121st ASEE Annual Conference & Exposition
Indianapolis, IN
June 15-18, 2014

Paper ID #9492

The Use Of Peer Evaluations In A Non-Traditional First Year System Design Class

Mr. Joseph Pow, Chester F. Carlson Center for Imaging Science, RIT

Joe Pow is the Associate Director of the Chester F. Carlson Center for Imaging Science at the Rochester Institute of Technology. He was the designer and first instructor of a new non-traditional project-based course for incoming Imaging Science freshmen which has had a transformational impact on the Center. Prior to his arrival at RIT, he was a project manager for the Department of Defense, where he was responsible for the development and production of a wide range of new technologies and systems.

Dr. María Helguera, Rochester Institute of Technology

María Helguera was born in Mexico city where she got a BS in Physics from the National Autonomous University of Mexico (UNAM). She also holds an MS in Electrical Engineering from the University of Rochester and a PhD in Imaging Science from the Rochester Institute of Technology (RIT) . Dr. Helguera is the principal investigator in the Biomedical and Materials Multimodal Imaging Laboratory in the Chester F. Carlson Center for Imaging Science (CIS), RIT. Dr. Helguera is also very interested in implementing novel pedagogies in science and technology and has been involved with the freshman imaging project since its inception . She teaches undergraduate and graduate courses at RIT. She is an active member of the CIS undergraduate curriculum committee and has been the undergraduate program coordinator since 2012.

Elizabeth Pieri, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology

Second-year Motion Picture Science student at Rochester Institute of Technology. From Syracuse, NY.

Ms. Sadie Wolters
Mr. Michael Glynn Augspurger, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology

I am a first-year imaging science major in Rochester Institute of Technology. I am part of the university's honors program, which has given me the opportunity to study the pedagogy of project based learning versus traditional, lecture based learning.

Briana A. Neuberger, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology

Second year Industrial/Systems Engineering and Imaging Science major at Rochester Institute of Technology. Background in Geospatial Information Science which stemmed from previous involvement in an organization called Technology Student Association. Currently pursuing research in imaging fields regarding spatial/temporal thresholds and tracking, project-based learning, and systems that integrate tactile learning structures into classroom technology. Interests in intelligence collection and system optimization and efficiency. Hometown is in Herndon, Virginia.

Victoria Scholl, RIT
Elizabeth Bondi, Rochester Institute of Technology

The Use of Peer Evaluations In A Non-Traditional First Year System Design Class

Abstract

In the fall of 2010 the Chester F. Carlson Center for Imaging Science, an imaging systems engineering department at the Rochester Institute of Technology, completely abandoned its traditional lecture based pedagogy for incoming freshmen and in its place implemented a radically different project based class for first year students. Similar to many existing senior level capstone experiences, this new approach challenged first year students to work together as a single integrated multidisciplinary team for a full academic year to design, develop, build, and test a unique, fully functional imaging system from scratch. Now in its fourth year, all indications are that this pedagogy has been transformational, not only for the freshmen who have taken the class, but for the department as a whole. It has changed long held perceptions about the abilities of first year college students, and has led to a new understanding of the role of faculty in technical undergraduate degree programs.

One of the central ways in which this pedagogy differs from a traditional approach is in its desired student outcomes. Whereas the outcomes of the old pedagogy were primarily knowledge oriented, in the new class the outcomes are focused on the degree to which students begin to adopt the behaviors and practices of professional engineers. Consequently conventional assessment tools which measure only knowledge, such as quizzes, tests, and final exams, are of limited value. Instead, instructors in the new class must rely on other techniques to assess student growth and development. One of these is the use of formal peer evaluations. Although these peer evaluations were treated as mandatory assignments due at the end of each academic term, their scope and format were determined by the students themselves. The evaluations were submitted to the instructor, who sanitized them to preserve the anonymity of the evaluators and then compiled and distributed them to each recipient. In this way every student received feedback on how at least a portion of their classmates perceived their performance three times over the course of the year long project.

In this paper we examine the peer evaluations submitted by the first three cohorts to experience the new pedagogy in an effort to gain some insight into their use in a non-traditional classroom. Our analysis focuses on three primary aspects of the peer evaluation system. First, we look at how the scope and format of the evaluations evolved over time as a way of understanding which characteristics the students felt were most essential in their fellow team members. Next we examine the nature and quality of the feedback as a way to assess the perceived value of the peer evaluations. And lastly we draw from the records of the course instructor to see the degree to which the peer evaluations aligned with the perceptions of the instructors. Together these analyses can thus inform the use of peer evaluations as an assessment tool in engineering classes at other institutions.

Introduction - Freshman Imaging Project Overview

For the past several years research on STEM education has consistently revealed the benefits of non-traditional project-based experiences for students at all levels. As a leading sponsor of scholarly work in this area, the Association of American Colleges and Universities' Project

Kaleidoscope (PKAL) has been a vocal advocate for widespread STEM education reform. The themes emerging from PKAL research regarding undergraduate STEM education are clear and consistent:

- Learning should be experiential and steeped in investigation from the very first courses.[1]
- Learning should be personally meaningful for students and faculty, it should make connections to other fields of inquiry, and it should suggest practical applications related to the experience of students.[2]
- Learning should take place in a community where faculty see students as partners in learning, where students collaborate with one another and gain confidence that they succeed, and where institutions support such communities of learners.[3]
- Higher education should produce new frames of understanding by piloting new ideas, tools, and approaches to keep students' learning on the cutting edge.[4]

In 2010 the Chester F. Carlson Center for Imaging Science, an imaging systems engineering department at the Rochester Institute of Technology, developed and implemented a new freshman-level course, known as the Freshman Imaging Project, which embodies this pedagogical framework. While the architects of this new pedagogy wanted it to reflect the most recent research on STEM education, it was also built upon other fundamental beliefs. For example, the belief that first year students are capable of understanding advanced concepts, and their motivation is enhanced by giving them more independence and more control over their educational experience. The team which developed this experience felt strongly that if successful, this pedagogy would be transformational, and would not only challenge widely held perceptions of students' abilities, but also the role of the faculty in undergraduate STEM education.

The new curriculum is a year-long (three academic quarters) sequence of courses in which the students work together as a single integrated multidisciplinary team to design and build a different functional imaging device from scratch each year. The general type of device is specified by the department faculty but the students are responsible for establishing technical performance parameters by assessing the needs of prospective users of their system. Once those performance parameters are established, the students are responsible for creating their own work breakdown structure, as well as planning and executing the entire design and development effort. The only major milestones the students are required to meet are two formal design reviews for external evaluators at the end of the fall and winter quarters, and a public demonstration of the finished product at an annual campus-wide innovation festival at the end of the academic year.

An instructor of record is assigned responsibility for the course but there are no required textbooks or formal lectures. The students jointly construct a common understanding of new concepts by researching in the published literature any topics they need to investigate, and then share their interpretations with their classmates. As necessary, they seek assistance from subject area experts in the faculty, from upper class students, and from outside sources. Scheduled class meetings (two per week for two hours each) take place in a dedicated 800 square foot laboratory configured specifically for this purpose which is available on a 24 hour basis to freshmen enrolled in the course. No other classes are scheduled in this room.

The sequence of courses that make up this year-long experience is required for imaging systems engineering majors, but in an effort to maximize the authenticity of the experience, freshmen from other degree programs are encouraged to enroll in any or all of the three courses. Students continuing in the course from previous quarters are responsible for orienting and integrating any new students into the design team. Although interaction with upper class students is strongly encouraged, formal enrollment in the course is restricted to only first year students.

Since one of the primary outcomes from this pedagogy is to have the students adopt the behaviors of professional scientists and engineers, particular attention is given to providing opportunities for the students to share their experiences with a variety of audiences in both written and oral formats. For example while doing their initial research the students compile a collection of written précis which help them construct a common understanding of key technical concepts. Their written products also include requests for purchases of equipment, user's manuals for their systems, and responses to action items raised at the formal design reviews. Those design reviews are the primary oral presentations each quarter, but presentations are also given to a variety of undergraduate student groups such as the Society of Imaging Science and Technology, the Society of Mechanical Engineers, and the university's College of Science Undergraduate Research Symposium.

Because this pedagogy represents such a radical departure from any the department has previously used, it is important to evaluate its effectiveness. To do this, the department initially planned to enlist the aid of external evaluators to conduct a formal assessment. However anticipated funding to support this effort did not materialize, so a rigorous evaluation has not yet been performed. Additionally, the desire to draw any clear conclusions regarding its effectiveness is also hampered by the small sample size. To date, only four cohorts – a total of 84 students including those who are currently enrolled – have taken this class. And since the students from the first cohort have yet to graduate, the full impact of this pedagogy on their academic careers is just now being assessed. More data must be collected and a more comprehensive analysis must be done before any definitive conclusions can be drawn about the effectiveness of this approach.

In spite of this, the department is attempting to identify any early indications in student attitudes and behaviors which may be attributable to this pedagogy. Indeed, some appear to be emerging. For example, student feedback on standard course evaluations and other informal surveys has been overwhelmingly positive. When asked "What is your overall rating of this course?" 72% of the responding students have given it the "best possible rating," another 15% rated it "above average." No students have ever given ratings of "below average" or "worst possible."

The Use of Peer Evaluations

One of the primary challenges associated with the implementation of a non-traditional pedagogy, especially one like the Freshman Imaging Project in which there are no quizzes, tests, or final exams, is student assessment. In this class, the final grade is based solely in each student's attitude, effort, and contribution to the group. In order to ensure that the students always know where they stand with respect to their performance, they are required to individually meet with the instructor at least every three weeks. During these meetings the student and the instructor

talk about the grade they can expect if they continue on their current "trajectory," and specific actions they can take if they want to modify that "trajectory" to improve their grade. They also get feedback at the end of each quarter through formal, written peer evaluations. These evaluations are submitted to the faculty member, who first sanitizes them by removing the identity of the rater, and then compiles all of the ratings and comments provided for each individual regarding their performance.

The use of peer evaluations in conjunction with group projects is not new. A comprehensive literature review by Fellenz[5] identified a variety of reasons why faculty have used peer evaluations, including differentiating between individual contributions, increasing fairness and accuracy of grades, establishing a positive learning environment, and developing professionally relevant skills. The objectives of incorporating peer evaluations into the Freshman Imaging Project included all of these to varying degrees.

From a faculty perspective the evaluations influence assessment in two ways. First they provide the instructors with additional insights and information about student contributions that may not have otherwise been available. With a class of 20 or more students, working in parallel on different aspects of the project in multiple locations simultaneously, it is impossible for the instructors to observe everything that happens in the class. The peer evaluations are a way to help ensure that students are receiving full credit for their contributions. However, recognizing that the scope and duration of the project would place some unique demands on the students, and challenge them in ways they had never been challenged before, the faculty anticipated that tension and stress might skew the participants' perceptions of their classmates' performance. It was therefore decided that the peer evaluations would never have a negative impact on a student's grade. Additionally, the peer evaluations served as a tool to motivate discussion during the one-on-one advising sessions. Comments in the evaluations could springboard a dialog about specific situations in the class or actions a student could possibly take to help move the project in a positive direction.

For the students, the peer evaluations were an initial exposure to the experience of giving and receiving professional feedback – an essential behavior that must be mastered on their journey to becoming practicing engineers. At the beginning of each quarter, after the peer evaluations from the previous quarter had been compiled and distributed, the instructors would discuss them with the class as a whole. In these open discussions the instructors would solicit reactions from the students on the feedback they had received, and offer a faculty perspective of the process and discuss ways to interpret and respond to that feedback. These sessions were an attempt to demystify the process of giving professionally relevant feedback to their peers, to have the students to reflect on their personal development, and to get them to the point where they can receive constructive criticism without feeling threatened.

The Purpose of This Study

One of the primary lessons learned by the faculty who have taught using this pedagogy is that the classroom becomes an extremely dynamic and complex place. Student behaviors in such an environment are influenced by a myriad of interrelated factors. Deconstructing student actions at both the individual and group level is enormously challenging. But doing so is essential to

assessing whether student development is moving in a positive direction in critical dimensions. To that end, everything associated with the class becomes data that could potentially illuminate student motivations, and help the instructors adapt to this extremely fluid environment

One of the richest data sets arising from the class is the collection of quarterly peer evaluations. With each cohort producing hundreds of these evaluations, they offer the opportunity to gain valuable insight into student attitudes and growth. By mining the peer evaluations from the first three cohorts of this class we hope to answer the following questions:

1. What attributes and behaviors did the students value most in their classmates? By giving each cohort the opportunity to design its own peer evaluation instrument we might be able to see what the students felt were the most important traits they needed to have in order to function effectively in this setting, and we should be able to see if those values change over the course of the year the students work together on the project.

2. What were the students' attitudes toward peer evaluation? Ideally, as budding engineers we would like the students to develop a healthy appreciation for the value of constructive professional criticism in their chosen discipline. However research has shown that peer evaluation is not always viewed in a positive light.[5,6] By carefully analyzing our collection of peer evaluations, we hope to glean how professionally the students approached this task.

3. Did the students' perceptions of their classmates' performance align with the instructor's perception? Again, several studies have been undertaken in this area (cite, and elaborate). Since our students submitted peer evaluations three times each year, our goal is to determine whether that alignment changed over time.

The Peer Evaluation Instruments and Their Evolution

While the students in the Freshman Imaging Project were required to submit peer evaluations, they were never told what form those evaluations should take. From the beginning, the students were given the opportunity to tailor an evaluation instrument specifically for their cohort, and were in fact encouraged to modify the instrument from quarter to quarter if they felt the need to do so. They were also given the opportunity to determine how the evaluations would be administered in terms of which and how many of their classmates each student would be required to evaluate. The only aspect of the peer evaluation process that was enforced by the instructor was the requirement to actually do it. Those students who failed to submit their peer evaluations by the end of each quarter were given grades of "incomplete." Once the late evaluations were submitted, the incomplete was changed to a letter grade, with no penalty imposed for late submission.

When possible, each cohort took advantage of the work the preceding cohorts had done on designing and administering a peer evaluation instrument. Since the students in the first cohort had no prior work on which to base the design of their instrument, they devoted time early in the year to researching best practices in this area. Their first attempt clearly exhibited their lack of experience in designing a suitable instrument and administering the evaluation process. Although there were only six students in the first cohort's fall quarter, three slightly different instruments were used. The differences were due to the fact that some of the students completed their evaluations before the design of the instrument had been finalized. All three designs had

the raters provide feedback on their classmates' timeliness, effort, and participation, as well overall performance. They were also able to comment on any other items not mentioned elsewhere on the form. A second version of the instrument added an evaluation of completeness and accuracy of work. A third version included all of these items as well as communication.

Just as the variety of instruments indicated the relative immaturity of the process, so too did the wording of the questions display a lack of experience in designing instruments which would provide the most meaningful and relevant feedback. For example, the question about timeliness, common to all of the fall evaluations in the first cohort, was worded "Did the group member turn in their work in a timely manner?" Obviously this question can be answered with a simple yes or no. There was no attempt to prompt the rater into indicating how often the ratee submitted work late, there was no opportunity to address the reasons the work may have been submitted late, and there was no opportunity to suggest corrective actions, if required. But in spite of their shortcomings, these three variants provide us with an initial glimpse into the behaviors and attributes these first few students in Cohort 1 initially valued in their classmates.

Table 1. Attributes included in Cohort 1 final peer evaluation instrument.

1. Communication. Take into consideration the individual's ability to communicate not only within his/her Work Breakdown Structure (WBS) group, but also with other groups in the project. Also examine the individual's clarity and concision when speaking with others.

2. Contribution to Discussions. Evaluate this individual's contributions to class discussions this quarter. Does he/she provide thoughtful and insightful comments in a group setting? Do this individual's comments encourage further discussion? Is this individual receptive to ideas of others?

3. Advancement of Knowledge. Evaluate the individual's effort in learning new material this quarter. Have they been proactive in doing research and talking to outside sources to learn more about their part of the project?

4. Attitude and Peer Relations. Evaluate the attitude of the individual. Is this person's mindset conducive to an efficient workplace? Are they supportive of others and open to the ideas of others? Do they work effectively in groups with people? Does this individual show high levels of professionalism in dealing with others?

5. Effectiveness of Leader. If this person was a leader, were they effective at their job? Were they able to clearly define group goals and objectives, and then assign tasks to group members to reach those goals? Was this person reliable and able to achieve results in a timely manner?

Other Comments. Use this space for any additional comments and constructive criticisms of this individual's contributions to the Freshman Imaging Project. If there are any areas in which this individual could improve, or areas in which this individual particularly excelled that were not previously mentioned above, those may be listed here.

In the second quarter of the Cohort 1 the students were given the chance to redesign the instrument but struggled to do so. In the end, when it came time to perform the peer evaluations, the class chose to simply have each student submit a summary paragraph on each of their classmates' performance without having to rely on specific prompting questions. While these evaluations provided the students with significant feedback, in the third quarter they decided to return to a more structured instrument. One student designed a peer evaluation form which included Likert scale ratings of five specific attributes as shown in Table 1 above. This modified instrument was adopted by the class as a whole and used during the final quarter of Cohort 1.

At the beginning of Cohort 2 the new group of students was given the final format of the peer evaluation used by Cohort 1 to use as a point of departure for designing their own unique instrument, tailored to reflect their thoughts on how to provide classmates with meaningful feedback. They used this instrument unchanged in the fall quarter. Since each cohort designs and builds a different device, and is given the freedom to pursue its development in whatever way they wish, circumstances within the class may dictate changes in the form or administration of the peer evaluations. This became apparent to the students of Cohort 2 as they approached the end of the winter quarter. During that quarter they had divided themselves into five groups based on the five technical approaches they wanted to evaluate for their system. Each of these groups functioned independently of the others during this phase of the program, so as they approached the end of the quarter they realized that they wouldn't be able to provide meaningful peer evaluations to all of their classmates. They decided to evaluate only the members of the group to which they belonged. They used the same instrument as the one they used in the fall, with the only addition being a requirement to indicate which group the ratee worked in.

Table 2. Attributes included in Cohorts 2 and 3 final peer evaluation instrument.

1. Communication. Take into consideration the individual's ability to communicate not only within his or her group, but also other groups within the project. Also examine the individual's clarity and concision when speaking with others.

2. Contribution to Project. Has this person chosen to take on tasks relevant to the project? Do they put forth effort inside and outside of class that is relevant to the project?

3. Dedication to Assigned Tasks. Is this person gainfully employed during class? Do they put effort into what they are working on in the classroom?

4. Attitude and Peer Relations. Evaluate the attitude of the individual. Is the person's mindset conducive to an efficient workplace? Are they supportive of others and open their ideas? Do they work effectively in groups with people? Does this individual show high levels of professionalism in dealing with others?

Comments. Please use this space for any additional comments.

By the spring quarter the students of Cohort 2 had completed the tradeoff study of the five technical approaches and once again came together as a single group to build one prototype system and prepare for their public demonstration. Prior to the end of the year they made one

last change to the peer evaluation instrument. As the project moved from research and tradeoff studies into final system integration, they changed the specific areas addressed in the instrument. They eliminated the items relating to the "contribution to discussions," "advancement of knowledge," and "effectiveness of leader," and added two new items, one which addressed "contributions to the project," and another on "dedication to assigned tasks." The final instrument is shown in Table 2 above.

When the new students arrived the following fall at the beginning of Cohort 3, they were again given the most recent version of the peer evaluation instrument with the understanding that they would be able to make whatever changes they wanted prior to actually rating their classmates at the end of the quarter. They chose to make no changes to this instrument, and in fact used it in the same form throughout the entire academic year.

Looking across all three cohorts then we see that the peer evaluation instrument evolved the most in the first year, with the students using a different one in each of their three quarters. In the second year two different instruments were used. And in the third year only one was used. This convergence over time toward a stable instrument may be an indication that, for now at least, the students have settled on the behaviors or attributes they value most in their classmates in this non-traditional setting.

Analysis of the Peer Evaluations

Armed with an understanding of the behaviors that students valued most, we can now begin to examine how the peer evaluations were used in practice by the students and how they perceived their fellow classmates. To do this we performed a mixed-methods analysis of all of the peer evaluation instruments submitted by the students in Cohorts 2 and 3. The evaluations from Cohort 1 were not included in the analysis for two reasons. First of all, the original evaluations are no longer available. All that remains from Cohort 1 are the feedback summaries given to each student by the instructor after compiling all of the instruments submitted for that person. Therefore, there is no longer any way to attribute a particular evaluation to the person who wrote it. Second, as was already mentioned, there were three different instruments used by Cohort 1. Because of the lack of continuity from quarter to quarter, it was impossible to assess perceived improvements in a given area in that first year. In contrast, the instruments used by Cohorts 2 and 3 were relatively stable, and had sufficiently aligned content that the datasets from the two groups can be analyzed.

All of the data analyzed had originally been submitted to the instructors in an electronic format. In order to preserve the confidentiality of both the raters and the ratees throughout the analysis, the instructors who taught the class sanitized the data by stripping out all identifying information. This consisted of replacing the names of all individuals, whether they were part of a file name or embedded in the evaluation instruments, with a numerical code. The key to the code was maintained by the course instructors. None of the research assistants who helped with the analysis were thus able to identify either the rater or the ratee. The only things they knew for sure about a given evaluation were the cohort and quarter in which it was written.

The population which produced the peer evaluations that were analyzed in this study consisted of 45 freshmen (27 males, 18 females). Of this group, 31 were enrolled in the class all three quarters of their freshman year, eight were enrolled for two quarters, and six were enrolled for

just one quarter. During the time they participated in the class, 41 had declared majors in six different disciplines, while four were enrolled in undeclared "exploration" programs in two areas. This population of students produced a total of 691 peer evaluations – 441 in Cohort 2 and another 250 in Cohort 3. The numbers for Cohort 3 are lower because in this class each student evaluated 5 others at the most, whereas in previous years each student evaluated an average of 10 peers.

The analysis of these data began with a statistical investigation of the Likert scores associated with the various questions on the evaluation instruments. The goal of this part of the analysis was to determine whether the students in each of the cohorts perceived an improvement in the performance of their classmates over the course of the year. For Cohort 2 this is only possible in two areas: communication and attitude/peer relations. Since the instrument was modified prior to the spring quarter, these are the only two factors that remained common to all instruments throughout the year. The instruments used in Cohort 3 were never modified, thus allowing us to assess perceived improvement in all areas throughout the year.

To assess perceived improvement we calculated the mean Likert scores for each of the common questions in Cohorts 2 and 3, then performed paired sample t-tests to determine whether the changes in the means from quarter to quarter were significant ($p<0.05$). The results were inconclusive. In Cohort 2 there was a significant improvement in communication skills only during the winter quarter. There was no significant change in attitude and peer relations. For Cohort 3 the analysis also shows a significant improvement in communication skills in the winter quarter but not in the spring. Perceived contributions to the project saw a significant improvement only in the spring quarter. There was no significant change in dedication to assigned tasks or attitude and peer relations.

Table 3. Rubric for analysis of comments in peer evaluations.

1. How many times did a rater provide any comments in response to a question on the instrument (comments were always optional)?
2. How many of the comments were performance-related (either positive or negative)?
3. How many of the comments were positive?
4. How many of the comments were negative?
5. How many of the comments included specific examples of observed behavior?
6. How many of the comments offered specific examples of actions to improve performance?
7. How many of the comments were written as if the rater was communicating to the instructor?
8. How many of the comments were written as if the rater was communicating directly to the ratee?

With no clear trends emerging from the Likert data we moved on to an analysis of the written comments that were provided with the peer evaluations. The first step in this part of the analysis involved simply tallying the comments according to the rubric shown in Table 3.

These numbers were compiled with the assumption that if the students were beginning to adopt the attitudes and behaviors of professional engineers over time that we would see an increase in the use and quality of comments in the peer evaluations. In this case "quality" comments would be those which were balanced in terms of positive and negative feedback, which included specific examples of behaviors which prompted the comments, and which offered suggested strategies for improving in any areas perceived as weak. An analysis of these numbers revealed that between the fall and the spring:

1. The relative number of comments (with respect to the number of opportunities to comment) went up 19%. Each peer evaluation offered the rater several opportunities to provide written comments. In addition to having the rater assess performance on a Likert scale, each question also offered the opportunity to elaborate on the rating with written comments. So the total number of opportunities to provide comment was the number of peer evaluations submitted, times the number of questions (including the open-ended opportunity at the end of the instrument). In Cohort 3, for example, this came to 560 opportunities in the fall, 360 in the winter, and 375 in the spring. Comparing the actual number of comments provided to the number of opportunities available indicated that over half took advantage of these opportunities in the fall, and that number continued to grow with each successive quarter.

2. The relative number of performance-related comments stayed consistently high. While there were instances of comments which had no constructive value ("Have a great vacation") the data show that the overwhelming majority of the comments actually focused on relevant performance-related items.

3. The relative number of positive comments (which started out very high) went up 20%, while the relative number of negative comments (which started out very small) went down 8%. By the end of the year, virtually all of the comments were positive and virtually none of the comments were negative.

4. The relative number of comments which gave specific examples of behavior went up 20%. Rather than just stating, for example, "you need to improve your communication skills," over the year the raters became slightly better at citing specific behaviors that motivated the Likert rating, such as stating "when you give a presentation you refer too often to your notes."

5. The relative number of comments which recommended specific corrective action (which was always small) went down 6%. After the first quarter only about 1 in 10 comments offered suggestions about things the ratee could do to improve their performance. By the end of the year there were almost none.

6. When providing written comments, the rater almost always talk *about* the ratee, and virtually never talk *to* the ratee. This was an interesting observation. Although the students all knew the purpose of the peer evaluation was to give some of their classmates feedback about their performance, the comments were virtually always written as if the rater was communicating with

their instructor, and not their classmate. So rather than writing, for example, "*you* could have done a better job preparing for the design review," almost all raters would have written "*Dan* could have done a better job preparing for the design review."

The final area of analysis was an attempt to determine, using the peer evaluations, the degree to which the students' perceptions of their classmates' performance aligned with that of the instructors. Given the non-traditional nature of this class and the lack of quantitative performance indicators, the instructors were generally very lenient in their grading. Grades of "A" were the norm (74% of all grades), "B's" (17% of all grades) were an indication that a student needed to improve, and "C's" (9% of all grades) meant that the recipient's attitude, effort, and contributions were significantly lacking. No grades of "D" or "F" were ever given. This very narrow distribution of grades complicates this portion of the analysis. Had a broader distribution of grades been assigned, the mapping of the instructors' perceived effort onto the five-level Likert scale in the students' peer evaluations would have been more straightforward. But given the lack of fine resolution in the assigned grades (this university does not use +/- when grading), this analysis looked only at the mean Likert scores of those students who received grades of "B" and "C" – the bottom 26% of the grades issued – to see how they compared to the class mean as calculated from the peer evaluations. If the students' assessment of their classmates' performance aligned with that of the instructors, we could expect that all of those receiving grades of "B" and "C" would be significantly below the class mean Likert score for the quarter in which the grades were issued. However, our analysis of the peer evaluations shows that only 53% of those students were significantly below the class mean ($p<0.05$), which indicates that about half of the class disagreed with the instructor. This result contradicts the literature review done by Ryan et al.[7] which cited multiple studies in which there was a strong correlation between peer and instructor grades. An examination of the comments which accompanied the peer evaluations for which there was a disparity between the instructor and the students reveals that having a pleasant personality may carry more weight with the raters than the actual contributions of the individuals being evaluated.

Interpretation of the Data

Having done both quantitative and qualitative analysis of the peer evaluations, the challenge now is to extract useful information from the results. As previously mentioned, this is an extremely difficult task, due to the many confounding factors that influence student behavior in this non-traditional environment.

Our first objective with this study was to gain an understanding of the attributes and behaviors that the students valued most in their classmates, and to see if those values changed over time. Ideally those values would be represented explicitly by the specific questions the students included in the peer evaluation instruments. If that is the case, then the fact that over a three year period the questions in the instrument have fully stabilized might, at first glance, mean that in that time they have actually found a comprehensive set of questions that address all of the attributes needed to succeed in this class.

But upon closer examination, the final stable instrument reveals areas where there may still be opportunities for improvement. Specifically, the most mature peer evaluation instrument (used in the spring quarter of Cohort 2 and throughout Cohort 3) contains factors including "Contribution to Project" and "Dedication to Assigned Tasks." Without any further explanation

it might be assumed that the former addresses whether or not the rated individual had a tangible impact on the design and development of the system, while the later only looked at effort, regardless of whether it contributed to moving the project forward. However these questions *are* accompanied by short rubrics (see Table 2) which describe what the rater should consider when assigning a value or making comments in these areas, and those rubrics appear remarkably similar, to the extent that they use some of the exact same words ("Do they put forth effort inside and outside of class…" vs. "Do they put effort into what they are working on in the classroom"). The similarity of these two factors can only be attributed to one thing – the students aren't giving sufficient critical thought to these items.

By looking closely at the results of the quantitative analysis we can also infer that the students don't view the peer evaluations as an opportunity to give relevant, meaningful feedback to their classmates. This can be seen in the discrepancy between the mean Likert scores, which rarely show significant improvement over time, and the tone of the written comments, which get progressively more and more positive to the point where virtually no negative feedback is given by the end of the year. If the students were taking the peer evaluations seriously, one would expect that the nature of the comments would vary with the mean Likert scores – a drop in the mean would result in less positive and more negative comments, while an improvement in the mean would have the opposite effect. Since the mean scores don't often show significant improvement, the clear increase in the percentage of positive comments might indicate that the students aren't putting in the effort needed to offer constructive criticism.

This assessment of student attitudes toward the peer evaluations is also reinforced by the disparity between the mean Likert scores and the feedback given by the instructor in the regular one-on-one meetings. Here the contrast is dramatic. While only 4%-10% of all written comments in the peer evaluations address specific actions the ratee can take to improve, the instructor's notes show that almost every one-on-one meeting included a discussion of what the student could do to move themselves, the group, and the project forward. Granted, the instructors have far more experience at assessing student performance and giving clear, concise feedback, but even though they regularly model these behaviors it appears that the students make little effort to adopt them as their own.

The lack of critical assessment of their classmates is also evident from the disparity between the students and the instructors when it comes to perceived performance. The fact that only half of the time will students acknowledge on the peer evaluations that a classmate who is dramatically deficient needs to improve challenges the assertion made by some researchers[8] that in classes such as this the students themselves are in the best position to provide each other with meaningful feedback on their performance. On the contrary, this disparity is yet another indication that a peer evaluation process which is fully designed and administered by students may not produce the desired outcomes.

Conclusions

When we implemented a new non-traditional pedagogy for our incoming system design students, our primary objective was to have them begin to adopt the behaviors of professional engineers by the end of their first academic year. One element of that pedagogy included the regular use of formal peer evaluations. It was hoped that by incorporating peer evaluations into the experience that the students would learn how to give and receive constructive criticism in a non-threatening

fashion, and to reflect on and improve their own performance – behaviors that are central to success as a professional engineer.

Our goal in undertaking this study was to determine how effective the peer evaluations were in achieving those desired outcomes. Using information extracted from quarterly peer evaluations, we attempted to identify the attributes the students valued most in their classmates, to assess the students' attitudes toward the peer evaluation process, and to see whether the students' perceptions of their classmates' performance aligned with their instructors. This study was complicated by a number of factors, not the least of which was the fact that the students themselves controlled the entire process. As a result, we had to deal with challenges such as a dataset that includes multiple instruments, and populations that change from quarter to quarter and year to year. Consequently, our results are all over the board and it's almost impossible to find any consistent themes. In the end, it may be that the only conclusion that we can draw from this effort is that it may not yet be possible to draw any firm conclusions.

While we may have fallen short of our objectives, this experience has done much to help set the stage for more rigorous and meaningful analysis of peer evaluations in the future. For example it may be helpful to encourage the students to think more critically about the development of the peer evaluation instrument and process. Suggesting to them at the beginning of the year that they investigate more deeply the research on peer evaluations, and engaging them in a dialog about their findings could result in changes which would provide the instructors with a dataset that lends itself more readily to the kind of analysis we are trying to undertake. The challenge here is to stimulate student effort in this direction without undermining the high degree of autonomy that they are given in this class. Ultimately, when viewed from a "big picture" perspective, it is far more important to have the students genuinely feel that they own all aspects of the year-long experience than it is to take some of that ownership away in the interest of providing the instructors with a dataset which is more conducive to analysis.

With that in mind, this research has motivated us to identify ways to collect additional data which could complement the peer evaluations and facilitate their analysis without compromising the integrity of the pedagogy. Specifically, it has become clear that the one-on-one meetings the instructors have with each of the students afford the opportunity to gain valuable insights about their reactions to the feedback they are receiving from their peers. This information has been sorely lacking in the analysis to date. By ensuring that peer evaluation feedback is routinely discussed with the individual students during the one-on-one sessions we will be able to gain a deeper understanding of the rationale behind the ratings that are given, and the impact they have on the recipient. Armed with this additional data we will be well positioned to revisit the question of how to foster more constructive feedback between members of the design teams.

References

1. Healey, Mick, and Alan Jenkins. "Strengthening The Teaching-research Linkage In Undergraduate Courses And Programs." *New Directions for Teaching and Learning* 2006.107 (2006): 43-53. Print.
2. Narum, J. "Promising practices in undergraduate STEM education."*Commissioned paper presented at NRC workshop on Evidence on Selected Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education*. 2008.

3. Fairweather, James. "Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education."*Board of Science Education, National Research Council, The National Academies, Washington, DC* (2008).
4. Barnett, R.. "University Knowledge in an Age of Supercomplexity." *Higher Education* 40 (2000): 409-422. Print.
5. Fellenz, M. R.. "Toward Fairness In Assessing Student Groupwork: A Protocol For Peer Evaluation Of Individual Contributions." *Journal of Management Education* 30.4 (2006): 570-591. Print.
6. Brooks, Charles, and Janice Ammons. "Free Riding In Group Projects And The Effects Of Timing, Frequency, And Specificity Of Criteria In Peer Assessments." *Journal of Education for Business* 78.5 (2003): 268-272. Print.
7. Ryan, G. J., L. L. Marshall, K. Porter, and H. Jia. "Peer, Professor And Self-evaluation Of Class Participation." *Active Learning in Higher Education* 8.1 (2007): 49-61. Print.
8. McGourty, Jack, Peter Dominick, and Richard Reilly. "Incorporating Student Peer Review Feedback into the Assessment Process." *FIE'98: 28th Annual Frontiers in Education Conference : Moving from "teacher-centered" to "learner-centered" education : conference proceedings : Tempe Mission Palms Hotel, Tempe, Arizona, November 4-7, 1998*. Piscataway, N.J.: IEEE, 1998. 14-18. Print.