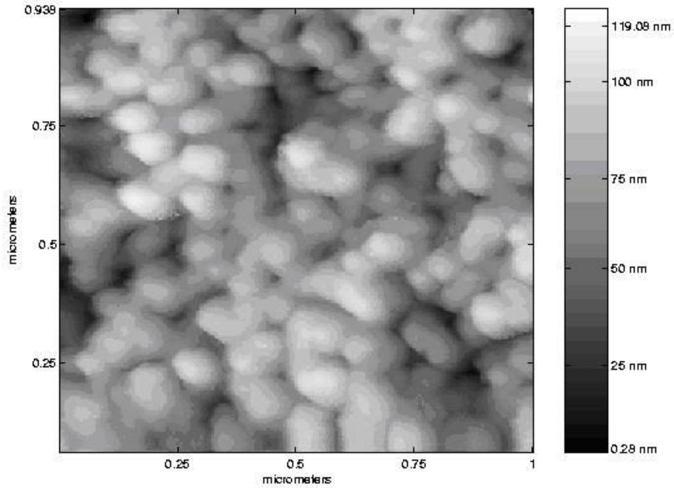
Please reply in a memo with the following information:

- The series of steps that can be used to measure roughness of the nanoscale material using the AFM images.
- A description of how the procedure would work by applying it to gold samples A, B, and C that are attached to this memo.
- A description of what information your team would need in order to improve your procedure to quantify the roughness of the gold.

Thank you for your team's efforts in this endeavor.

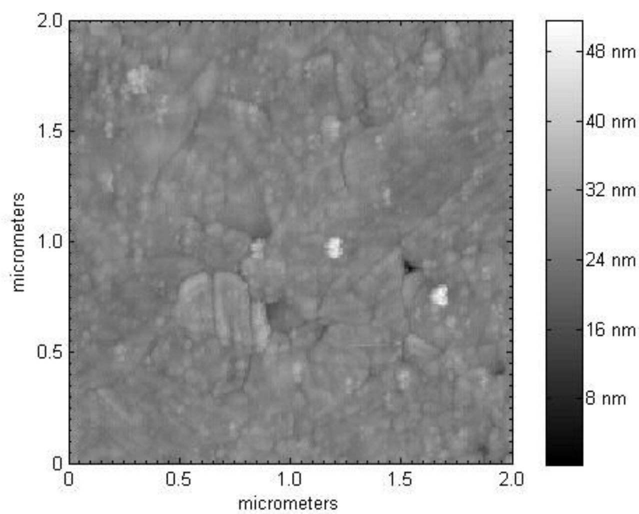
Kerry Prior, VP Research

Figure 2 - Sample B (Courtesy of the Reifenberger Nanoscale Physics Lab at Purdue University)



No changes were asked of students for their second draft however the third draft included additional test cases. These cases were all selected to visually look different than the original samples and have surfaces with lower ranges of peak/valley heights and lower standard deviations of surface. One was selected with nearly all the same height to challenge their process. This can be seen in Figure 3.

Figure 3 - Sample F (Courtesy of the Reifenberger Nanoscale Physics Lab at Purdue University)



High Quality Solution Characteristics

Solutions of the highest quality should have addressed each of the following issues:

- A sampling method that meets the client's needs for a quick and easy-to-use method but also provides adequate data to perform any subsequent statistical analysis (for instance, mean or standard deviation).

- One or more statistical measures (e.g. maximum, range, standard deviation) of height (surface elevation) are used to quantify the roughness of the image. The measure(s) selected are aligned with a clearly stated definition of roughness.
- Frequency, 2-d size, and/or distances between significant features in the images is addressed. Procedures that address these issues must also use a measure related to height to quantify roughness. This is necessary as measures of frequency, 2-d size, and distance between features alone cannot define roughness. Either the procedure accounts for these issues or a rationale is provided for not considering these issues within the procedure.
- The fact that AFM images can be of different sizes is addressed. Either the procedure accounts for image size or there a rationale is provided for not considering the size within the procedure.
- Critical steps that need justification / rationale:
  - Sampling method
  - Each measure contributing to the quantification of roughness
  - Adjustments for size of image
- Student teams should state that the procedure is designed to be used on AFM images with an x-y scale on the image and an associated colorbar indicating the height of the surface.
- Students should indicate limitations of their procedure. Limitations may arise if the team hard-codes values in their procedure (e.g. sampling method).
- The client requires a quick and easy-to-use procedure. If this has not been delivered, the solution is not high quality work. If you, as a representative of the client, cannot replicate or generate results, the solution is not high quality work.
- Results of applying the procedure must be free of unit problems or orders of magnitude issues.

Author Qualifications

The first author was a teaching assistant for ENGR 126 for 7 semesters between August 2003 and December 2006. In that time, he graded approximately 500 MEAs. In the summer of 2005, he wrote an MEA which has since been used three times in ENGR126 and has been discussed in a follow-up course offered by the computer science department. Additionally, he has helped lead two workshops and written a conference paper on the teaching assistant experience of using MEAs.

The second author was the course coordinator for ENGR126 from August 1998 to December 2006. In her role as course coordinator, she overhauled the curriculum to focus on solving more realistic engineering problems. Part of that overhaul included the development and incorporation of MEAs as a standard component of the curriculum. She has authored numerous papers on the subject, obtained funding from multiple sources to continue research into their value, and continually pushed MEAs in new directions.

Selection of the Case

Approximately one quarter of the 402 of student team responses from Fall 2007 were reviewed to find a case that exemplified the improvement properties desired for this discussion, namely to demonstrate the continual improvement and incorporation to feedback from both the teaching assistant and peer reviews. The case selected is atypical in that it results in a procedure that is of much higher quality than those generally found. It was selected because the changes seen across the three drafts were readily attributable to comments found in the feedback.

Case Study #28-4

Table 3 presents the three drafts and feedback for the procedure given by a single team. While the feedback includes both comment sections as well as Likert items, only specific comment sections related to recommendations for improvement are included. For the peer feedback, the team received feedback from four peers. Of the four, three provided detailed comments while the fourth provided feedback of only minimal value. The comments from the fourth individual have been excluded.

Draft 1 TO: Kerry Prior, VP of Research, Liguore Laboratories FROM: Team 4 RE: Nanoscale Roughness Heuristics You've asked us to create a procedure with which our scientists can quickly and easily quantify roughness in the lab. The following procedure uses the sharpness of the surface bumps, and the percentage of the surface area covered by significant surface bumpsto calculate an approximate but quantified value for the "roughness" of a surface. The only input this procedure can use is a topographical image of a few square microns of the surface. The surfaces evaluated will discrete bumps which are large enough to measure, given the resolution of the images. To evaluate the roughness of the surface, first make a general visual evaluation of what constitutes a significant bump. There may be smaller bumps as well, but these are insignificant and do not affect roughness. Choose a typical bump, and visually approximate the width and height of the bump. Divide the height of the bump by the width of the bump to calculate the "sharpness" of a typical bump. Next, visually approximate the percentage of the area covered by significant bumps. The inverse of this percentage (one divided by the percentage) is the "sparseness" of the surface. Multiply the sharpness of a typical bump by the sparseness of the bumps to calculate the "roughness" of the surface.
Comments from the Teaching Assistant about Draft 1 ***Mathematical Model*** – First, you did good job defining roughness although it is sort of vague. Now, you need to try and use the definition of roughness in your memo more thoroughly. Sharpness of the image is hard to constantly find. You need to provide more guidance to the client because as of now, it is very subjective to what is light and dark and anywhere in between. Remember this needs to be quick and easy to use so make sure everything is clear and concise and explained through rationales. Are you using the "colorbar" here and how did you account for the varying size of it? Providing some rationales and guidelines for defining the bumps would be

good.

When you say "Choose a typical bump, and visually approximate the width and height of the bump" how would you more consistently quantify this? What about the standard deviation of the height peaks? Low peaks? Average peaks? You just mention your definition, now try to more directly apply and use it. Remember, here you are not comparing the images, you are using them to come up with a way to define what means too rough or ok. Also, providing clear rationales for what you are doing is a must too. So, in your procedure, what would be rough (a number or ratio or range)? What would not be rough (a number or ratio or range)?

I am kind of confused about your sampling method. As of now, you just expect the client to pick bumps? Could you please provide a more definite idea as to when to count a bump and not? Remember, this needs to be a re-useable procedure. Also, don't assume that ever image is at square (which is not the case for sample B), how do you account for Non-square images? Smaller sizes? Larger sizes? Try to give a more detailed explain for how you got the numbers you did and the rationales behind it.

Does the frequency, size, and distance between features on the image matter? This issue needs to be addressed in your procedure and / or rationale.

Re-Usability/Share-ability – No assumptions are provided. Need to provide the client with clear information about the necessary conditions for this procedure to be used. At a minimum, assumptions need to be built around the fact that only a hardcopy of the image is available. These assumptions should articulate what these images must entail.

Audience – First, your memo needs to have results in it and they need to include the RIGHT amount of significant figures and all answers need units.

The procedure does not meet the client's needs if there are images for which the procedure cannot be applied. By quantifying what rough is and giving a more constant re-useable sampling method, then the client will have a more easier time replicating your results (if you had some). Have you thought about other possible statistical methods than the difference and mean? What about the std dev to relate the height points?

You need to make this quick and easy to use, so make sure you clearly define all rationales and assumptions and using the memo outline provided in the lab would help your team hit all of the required points that need to be put in the memo. Also, don't forget to sign your memo.

Draft 2

TO: Kerry Prior, VP of Research, Liguore
FROM: Team 4
Dear Kelly Prior of Ligoure labs,

You have asked our team to revise our original procedure of quantifying the roughness of nanoscale materials using hardcopies of AFM images. We have revised our procedure to better meet the needs of your researches and have made it quick and easy to use. The re-usable and share-able procedure we have created is designed to quantify the roughness of the diamond coating samples as they are produced. Our procedure requires hardcopies of AFM images that have an "x" and "y" axis in micrometers. Each

AFM image must also have a depth intensity scale (colorbar) in nanometers.

Here is the procedure in steps:

1. Divide the AFM image into a 4 x 1 grid of x and y.
2. Moving along the row from left to right at a y-axial midpoint, determine how many times the surface is concave up and for how long. There will be concave down interval when the "colorbar" indicates, under the certain interval, the depth is under 45% of the maximum and then proceeds to go above 45% after a certain distance.
3. Calculate the sum of how many times the image becomes concave up and the sum of the distance it concave up.
4. Repeat steps 1-3 for all rows.
5. .20 < (Concave up distance / length of x) < .80 , then the surface is rough. Also, at .5 the surface has optimal roughness.
6. To determine the percent roughness of surface: Concave up distance / length of x)*100
7. A concavity change means that there is a bump on the surface, therefore:
(concavity changes + 1)/1 unit of length = bumps/unit of length
8. After the roughness has been computed, calculate the roughness of the entire surface by averaging the roughness of each row.

Results:
Sample B

Row 1 : Concavity changes = 5	Bumps = 6	Concavity Distance= .21
Row2: Concavity changes= 4	Bumps= 5	Concavity Distance= .24
Row3: Concavity changes=5	Bumps: 6	Concavity Distance= .28
Row4: Concavity changes= 5	Bumps: 6	Concavity Distance= .22

Comments from the peer reviewers about Draft 2

Mathematical Model – This procedure seems very difficult to use when you have an image with many different "bumps" on the surface. Concavity and distance of that concavity are very hard to measure when an image has so many small dots packed together. I had trouble determining where the "bumps" were because the step describing finding whether or not a part was concave or not was too vague.

You need to explain why this method defines roughness. Why do you use the formulas that you use. Where do they come from? You need to explain why you chose 45% as well.

While this method is relatively complex, the situation calls for it. There are few ways to perfom such a subjective task with precision.

Overall, very good procedure. Perhaps explaining the procedure in greater detail would ensure comprehension.

I would better explain to the user what you mean when you are talking about using concave up. Also explain to them better how to calculate that distance that it is concave up.

You need to better explain what you mean by roughness. There isn't a very good explanation of what your definition of roughness is anywhere in your memo.

Re-Usability/Share-ability – You need to include some type of assumptions. You may want to assume that this sample image represents the actual surface of the entire sample it was taken from.

Explain what an AFM image is and more clearly explain how to obtain precise data from it.

Provide more assumptions about the scale that they give you with the sample. Also provide assumptions about the type of sample that they give you. Are you assuming that the sample they give you will always be square or rectangular? Are you assuming that sometimes they might give you a circular sample?

Audience – The procedure being broken up into steps makes it easier for a user to follow. The steps about calculating concavity are a little vague and need more explanation.

Very reusable already. Wording of sentences could be improved to enhance readability and fluidity, thus bettering the client comprehension. Examples could be organized better.

I think that you should simplify your method a little so that the user will better understand by what you mean. There are some people who wouldn't understand the method that you used.

Draft 3

TO: Kerry Prior, VP of Research, Liguore
FROM: Team 4
Dear Kelly Prior of Ligoure labs,

You have asked our team to revise our original procedure of quantifying the roughness of nanoscale materials using hardcopies of AFM images. We have revised our procedure to better meet the needs of your researches and have made it quick and easy to use. The re-usable and share-able procedure we have created is designed to quantify the roughness of the nanoscale material samples as they are produced.

Assumptions: Our procedure requires hardcopies of square or rectangular AFM images of nanoscale materials that have an “x” and “y” axis in micrometers. Each AFM image must be in black and white and also have a depth intensity scale (colorbar) in nanometers. The depth intensity bar determines how deep or elevated a certain point on the surface is.

Here is the procedure in steps:

1. Divide the AFM image into a grid of x and y. Divide the AFM image into rows. This organizes the image for easier testing. More rows equal more accurate results.
2. When you run your finger over a bump, at first you feel smoothness, then your finger starts to become elevated because of the increasing curve of the bump. At the maximum of the bump’s curve, the curve then begins to decrease and then you feel smoothness again. Something is smooth when there are very few bumps, or a lot of them (the bumps are so close together, it would be very hard to feel the difference). In this step, we are finding out how long the surface dips under 45% of the maximum peak (determined from the colorbar) for each row.

So, in order to do this:

Moving along a row from left to right at a y-axial midpoint (halfway between the first row and the beginning of the second row), determine how long the surface is concave up. There will be concave up interval when the "colorbar" indicates, under the certain "x" interval, the depth is under 45% of the maximum and then proceeds to go above 45% after a certain distance.

3. Calculate the sum of the distances it's concave up. We are doing this because if the image is concave only for a very short period , then the image is relatively smooth (same for having a ton of concave up intervals).

4. Repeat steps 1-3 for all rows.

5. .30 < (Sum of Concave up distance / length of x) < .70 , then the surface is rough. Also, at .5 the surface has optimal roughness. .5 is optimal roughness because it is just enough so that there are not too many concave up intervals (so that you couldn't tell the difference if you ran your finger over it) but also enough to feel the roughness. If the amount of distance along the surface is in between 30% and 70%, then the surface can be called rough. The intensity of the roughness is determined how close the percent from the above equation is to .5.

6. To determine the percent roughness of surface: Concave up distance / length of x)*100

7. After the roughness has been computed, calculate the roughness of the entire surface by averaging the roughness of each row.

Results:

Sample A: Overall Roughness = 32.5%

Row 1 :	Concavity Distance= 2.4	roughness= 40%
Row2:	Concavity Distance= 2.2	roughness= 37%
Row3:	Concavity Distance= 2.8	roughness= 30%
Row4:	Concavity Distance= 1.5	roughness= 25%
Row5:	Concavity Distance= 2.4	roughness= 40%
Row6:	Concavity Distance = 1.4	roughness = 23%

Sample B: Overall Roughness = 23.75%

Row 1 :	Concavity Distance= .21	roughness= 21%
Row2:	Concavity Distance= .24	roughness= 24%
Row3:	Concavity Distance= .28	roughness= 28%
Row4:	Concavity Distance= .22	roughness= 22%

Sample C: Overall Roughness = 21.25%

Row 1 :	Concavity Distance= .5	roughness= 25%
Row2:	Concavity Distance= .3	roughness= 15%
Row3:	Concavity Distance= .2	roughness= 10%
Row4:	Concavity Distance= .7	roughness= 35%

Sample D: Overall Roughness = 53.33%

Row 1 :	Concavity Distance= 1.9	roughness= 63%
Row2:	Concavity Distance= 1.2	roughness= 40%
Row3:	Concavity Distance= 1.7	roughness= 57%

Sample E: Overall Roughness = 23.33%

Row 1 :	Concavity Distance= .5	roughness= 17%
Row2:	Concavity Distance= .9	roughness= 30%
Row3:	Concavity Distance= .7	roughness= 23%

Sample F: Overall Roughness = 8.13%

Row 1 :	Concavity Distance= .2	roughness= 10%
Row2:	Concavity Distance= .1	roughness= 5%
Row3:	Concavity Distance= .05	roughness= 2.5%
Row4:	Concavity Distance= .3	roughness= 15%

Discussion

The progression across the three drafts can be most easily examined by reviewing each of the three rubric dimensions in light of the qualities of a high quality solution. The progression for each of the properties of high quality solutions is tracked in Table 3. A detailed discussion of the progression follows.

Table 3 –Properties of High Quality MEAs

		Draft 1	Draft 2	Draft 3
Mathematical Model	A sampling method	S	Y	Y
	One or more statistical measures	Y	Y	Y
	Frequency, 2-d size, and/or distances between significant features in the images is addressed		Y	Y
	AFM images can be of different sizes is addressed	Y	Y	Y
	Justifications / rationales			Y
Re-usability Share-ability	States that it is designed to be used on AFM images with an x-y scale on the image and an associated colorbar indicating the height of the surface		Y	Y
	Indicates limitations of their procedure		S	Y
Audience	Quick and easy-to-use procedure	Y	Y	Y
	Results of applying the procedure		S	Y
S = Somewhat, Y = Yes				

Mathematical Model - In their first draft, the team attempted to quantify the roughness based partly on a visual approximation. While their approach still contained quantified elements, it lacked the reproducibility necessary in a good model. Terms like "visually approximate" and "general visual evaluation" can easily be interpreted differently by different individuals. At the core level, this is an issue of sampling and feature properties. The team did not have a consistent sampling mechanism and therefore their method did not adequately account for the properties of the features of the image. In the comments, the teaching assistant noted, "I am kind of confused about your sampling method. As of now, you just expect the client to pick bumps? Could you please provide a more definite idea as to when to count a bump and not?" The teaching assistant also noted that the memo lacked appropriate rationales for the various steps. While they describe how to calculate the "roughness", they don't explain why each step produces a value appropriate to the answer.

In response to the feedback from the teaching assistant, the team drastically overhauled their model for the second draft. Recognizing that their original approach could be interpreted in multiple ways, they developed a mechanism for measuring the number of "bumps" found along a line. Repeating this process for four lines and taking an average yields an overall roughness metric for the entire image. Collectively, this resolved the issues of sampling and feature properties.

Despite resolving the concerns by the teaching assistant about concreteness, draft 2 still lacked sufficient rationales. Only one of their four peer reviewers felt that there was sufficient rationale for each of the steps, and that reviewer still wanted more details. In their third draft, the team maintained the same basic mathematical model but included more rationale for their steps. In response to a question about what changes they made to their procedure in the third draft, the team responded, "We added a lot more rationales and assumptions. That was the only drawback to our old draft."

Re-usability/Share-ability – One the biggest elements associated with the re-usability/share-ability dimension is the description of the assumptions imbedded/underlying their procedure as well as the explicit limitations of the procedure. As with any good engineering solution, it is important to know the constraints of when the solution applies to a problem. The first draft included no assumptions, and the teaching assistant was quick to comment on this. "No assumptions are provided. Need to provide the client with clear information about the necessary conditions for this procedure to be used. At a minimum, assumptions need to be built around the fact that only a hardcopy of the image is available. These assumptions should articulate what these images must entail." Reactionary to this, the team tried to be more explicit about their assumptions in the second draft. Unfortunately, they did not provide any assumptions beyond the realm of those introduced by the teaching assistant. The prod given to them was not enough to get them to analyze the limitations of their procedure. It took all four peer reviewers commenting on a lack of assumptions to begin to push the team towards exploring their assumptions. While they are weak assumptions, the inclusion of an assumption about the orthogonality of the image demonstrates some critical analysis of their procedure. It clearly took two drafts for this team to begin to analyze the limits on their procedure, and even then, their analysis was weak.

Audience – The purpose of the audience dimension is primarily to answer the question, "Did they include everything that was asked for?" In that regard, there is one major flaw in this dimension; a complete lack of results. The client clearly asks for the results of applying the procedure to the three samples provided, yet the team does not include any results. According to the teaching assistant, "First, your memo needs to have results in it and they need to include the RIGHT amount of significant figures and all answers need units." For their second draft, the team only included results for a single sample, despite the instructions asking for results on all three samples. It took one of the peer reviewers marking a Likert item indicating that they did not include results from applying the procedure to push the team to provide all of the requested results (the three originally requested and three additional samples requested for draft 3).

Overall – Over the span of three drafts, the team produced a unique procedure to consistently quantify the roughness of an image. The second draft helped get the procedure into a functional format that produced a result of the appropriate type, mainly a reproducible quantity describing roughness. While an argument could be made that the team had an acceptable procedure in draft two, the third draft allowed the team to produce a complete solution that addressed all of the aspects of the modeling problem. The third draft was necessary in getting the team to provide appropriate rationales and results for all of the provided samples.

<u>Implications and Future Research</u>

The Need for Iteration - In reviewing cases for this paper, a number of issues, as well as a number of potential research questions, became apparent. First, it became evident that students really needed three drafts to achieve a high quality solution. The first draft got them involved in the context, but typically produced a poor model. The second draft resulted in a moderately acceptable model, but usually lacked the finer attributes that resulted in a truly high quality solution. This was what the third draft was for. It was rare to see significant changes between drafts two and three, but usually the changes that were made were critical. Indeed, the average memo increased 72% in length between drafts one and two, but only increased 24% between drafts two and three. A portion of that 24% also had to be used to display the results associated with the three additional test cases.

Peer Feedback - One aspect in which additional research is needed is related to the peer feedback following draft two. There is a two-pronged problem. First is getting students to provide good feedback. With the case study, three of the four peers gave acceptable feedback, however the fourth peer's feedback was poor, often limited to simple phrases and one word responses. In part a motivation issue and in part a training issue, many students give poor peer feedback. This is a common problem within the peer review literature [6, 7]. Training is viewed as one of the key approaches to resolving this. With the MEA, students do receive training in the form of a calibration exercise. Before they review their peers, they review a piece of stock work and compare their review to that of an expert. Despite this exercise, the reviews being generated still need work. Developing and testing a new training mechanism will be essential in improving the quality of peer reviews.

The second issue is in how teams interpret the reviews from their peers. Many teams do not view peer reviews as being of value. An adjustment to the training system will most likely help this problem, but additional mechanisms must be sought out to assure students that their peers are capable reviewers. Additionally, investigating how teams handle conflicting feedback is a question that must be addressed.

Feedback Attributes - In tandem with investigating issues of peer feedback, it would be beneficial to know what attributes of feedback evoke changes along the dimensions needed for a high quality solution. Teaching assistants who left better feedback naturally produced teams with better procedures, so investigating how to harness the positive attributes of that feedback is a critical step in learning to train both teaching assistant and peers to be better reviewers.

Test Cases - Additional work needs to be done on the selection of test cases, specifically, what is the impact of test case selection on quality of student solutions during the iteration process? The three additional test cases were explicitly selected to cause problems in the most common solution paths. For example, many teams elected to use standard deviation of a selection of points as a measure of roughness. By including some of the unique images added to draft three's requirements, the goal was to get students to reevaluate if standard deviation was an adequate measure of roughness. Despite trying to push them in a different direction, most teams elected to stay the course and continue to use standard deviation even when it did not map well to the samples. There was a cognitive dissonance effect, wherein teams rationalized that it must be how they are interpreting the sample and not that their methodology was flawed. An area for further research is in what features of test cases are beneficial in evoking change towards unique solution paths.

Sequencing - Finally, with regards to the iterative process, additional work needs to be done to find the best sequencing. Given the time constraints of the lab component, would it be beneficial to allow students to revise their first draft with no external feedback? Would peer feedback on the first draft evoke the same level of change as teaching assistant feedback? How would the quality of the final draft change if the sequence were spread out over a longer period of time? While the 2005 efforts to increase the number of iterations a team worked on an MEA have generally resulted in improved solution quality, no research has been done to analyze how to best organize those iterations to further improve solution quality.

Conclusions

Iteration has become an essential component to MEAs. Without the second and third drafts, the memos teams produce would suffer from poor repeatability and lack the rigor needed for a good engineering procedure. Through the selection and analysis of this case study, a number of issues came to light that require further research. One thing that became clear is the necessary value of the iterative process. Over the span of multiple iterations, teams gradually moved their procedures closer to the level of quality desired.

References

[1] R. Lesh and H. Doerr, "Beyond Constructivism: Models and Modeling Perspectives on Mathematics Problem Solving, Learning, and Teaching," Mahwah, NJ: Lawrence Erlbaum, 2003.
[2] T. Moore and H. A. Diefes-Dux, "Developing Model-Eliciting Activities for Undergraduate Students Based On Advanced Engineering Content," in *Frontiers in Education* Savannah, GA, 2004.
[3] J. Zawojewski, H. A. Diefes-Dux, and K. J. Bowman, "Models and modeling in Engineering Education: Designing experiences for all students," Rotterdam: Sense Publications, In Review.
[4] H. A. Diefes-Dux, D. Follman, P. K. Imbrie, J. Zawojewski, B. Capobianco, and M. Hjalmarson, "Model Eliciting Activities: An In-class Approach to Improving Persistence and Retention of Women in Engineering," in *American Society for Engineering Education National Conference* Salt Lake City, UT, 2004.
[5] H. A. Diefes-Dux and P. K. Imbrie, "Modeling Activities In A First-Year Engineering Course," in *Models and Modeling in Engineering Education: Designing Experiences for All Students*, J. Zawojewski, H. A. Diefes-Dux, and K. J. Bowman, Eds.: Sense Publishers, In Review.
[6] K. Topping, "Peer Assessment Between Students in Colleges and Universities," *Review of Educational Research,* vol. 68, pp. 249-276, 1998.
[7] N. Falchikov and J. Goldfinch, "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks," *Review of Educational Research,* vol. 70, pp. 287-322, 2000.