

AC 2008-674: COMPARISON OF TWO PEER EVALUATION INSTRUMENTS FOR PROJECT TEAMS

Kerry Meyers, University of Notre Dame

Kerry L. Meyers is an Associate Professional Faculty Member and Co-coordinator of the First-Year Engineering Program at the University of Notre Dame.

Matthew Ohland, Purdue Engineering Education

Matthew W. Ohland is an Associate Professor and Director of First-Year Engineering in the School of Engineering Education at Purdue University and is the Past President of Tau Beta Pi, the engineering honor society. He received his Ph.D. in Civil Engineering with a minor in Education from the University of Florida in 1996. Previously, he served as Assistant Director of the NSF-sponsored SUCCEED Engineering Education Coalition. In addition to this work, he studies peer evaluation and longitudinal student records in engineering education.

Stephen Silliman, University of Notre Dame

Dr. Silliman is a Professor of Civil Engineering and Geological Sciences and an Associate Dean of Engineering at the University of Notre Dame. His research focuses on groundwater hydraulics and includes educational and research initiatives in developing countries.

Leo McWilliams, University of Notre Dame

Dr. Leo H. McWilliams has been a Course Coordinator for the First-Year Engineering Program at the University of Notre Dame since 2001. Prior to joining Notre Dame in this capacity, he worked as a principal engineer at Honeywell International. Dr. McWilliams received his B.A. in economics, B.S.E.E., M.S.E.E. and Ph.D. from the University of Notre Dame.

Tracy Kijewski-Correa, University of Notre Dame

Dr. Kijewski-Correa is the Rooney Assistant Professor in the Department of Civil Engineering and Geological Sciences at the University of Notre Dame. Her research focuses on structural dynamics, wind effects on structures, and the use of novel sensors to assess structural health and integrity.

Comparison of Two Peer Evaluation Instruments for Project Teams

Abstract

The College of Engineering at the University of Notre Dame has utilized a paper-pencil instrument for peer evaluations since 2005 as a portion of the assessment of project team efforts (typically 4-5 students per team) in its First Year Engineering Course. The College was considering moving from paper-pencil peer evaluations to an on-line, behaviorally based evaluation instrument, CATME¹. The instructors at Notre Dame conducted a comparative study of student feedback on these two instruments during the fall 2007. During the fall semester, the students (~380) within the first year course were divided into two groups, one group using the paper-pencil instrument and the second group using CATME, both groups of approximately equal size. After completion of peer evaluations for a seven-week course project, the students were required to complete a survey providing their reaction to the instrument they used in terms of perceived simplicity, comfort, confidentiality, usefulness of feedback, and overall experience. Comparison of results from the surveys provided insight into both the relative merit and drawbacks of the two administrations. Several of the follow up survey questions comparing the instruments did not show statistically significant differences in the sample means. In spite of the confounding of the instrument design and the administration method, useful results emerged. The biggest differences in student survey results were seen in the areas of feedback and overall experience, both of which were higher for CATME. Student confidence in instructor confidentiality (keeping their comments confidential) was high for both instruments, but it was slightly higher for the paper-pencil instrument. Because student perception of the quality of the feedback is critical to both rater accuracy and the student learning experience, this study enabled the College to make a data-driven decision to use the CATME instrument in future offerings of the first year course.

Introduction

College students, regardless of their field of study, commonly work collaboratively in groups on course assignments. The benefits of collaborative learning have been well documented^{2,3,4} and are rarely disputed. However, collaboration can lead to difficulties in evaluating the work of individual students. For example, how can instructors ensure that all students are contributing appropriately towards the completion of a project? There are often concerns over hitchhiking, a phenomenon wherein a student does not contribute adequately towards the project goals and allows teammates to do the majority of the project work. There is a disconnect because the instructor is not typically present for much of the time the group spends working on a project outside of class, yet the instructor must assign individual course grades. Social dominance is another potential issue, wherein a student takes over a project and does not allow other group members to contribute to project goals in a meaningful way. Given these challenges, finding an effective method to assess and assign individual contributions to group work is a topic of much research and debate within the education community, with substantial attention being paid to the benefit, and possible limitations, of peer evaluation methods^{5,6,7}.

Feedback received through peer evaluation has generally been viewed by students as useful and beneficial to improving performance^{8,9}. Peer evaluations are less likely to be lenient (higher than actual performance)¹⁰ or harsh (lower than actual performance)¹¹ when they are made for the purposes of feedback, development, or research rather than a strictly administrative purpose. Thus it is critical to assess the student perception of the feedback from each instrument. It has also been found that confidentiality (but not anonymity) makes raters more accountable and thus more accurate in their ratings¹².

For the past two school years, 2005-2007, the first year engineering students at the University of Notre Dame have used a paper-pencil peer evaluation to document individual contributions to a design project. This evaluation was developed empirically by three faculty members in the College of Engineering. Consideration was being given to replace this paper-pencil instrument with an on-line instrument that is theoretically based and eases administrative burdens. The motivation for this study was to compare the paper-pencil on-line instrument called the Comprehensive Assessment of Team Member Effectiveness¹. CATME was designed to provide automated feedback and reporting to students and to ease administrative burdens. These two instruments are different both in the content and administration methodologies. Two studies comparing teaching evaluations have shown that students perceive on-line administration to be less anonymous than written evaluations^{13,14}. Hence, the purpose of the present study is to compare student response to these two instruments to determine whether there are any advantages or disadvantages (from the student perspective) to making a change to the on-line, behaviorally based instrument.

The specific areas for comparison include student perceptions of: simplicity, comfort, confidentiality, usefulness of feedback, and overall impression, specifically to answer the research questions:

- Which instrument do students perceive to be simpler (straight-forward and easy)?
- With which instrument are students more comfortable (at ease or content)?
- Which instrument do students perceive to ensure a higher confidentiality? Do students express any concern regarding confidentiality of their responses?
- Which instrument do students perceive the feedback from peer evaluation to be more useful?
- For which instrument do students report a more positive overall experience?

The present study parallels previous studies comparing teaching evaluations administered in class and on-line^{14,15}. In the study by Dommeyer et al. (2002), approximately half of the student participants conducted their evaluation on-line and half conducted their evaluation in-class. A follow up survey was administered to assess student perceptions of the evaluation methods. Dommeyer et al. found that students perceived the on-line instrument to be less anonymous than the paper instrument. The present study is similar to that of Dommeyer et al., but looks at administration methods for peer evaluation rather than teaching evaluations. Further, the two instruments being compared within the present study are founded on fundamentally different approaches to peer evaluation. The CATME instrument uses behavioral anchors as descriptions of team member

performance, whereas the paper-pencil evaluation uses Likert scales and rankings based on empirically derived questions.

Data Collection & Methods

Students enrolled in the first year engineering course at the University of Notre Dame during the fall of 2007, had the same requirements as in previous offerings of the course, which includes participating in a half-semester group design project. After the completion of the project, all students are required to fill out a peer evaluation to document their perception of individual contributions by their fellow group members. During the most recent offering of this course, approximately half of the students used CATME rather than the traditional paper-pencil instrument. The other half of the students completed the traditional paper-pencil instrument. Appendices A and B provide the content of these two instruments. We acknowledge that the administration (online vs. paper-and-pencil) and the design (theoretical vs. empirical) of the instruments are confounded, but we do not perceive that as a threat to our study, which focuses on the perception of students. Regardless of how the evaluations are designed and administered, student perceptions of the instruments, and particularly the feedback the instruments provide, are critical to the rating accuracy. After the peer evaluations were completed and feedback provided to the students, the students were required to complete an 18-question survey to assess their perceptions of the evaluation process and instrument they used. This survey focuses on simplicity, comfort, usefulness of feedback, and perceived confidentiality of the instruments (see Appendix C for survey questions). The results of the survey were compared for the two instruments with statistically significant differences in responses highlighted.

In addition to the student reactions, differences in the two instruments were studied via documentation of the administration process as collected from the two faculty members who administered the course evaluations. Specifically, the faculty members were asked to document how the two instruments were used to influence individual student grades, Appendix D shows the responses from the two faculty members.

Participants

All students in Introduction to Engineering Systems I at the University of Notre Dame were participants in the study. With few exceptions, all subjects are traditional freshman engineering students approximately 18 years of age. Upper-class students and students from two local colleges (Saint Mary's or Holy Cross) together comprise less than 3% of the enrollment of approximately 380 students.

The first-year engineering course is composed of 13 sections, each containing up to 30 students. Students were assigned essentially randomly to sections prior to their arrival at the University. (Students have no control over which section they are assigned to by their first year advisor, the only influence on which section they are assigned are the other courses they are registering for). Two faculty members each teach 6 sections (consisting of approximately 175 students), so these two faculty members cover 12 of the 13 sections and approximately 350 of the 380 students enrolled. The 13th section (taught by a separate

lecturer) is not considered in this evaluation. The 12 sections were assigned to the different evaluation instruments such that half of each instructor's sections were designated to use CATME and half were designated to use the paper-pencil instrument, so it was possible to control for any instructor effect and any interaction between the instructor and the instruments.

Statistical Methods

All data collected for the student survey of the two instruments were collected using Survey Monkey¹⁶ during a two week period in October of 2007. The analysis of these data was conducted on the raw data and includes descriptive statistics, t-tests, chi-squared, correlation, multiple regression, and decomposition of variance. The descriptive statistics and t-tests were considered for the overall data collected, by method (on-line or paper-pencil), and by instructor and these methods are the presented in this paper. All analysis was conducted using a statistical software package, STATA¹⁷.

Table 1 provides a summary of the various variables involved in evaluating the individual instruments and additional grouping variables used in our analysis. The grouping variables included: (i) the evaluation method used by the respondent (CATME or paper-pencil), (ii) the course instructor, and (iii) a question was asked regarding the student's general preference for on-line versus paper administration of course materials. Within the instrument variables, the time to complete the instrument is a discrete variable, with an option to select from among 1-minute increments between 0 and 25 minutes. The remaining questions are shown in Table 1 below and are all based on a Likert Scale where 1 is strongly disagree and 5 is strongly agree. While it is recognized that the data are neither continuous nor normally distributed, the population sizes are large enough to assume that the sample statistics approach continuous, normally-distributed random variables.

Table 1. Summary of Survey Questions

<u>Categories</u>	<u>Survey Questions</u>
Evaluation Method	<ul style="list-style-type: none"> ▪ The evaluation method was simple to use ▪ The evaluation method was easy to understand
	<ul style="list-style-type: none"> ▪ How my contribution to the project affects team performance is clear
Course Instructor	<ul style="list-style-type: none"> ▪ The criteria I used to evaluate my teammates on the peer evaluation was clear ▪ The criteria that my teammates used to evaluate me was clear ▪ I am comfortable evaluating my teammates ▪ I am comfortable being evaluated by my teammates ▪ My ratings of my teammates is kept confidential by the instructor ▪ Other student's ratings of me are kept confidential by the instructor ▪ How my instructor uses this information is clear ▪ The feedback I received resulting from the peer evaluation was helpful

	<ul style="list-style-type: none"> ▪ I will use the feedback from the peer evaluation to improve my performance in future team efforts. ▪ Generally speaking, I enjoy working with groups on course related activities ▪ Overall my experience with peer evaluation in EG10111 was positive
--	--

Results

The descriptive statistics, including mean and standard deviation, are shown for each instrument variable in Table 2. The results are shown for all students and also for each instructor.

Table 2. Descriptive Statistics Mean and Standard Deviation Separated by Method & Instructor

Question	ALL		Instructor 1		Instructor 2	
	On-Line Mean / SD	Paper Pencil Mean / SD	On-Line Mean / SD	Paper Pencil Mean / SD	On-Line Mean / SD	Paper Pencil Mean / SD
3 : Time to complete peer evaluation	11.82 / 4.14	11.18 / 3.21	12.01 / 4.51	11.86 / 3.60	11.63 / 3.73	10.57 / 2.71
4 : Preferred Method Course Materials 1: On-line / 2: Written	1.32 / 0.469	1.45 / 0.50	1.29 / 0.458	1.47 / 0.502	1.35 / 0.479	1.42 / 0.497
5 : Simple	4.25 / 0.642	4.23 / 0.653	4.21 / 0.668	4.16 / 0.661	4.30 / 0.615	4.28 / 0.644
6 : Easy	4.15 / 0.680	4.26 / 0.680	4.03 / 0.713	4.17 / 0.728	4.28 / 0.624	4.33 / 0.628
7 : Contribution effect on team performance	3.89 / 0.751	3.97 / 0.823	3.82 / 0.772	3.90 / 0.856	3.98 / 0.723	4.02 / 0.793
8 : Criteria to evaluate clear	3.84 / 0.782	3.95 / 0.843	3.75 / 0.760	4 / 0.790	3.94 / 0.795	3.91 / 0.889
9 : Criteria being evaluated on are clear	3.8 / 0.789	3.8 / 0.924	3.70 / 0.752	3.85 / 0.903	3.91 / 0.815	3.75 / 0.944
10 : Comfortable evaluating	3.68 / 0.984	3.77 / 1.02	3.71 / 0.946	3.81 / 0.900	3.66 / 1.02	3.74 / 1.12
11 : Comfortable being evaluated	3.91 / 0.809	3.84 / 0.953	3.92 / 0.682	3.91 / 0.888	3.91 / 0.993	3.77 / 1.01
12 : My evaluation kept confidential	4.3 / 0.792	4.47 / 0.657	4.13 / 0.862	4.46 / 0.616	4.47 / 0.676	4.48 / 0.693
13 : Others evaluations kept confidential	4.25 / 0.812	4.47 / 0.657	4.10 / 0.838	4.41 / 0.612	4.40 / 0.760	4.52 / 0.693
14 : Instructor usage clear	3.34 / 0.996	3.41 / 1.12	3.21 / 0.934	3.35 / 1.05	3.46 / 1.05	3.46 / 1.18
15 : Feedback helpful	3.43 / 0.916	2.95 / 0.910	3.39 / 0.804	2.87 / 0.895	3.47 / 1.02	3.02 / 0.915
16 : Will use feedback	3.77 / 0.789	3.63 / 0.963	3.75 / 0.747	3.61 / 0.953	3.79 / 0.832	3.65 / 0.976
17 : Enjoy group activities	3.92 / 0.978	3.94 / 1.02	3.85 / 1.01	3.96 / 1.00	3.98 / 0.947	3.92 / 1.04
18 : Overall experience	3.96 / 0.728	3.79 / 0.934	3.94 / 0.768	3.87 / 0.858	3.97 / 0.690	3.71 / 0.995

T-test analyses were conducted on the sample means for each question by method and for each instructor in order to assess whether differences in the means were statistically significant. The results of this analysis are summarized in Table 3. These results lead to a number of observations about the differences between the two instruments as a whole and differences in the administration of the two instruments by the two instructors.

Table 3. T-test statistics of sample means by Method and Instructor

Question	ALL		Instructor 1		Instructor 2	
	CATME Mean	Paper-Pencil Mean	CATME Mean	Paper-Pencil Mean	CATME Mean	Paper-Pencil Mean
3 : Time	11.820	11.175	12.009	11.862	11.630	10.566
ttest by instructor	-2.303		-0.675	-2.897	-0.675	-2.897
ttest by method	1.768		0.255		2.382*	
4 : Preferred Method Course Materials 1: On-line / 2: Written	1.320	1.445	1.294	1.468	1.352	1.425
ttest by instructor	0.282		0.916	-0.616	0.916	-0.616
ttest by method	-2.586**		-2.591**		-1.089	
5 : Simple	4.250	4.225	4.211	4.160	4.296	4.283
ttest by instructor	1.621		0.978	1.337	0.978	1.337
ttest by method	0.448		0.550		0.154	
6 : Easy	4.150	4.255	4.028	4.170	4.278	4.330
ttest by instructor	3.183***		2.750**	1.668	2.750**	1.668
ttest by method	-1.544		-1.408		-0.612	
7 : Contribution effect on team performance	3.890	3.965	3.817	3.904	3.981	4.019
ttest by instructor	1.861		1.624	0.983	1.624	0.983
ttest by method	-0.861		-0.768		-0.361	
8 : Criteria to evaluate clear	3.840	3.950	3.752	4.000	3.944	3.906
ttest by instructor	0.731		1.820	-0.789	1.820	-0.789
ttest by method	-1.283		-2.275*		0.336	
9 : Criteria being evaluated on are clear	3.800	3.800	3.697	3.851	3.907	3.755
ttest by instructor	0.755		1.975*	-0.735	1.975*	-0.735
ttest by method	0.022		-1.324		1.267	
10 : Comfortable evaluating teammates	3.680	3.770	3.706	3.809	3.657	3.736
ttest by instructor	-0.585		-0.366	-0.501	-0.366	-0.501
ttest by method	-0.896		-0.786		-0.534	
11 : Comfortable being evaluated	3.910	3.840	3.197	3.915	3.907	3.774
ttest by instructor	-0.870		-0.091	-1.046	-0.091	-1.046
ttest by method	0.839		0.023		1.014	
12 : My evaluation is confidential	4.300	4.470	4.128	4.457	4.472	4.481
ttest by instructor	2.743**		3.267***	0.254	3.267***	0.254

ttest by method	-2.381*		-3.083**		-0.095	
13 : Other evaluations are confidential	4.250	4.470	4.101	4.415	4.398	4.519
ttest by instructor	2.909**		2.735**		1.118	2.735**
ttest by method	-3.041**		-3.006**		-1.213	
14 : Instructor usage clear	3.340	3.410	3.211	3.351	3.463	3.462
ttest by instructor	1.807		1.873	0.699	1.873	0.699
ttest by method	-0.709		-1.004		0.005	
15 : Feedback helpful	3.430	2.950	3.385	2.872	3.472	3.019
ttest by instructor	1.083		0.698	1.142	0.698	1.142
ttest by method	5.358***		4.302***		3.424***	
16 : Will use feedback	3.770	3.630	3.752	3.606	3.787	3.651
ttest by instructor	0.405		0.324	0.326	0.324	0.326
ttest by method	1.624		1.222		1.098	
17 : Enjoy group activities	3.920	3.935	3.853	3.957	3.981	3.915
ttest by instructor	0.481		0.966	-0.292	0.966	-0.292
ttest by method	-0.183		-0.736		0.488	
18 : Overall experience	3.960	3.785	3.945	3.872	3.972	3.708
ttest by instructor	-0.856		0.275	-1.247	0.275	-1.247
ttest by method	2.124*		0.636		2.265*	

(* denotes $p < 0.5$, ** denotes $p < 0.01$, *** denotes $p < .001$)

Overall Differences Between CATME and Paper-Pencil:

The most significant finding of this survey was that the students rated the feedback from the on-line instrument as more useful than the feedback from the paper-pencil survey. This result is expected as only the on-line instrument provided formal feedback (other than a grade) to the student, the logistics of the paper-and-pencil instrument make it a challenge to provide timely feedback. It is encouraging that the students recognize the increased feedback, particularly since the on-line instrument is substantially easier to administer (as indicated in the faculty description of the administration of these methods). Hence, it is clear that there is no loss of information to the student in moving from this particular paper-pencil instrument to this particular on-line instrument.

The survey also indicated a statistically significant difference in response to the question regarding whether the students preferred course material to be administered on-line versus via paper. Interestingly, the students who used the on-line instrument provided a greater preference for on-line administration of course material despite the fact that the only difference in the administration of the course when the surveys were collected was the on-line versus pencil-paper peer evaluation. Further, the two groups were not provided a formal period for comparison of these peer-evaluation methods prior to completing their surveys. This result appears consistent with the observation that the students completing the on-line instrument were more satisfied with the overall peer evaluation process. Additional study would be required to determine if results for these two questions are related and to determine the reasons for the higher satisfaction, in general, with on-line presentation of course materials.

A third significant overall difference between the two instruments is observed in the area of confidentiality. Consistent with the work of both Layne et al. and Dommeyer et al. who argued for a difference in student confidence in on-line versus paper-pencil instruments in terms of confidentiality, these results indicate that the students appear to have greater confidence in the confidentiality of the paper-pencil instrument than the on-line instrument. While this result is of some concern in terms of adopting the on-line instrument, it is noted that students assume a high level of confidentiality for both instruments. So, while the difference is interesting, we do not believe it imposes a serious limitation on moving to the on-line instrument.

Differences Between Responses As Identified By Instructor:

A number of differences were observed in the responses obtained when separated by instructor. For example, student responses to the question related to ease of use of the two instruments demonstrate that the students who had instructor 2 provided an overall higher (easier) rating to the instruments than those who had instructor 1. This difference can be tied directly to the student opinion of ease of use of the on-line instrument (where the difference between instructors is significant at the 99% level of confidence). Hence, there appears to be a significant difference in how the two instructors prepared the students for the instruments and/or for the survey.

This difference between instructors is reflected in other responses. For example, the students who had instructor 2 indicated statistically higher confidence in the confidentiality both of their own responses (regarding the other students) and the responses of their team members (other student's evaluation of the responding students). As noted above, this difference can be tied directly to a difference in confidence relative to the on-line instrument. A statistically significant difference in time required to complete the peer evaluations was also observed in an increased time committed to the on-line instrument for instructor 2. CATME users reported a higher recognition of the criteria used to evaluate their performance in instructor 2's class compared to instructor 1's. Finally, the students who had instructor 2 had a statistically significant preference for the on-line instrument in terms of overall satisfaction with the peer evaluation experience (this difference was not present for instructor 1).

Each instructor was responsible for administering the peer evaluation instruments and the survey to their respective sections, thus opening the possibility of slightly different information or delivery of that information to students. While the full reasons for the difference in responses based on instructor is unknown, an initial decomposition of the variance between and within instructor groups shows that over 99% of the variance for overall experience is within instructor groups (and not between instructor groups). Thus, while there is evidence that the instructor affects the outcome of the student experience (as would expect), the researchers were still able to capture differences that were independent of the instructor.

Preference for On-Line versus Paper Course Administration:

One of the survey questions asked students, "In general, what is your preferred method for completing course related tasks (Example: Homework, Quizzes, Surveys, etc.)?" The

objective of this question was to determine whether students might have a predisposition towards an on-line or a paper-pencil instrument. Controlling for response to this question, the overall student satisfaction with the peer evaluation process was evaluated.

62% of the students reported a preference towards completing course materials on-line versus 38% that prefer paper materials. Table 4 shows the percentage breakdown in each category ranking for students in both CATME and paper-pencil groups. While the relative distributions between CATME and paper-pencil administrations are not the same, they both follow the same trend. Specifically, the largest percentage of students fell into the 'agree' category, indicating that they had a positive experience overall with peer evaluation. It is interesting to note that, although a small percentage of all students, students that had a negative experience (rating of disagree or strongly disagree) in the process were far more likely to be students who completed the paper-pencil instrument while indicating a preference for on-line administration of course materials. Dissatisfied students were more than three times as likely to be in that group than any other grouping of preference and administration.

Table 4. Contingency Table for Student's Overall Experience Controlling for Preference

Question: Overall my experience with peer evaluation in EG10111 was positive	CATME		Paper-Pencil	
	Student Preference: On-line	Student Preference: Written	Student Preference: On-line	Student Preference: Written
Strongly Agree	34 (23.1%)	12 (17.1%)	24 (21.6%)	18 (20.2%)
Agree	85 (57.8%)	38 (54.3%)	52 (46.8%)	44 (49.4%)
Neutral	24 (16.3%)	17 (24.3%)	20 (18.0%)	24 (27.0%)
Disagree	4 (2.72%)	3 (4.3%)	10 (9.0%)	3 (3.4%)
Strongly Disagree	0 (0%)	0 (0%)	5 (4.5%)	0 (0%)
	147	70	111	89

Discussion

Two significant observations were made based on the student survey results. First, students using the CATME instrument indicated a significantly higher rating for the helpfulness of the feedback. Additionally, the overall experience of students using the on-line instrument was higher than for those students using the paper-pencil instrument. Thus, there is confidence that the new instrument will be viewed as equal to or better than the previous instrument. Second, the results indicated that in some survey responses there was a statistically significant difference in response based on instructor despite efforts to eliminate instructor as a variable of interest in this study. Specifically, an announcement was made during a lecture to all students early on in the project to explain that peer evaluation would be a course requirement. The announcement also indicated that there would be different methods used, and the results of their feedback would drive the decision for future evaluations. When it came time to administer the peer evaluations

within sections of ~30 students, the same Powerpoint slides were used by both instructors to explain the purpose and procedure for each instrument. For the students using the CATME instrument, all students, regardless of instructor, had the same four day window in which to go on-line and complete the evaluation. The mean grade assigned for students completing both peer evaluation instruments was normalized such that they had the same mean score. One must conclude, therefore, that in courses involving large student groups and multiple sections with different instructors, the reception by students of a particular peer evaluation instrument may be influenced by the instructor.

Gathering responses from as many students as possible on their impressions of the process was imperative to have a study with high statistical power. To encourage participation in the follow-up survey, students were given homework credit for completing the survey. The initial concern with this approach was that student ratings of confidentiality may be lowered overall because they were required to turn in a form that says that they completed the survey. This risk was viewed as relatively low and necessary to ensure a high response rate. In the end, the biggest threat to validity was that the students were not locked out of the follow up survey after initially completing it. It was administered by making a link available on the course website for a set period of time. This was done to avoid the risk of student perception of non-anonymity if they had to go through a login procedure to their overall responses on questions of confidentiality. As a result, a larger number of students responded to the survey than were enrolled in the course (approximately 16% additional responses received). This likely happened because the students had to print out the last page of the survey, the “Thank you for your participation” page, to get homework credit (the students were told in advance that this would be required). The intention was to ensure that students actually filled out the survey, but if a student completed the survey yet did not print the completion page at that time they completed the survey, they may have retaken the survey to print the sheet for homework credit. This lack of foresight on the administrator’s part did compromise the results. The survey results were reviewed to ensure that there were there were no responses that were same box checked for all questions, and that was not the case, so it appears that students in effect simply retook the survey with the same responses so some results were counted twice. We have no reason to suspect that this duplication of some student’s survey results would have been biased toward either administration method, but we are unable to document the true scope of the anomaly. Based on the number of students who received credit for the homework assignment, we calculate the duplication rate to be between 15-21%. Again, having no reason to suspect that a bias was introduced, we consider the results valid for the analyses provided.

In addition to these primary observations, the results showed consistency with previous studies^{13,14}, in demonstrating that students perceive the paper-pencil instrument to be more confidential than CATME / on-line instrument. This was again somewhat surprising considering the administration of the paper-pencil instrument took place in the class room (~30 students present), whereas the on-line instrument was performed at the student’s choice of location and time. This may point to an increasing level of distrust for on-line systems, a topic for future research. However; given that the questions on confidentiality were based on a Likert scale and responses for all instructors and administration methods

had mean values in the range of 4=Agree to 5= Strongly Agree, these results indicate that students collectively trust the faculty members to keep their responses confidential.

Conclusions

Referring back to research questions posed, the researchers found the following:

- Which instrument do students perceive to be simpler (straight forward and easy)?
 - No statistically significant difference observed
- With which instrument are students more comfortable (at ease or content)?
 - No statistically significant difference observed
- Which instrument do students perceive to ensure a higher confidentiality? Do students express any concern regarding confidentiality of their responses?
 - The paper-pencil instrument, by a statistically significant margin
 - No concerns expressed, and mean ratings for both instruments were high
- For which instrument do students perceive the feedback from peer evaluation to be more useful?
 - The CATME instrument, by a statistically significant margin.
- For which instrument do students report a more positive overall experience?
 - The CATME instrument, by a statistically significant margin.

This study was conducted to determine whether there was a significant difference in student response to the paper-pencil versus the on-line peer evaluation instruments. The purpose of this study was to determine whether it is reasonable for the University of Notre Dame to adopt the CATME system in future offerings of its first-year engineering courses. From the standpoint of the course instructors, the on-line instrument represents a significant savings in time and effort. Specifically, discussion with the two faculty members administering the peer evaluations confirmed that the administrative burden was significantly reduced using the CATME system. Hence, there are substantial administrative reasons for considering the change in peer evaluation instrument if the new instrument provides similar information and is well received by the students. The gain in student perceptions of useful feedback should improve the accuracy of student ratings, which is an opportunity for a gain for student learning as well.

Given the overall results of this study, combined with confidence expressed by the instructors in the course in the new instrument, a decision has been made to adopt CATME in future offerings of the first-year engineering course (starting spring, 2008) at the University of Notre Dame.

Appendix A:
 Comprehensive Assessment of Team Member Effectiveness (CATME)
<https://engineering.purdue.edu/CATME>

Contributing to the Team's Work

Student 1				Description of Rating
Student 2				
Student 3				
Student 4				
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<ul style="list-style-type: none"> Does more or higher-quality work than expected. Makes important contributions that improve the team's work. Helps teammates who are having difficulty completing their work.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Demonstrates behaviors described above and below.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<ul style="list-style-type: none"> Completes a fair share of the team's work with acceptable quality. Keeps commitments and completes assignments on time. Helps teammates who are having difficulty when it is easy or important.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Demonstrates behaviors described above and below.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<ul style="list-style-type: none"> Does not do a fair share of the team's work. Delivers sloppy or incomplete work. Misses deadlines. Is late, unprepared, or absent for team meetings. Does not assist teammates. Quits if the work becomes difficult.

Interacting with Teammates

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<ul style="list-style-type: none"> Asks for and shows an interest in teammates' ideas and contributions. Makes sure teammates stay informed and understand each other. Provides encouragement or enthusiasm to the team. Asks teammates for feedback and uses their suggestions to improve.
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Demonstrates behaviors described above and below.

				<ul style="list-style-type: none"> Respects and responds to feedback from teammates. Participates fully in team activities. Communicates clearly. Shares information with teammates. Listens to teammates and respects their contributions.
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<ul style="list-style-type: none"> Demonstrates behaviors described above and below.
				<ul style="list-style-type: none"> Interrupts, ignores, bosses, or makes fun of teammates. Takes actions that affect teammates without their input. Does not share information. Complains, makes excuses, or does not interact with teammates. Is defensive. Will not accept help or advice from teammates.
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

Keeping the Team on Track

				<ul style="list-style-type: none"> Watches conditions affecting the team and monitors the team's progress. Makes sure that teammates are making appropriate progress. Gives teammates specific, timely, and constructive feedback.
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<ul style="list-style-type: none"> Demonstrates behaviors described above and below.
				<ul style="list-style-type: none"> Notices changes that influence the team's success. Knows what everyone on the team should be doing and notices problems. Alerts teammates or suggests solutions when the team's success is threatened.
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<ul style="list-style-type: none"> Demonstrates behaviors described above and below.
				<ul style="list-style-type: none"> Is unaware of whether the team is meeting its goals. Does not pay attention to teammates' progress. Avoids discussing team problems, even when they are obvious.
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

Expecting Quality

				<ul style="list-style-type: none"> Motivates the team to do excellent work. Cares that the team does outstanding work, even if there is no additional reward.
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

				<ul style="list-style-type: none"> Believes that the team can do excellent work.
☐	☐	☐	☐	Demonstrates behaviors described above and below.
☐	☐	☐	☐	<ul style="list-style-type: none"> Encourages the team to do good work that meets all requirements. Wants the team to perform well enough to earn all available rewards. Believes that the team can fully meet its responsibilities.
☐	☐	☐	☐	Demonstrates behaviors described above and below.
☐	☐	☐	☐	<ul style="list-style-type: none"> Satisfied even if the team does not meet assigned standards. Wants the team to avoid work, even if it hurts the team. Doubts that the team can meet its requirements.

Having Related Knowledge, Skills, and Abilities

☐	☐	☐	☐	<ul style="list-style-type: none"> Demonstrates the knowledge, skills, and abilities to do excellent work. Acquires new knowledge or skills to improve the team's performance. Able to perform the role of any team member if necessary.
☐	☐	☐	☐	Demonstrates behaviors described above and below.
☐	☐	☐	☐	<p>Demonstrates sufficient knowledge, skills, and abilities to contribute to the team's work.</p> <ul style="list-style-type: none"> Acquires knowledge or skills as needed to meet requirements. Able to perform some of the tasks normally done by other team members.
☐	☐	☐	☐	Demonstrates behaviors described above and below.
☐	☐	☐	☐	<ul style="list-style-type: none"> Missing basic qualifications needed to be a member of the team. Unable or unwilling to develop knowledge or skills to contribute to the team. Unable to perform any of the duties of other team members.

Appendix B:

Written Evaluation Form Used from 2005-2007

**EG10111: Introduction to Engineering Systems II
Peer Review Form**

Your Name: _____ Section #: _____ Group #: _____ Date: _____

Please fill out the table below in evaluation of your other 3-4 group members. Please feel free to be honest as these forms will remain confidential. On a scale of 1-10 please

Name	Technical Contribution to Project	Ability to work w/ others in the group	Ability to discuss / adapt to new ideas	Was a positive influence in group discussion / decisions	Comments / Justification	Rank order (1 st – 4 th)

assign a score to each team member according to your view of their contributions / ability WITH RESPECT TO THIS PROJECT. (10 being Super Hero Status & 1 Did Not Contribute at All)

Self Evaluation: (Please comment on your contribution to this project overall. How did you fit / work with the group? Did you feel that you were able to contribute appropriately to the group's efforts?)

Other Course / Project Comments: (Ideas for improvement? Use the back if needed.)

Appendix C: Peer Evaluation Administration Procedure

Pre Administration Questions:

Documentation of procedure / written instrument:

- What is involved with data collection?

Instructor 1: Students are asked to complete a peer evaluation form asking them to score (1 – 10) each of their group members in the areas of technical contribution, ability to work with others, ability to discuss/adapt to new ideas and were they a positive influence in group discussion/decisions. An area for commenting/justification is provided as well. Each student is asked to rank order their group members. The form has space for students to provide a self evaluation and to provide other course/project comments. Class time is usually set aside for the students to complete the form and students are asked to not complete the form in the near vicinity of other group members.

Forms are sorted by group.

Instructor 2: The course instructors created a written sheet (instrument) that has been used for 4 semesters on all EG10111/ 2 students for peer evaluation. At the conclusion of each project (after all of the reports / presentations) have been completed and turned in, we devote 10 minutes in the main lecture for students to fill out the forms. This is not ideal b/c students may be sitting near other group members or feel like people are looking at their evaluations. Students are told that the evaluations are confidential and to feel free to rate their peers appropriately. After all of the sheets are filled out in class, a TA for the class sorts them by section number and group number and provides them to the appropriate faculty members.

- Once collected, how do you use data?

Instructor 1: Data are used to provide a numerical score for each group member. The project/comments section is used as an additional method for determining how well the module and/or course is progressing.

Instructor 2: Once the written evaluations were collected, I would read through the evaluations (one team from a given section at a time). Comments tended to be pretty consistent, so for the most part you would get similar rankings of team members. The ranking is really the part of the evaluation that I look closest at b/c it is an easy way to compare for consistency across group members. I would also look at the justification of the score / ranking.

- Does reported data ever result in meeting / contacting groups?

Instructor 1: I have not initiated any contact with a group based on the peer evaluations. However, individual students have requested meetings to discuss their assigned score from the peer evaluation.

Instructor 2: I have never called in a group after reading the evaluations. There are usually 1-2 students per project that will want to come in and discuss how their project group functioned and what their role was to the completion of that project. This usually occurs right before or right after the evaluations were completed in class. And it is virtually always someone that was in a group that was dysfunctional (although not reported previously). There are occasionally situations where a group is very clearly dysfunctional throughout the course of the project, and in those instances I have called groups in just to smooth things over and make sure they can get through the project.

- What is your method for assigning scores?

Instructor 1: The scores for each student are averaged with a possible adjustment based on the comments/justification, rankings and self evaluations.

Instructor 2: 10% of the student's project grade is based on this peer evaluation. For the most part,

the scores are in the 90-100 range (although there are always some in the 80's as well). Most groups report that their group worked well together and when that happens and all the evaluations are consistent then everyone gets a high score (all 100's or high 90's). When it is clear from rankings and comments that there is a slight difference in the work put into the project scores are a little more varied. Very low scores (20-60) do also occur when group members report that a member wasn't showing up or didn't complete their portion of the project.

- How much time do you spend reading evaluations / assigning a score to each group?

Instructor 1: It takes on average 8 minutes/group.

Instructor 2: It takes me 40-50 minutes per section to assign grades for the evaluations. There are 6 groups per section, so that breaks down to 6-9 minutes per group.

Post Administration Questions:

Document new procedure / CATME, on-line:

Faculty Members 1&2 worked collaboratively on this process.

- What is involved with data collection?

Student e-mail addresses and groups had to be entered into CATME.

Students had to create a password / login to the system.

Set open / close dates on the survey.

- Once collected, how do you use data?

The data was used to assign grades to each student.

The grades were normalized to have the same average as the paper-pencil system.

The feedback was released to the students (how they performed / rated themselves compared to their teammates) with the click of a button.

- Does reported data ever result in meeting / contacting groups?

This administration did not have any reported issues, so no groups were called in.

- What is your method for assigning scores?

Scores were scaled to 90% (average score from all previous paper-pencil peer evaluation administrations)

- How much time do you spend reading evaluations / assigning a score to each group?

The time spent was collectively under 1 hour (between 2 faculty members), and in the future it would likely be less than that.

References

1. CATME: Comprehensive Assessment of Team Member Effectiveness. Retrieved 1/3/2008, 2008, from <https://engineering.purdue.edu/CATME>
2. Savery, J. R., & Duffy, T. M. (1995). Problem Based Learning: An Instructional Model and its Constructivist Framework. *Constructivist Learning Environments: Case Studies in Instructional Design, Educational Technology Publications*, 1-10.
3. Panitz, T. (1999). *Collaborative versus Cooperative Learning: A Comparison of the Two Concepts Which Will Help Us Understand the Underlying Nature of Interactive Learning*.
4. Smith, K. A. (1995). *Cooperative Learning: Effective Teamwork for Engineering*. Paper presented at the Frontiers in Education Conference. from http://ieeexplore.ieee.org.lib-proxy.nd.edu/xpls/abs_all.jsp?arnumber=483059.
5. Kauffman, D. B., Felder, R. M., & Fuller, H. (2000). Accounting for Individual Effort in Cooperative Learning. *Journal of Engineering Education*, 89, 133-140.
6. Sheppard, S. C., Helen; Schaffer, Evvone; Steinbeck, Reinhold; Neumann, H; Ko, Pey-Jiun. (2004). *Peer Assessment of Student Collaborative Processes in Undergraduate Engineering Education (Final Report)*: National Science Foundation.
7. Gueldenzoph, L. M., Gary. (2002). Collaborative Peer Evaluation: Best Practices for Group Member Assessments. *Business Communication Quarterly*, 65(March 2002), 9-20.
8. Falchikov, N. (1995). Peer Feedback Marking: Developing Peer Assessment. *Innovations in Education and Training International*, 32(2), 175.
9. Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The Effects of Peer Feedback on Team Member Behavior. *Group & Organization Management*, 22(4), 508-520.
10. Jawahr, I. M., & Williams, C. R. (1997). WHERE ALL THE CHILDREN ARE ABOVE AVERAGE: THE PERFORMANCE APPRAISAL PURPOSE EFFECT. *Personnel Psychology*, 50(4), 905-926.
11. Bamberger, P. A. (In Press). Intentional Rating Distortion and Peer Evaluation in Management Education: Why and How to Identify "Game Players". *Journal of Learning in Higher Education*.
12. Bamberger, P. A., Erev, I., Kimmel, M., & Tali, O. (2005). Peer Assessment, Individual Performance, and Contribution to Group Processes: THE IMPACT OF RATER ANONYMITY. *Group & Organization Management*, 30(4), 344-377.
13. Layne, B. H., Decristoforo, J. R., & McGinty, D. (1999). ELECTRONIC VERSUS TRADITIONAL STUDENT RATINGS OF INSTRUCTION. *Research in Higher Education*, 40(2), 221-232.
14. Dommeyer, C. J., Baum, P., & Hanna, R. W. (2002). College Students' Attitudes Toward Methods of Collecting Teaching Evaluations: In-Class Versus On-Line. *Journal of Education for Business*, 78(1), 11.
15. Chang, T.-S. (2005). The Validity and Reliability of Student Ratings: Comparison between Paper-pencil and Online Survey. *Chinese Journal of Psychology*, 47(2), 113-125.
16. Survey Monkey. (1999-2007). Retrieved 1/3/2008, 2008, from <http://www.surveymonkey.com/>
17. STATA: Data Analysis and Statistical Software. (1996-2008). Retrieved 1/3/2008, 2008, from <http://www.stata.com/>