

PEER TO PEER: A NEW TOOL FOR STUDENT PEER EVALUATION

Christopher M. Scherpereel
Northern Arizona University
chris.scherpereel@nau.edu

ABSTRACT

Peer evaluations have a long history in the assessment of individual contribution to team performance in both experiential exercises and simulation games. Experience with different peer evaluation methods has revealed a number of significant implementation challenges including the structure of peer evaluations and the proper use of data collected. Many of these challenges have been discussed among the ABSEL membership; however, less attention has been given to the development of tools facilitating the collection of peer evaluation data and the feeding back of relevant data. The peer-to-peer assessment and feedback tool presented in this paper highlight the significant progress made toward the development of an Excel-based peer evaluation tool that can be easily and effectively implemented in team-based learning environments. Preliminary data is presented that shows students were generally favorable to both the peer evaluation input methods and the feedback they received.

INTRODUCTION

This paper offers a preliminary analysis of a student peer evaluation instrument currently under development in a large western business school. The instrument is based on best practices identified in a review of the business, engineering, and team management literature. Development has resulted in a prototype Excel-based tool that was beta tested in fall 2009. In the process of testing the peer evaluation instrument, data were collected confirming the usability and value of the peer evaluation process. The positive results obtained during beta testing have encouraged the author to present these preliminary results to the ABSEL membership. Discussion of progress in this area should be valuable to those interested in peer evaluation in team-based experiential learning. ABSEL also presents an opportunity to solicit feedback from experienced colleagues in both the design and implementation of this Excel-based tool, as beta testing, revisions, and additional data collection is expected to continue throughout the next several years.

A major international accrediting body for business schools (AACSB) highlights the growing pressure for team-based pedagogy.

The most effective learning takes place when students are involved in their educational experiences. Passive learning is ineffective and of short duration. Faculty

members should develop techniques and styles that engage students and make students responsible for meeting learning goals. Many pedagogical approaches are suitable for challenging students in this way – problem-based learning, projects, simulations, etc... Faculty members should find such approaches that are suited to their subject matter, and should adopt active learning methodologies (AACSB, 2005, p. 57).

The development of peer evaluation tools that are easy to administer is vital to effectively managing the data collection and feedback requirements of implementing the recommended team-based experiential pedagogy.

Due to accreditation pressures, many business schools are seeking opportunities to incorporate teamwork and team assignments into the curriculum (Hansen, 2006). Teamwork in an academic setting generally involves projects, case studies, practice sets, computer simulations, and presentations. Overall, the use of team-based learning is considered by educators as beneficial to students. Typical benefits identified in the literature include the opportunities to develop soft skills (written and oral communication, negotiation, conflict resolution) and to apply theory and concepts to promote a deeper understanding of discipline-specific topics. These benefits have been examined and confirmed in numerous studies (Price, 2004). While the value of incorporating teamwork into business curriculum is extensively researched and documented, the development of appraisal methods to assess individual performance (peer evaluation) in team environments has not been as extensively examined (Cederblom & Lounsbury, 1980; Hansford & Hattie, 1982; Morahan-Martin, 1996).

In addition to accreditation pressures, most business educators recognize that better team skills are correlated to future career success. Teams are now considered an essential component of the workplace (Acona, 1990) and many organizations design jobs specifically for teams (Feder, Battenhausen, & Davis, 1999). As a result, there has been an increasing interest in assessing team member performance and outcomes in business (Church & Bracken, 1997). In the business world “peer evaluation has become popular as an alternative to traditional performance appraisal. Research has described peer evaluation’s growing use, its potentially high validity and some of its related psychological mechanisms” (Peiperl, 1999). The trend toward greater use of peer evaluation in business should encourage a response by academia to provide students with the opportunity to give, receive, and implement peer evaluations.

Peer evaluations have long history in the assessment of experiential exercises and simulation games. Gentry et.al. (2003) highlighted the use of peer evaluations as the standard method for assessing individual contribution to team performance. However, classroom experience has exposed a number of significant implementation challenges including the structure of peer evaluations and the proper use of collected data. While these challenges have been discussed at ABSEL, tools to facilitate the collection of peer evaluation data and the feedback of relevant data has received less attention.

GOALS AND OBJECTIVES

The goal of this work is to develop an easy-to-use, valid peer evaluation system to be used in team- based classes. To achieve this goal, I have establish two objectives: 1) to develop a peer evaluation instrument that can be easily and effectively implemented in team-based learning environments and 2) to develop a process that can be used to teach students how to give, receive, and implement feedback from peer evaluations. This paper highlights the significant progress that I have made toward the first of these objectives and offers the mechanisms necessary to facilitate the second objective.

“There is no question that peer feedback can be very valuable to students and that learning how to give and take it is a crucial lifelong skill” (Nilsen & Campbell, 1993, p. 37). It is also known that peer feedback is vital to faculty trying to coach behavioral change in individual team members and offers the best method for evaluating individual contributions to team performance. According to published research, developing an effective peer evaluation instrument and process can be a daunting task. Best practice literature suggests that this effort will require “(1) building a foundation in the classroom that supports collaborative evaluation, (2) creating effective evaluation tools by articulating specific criteria and ensuring honest student participation, (3) implementing formative feedback during the collaborative experience, (4) formulating summative feedback at the conclusion of the experience, and (5) assessing the collaborative evaluation process” (Gueldenzoph & May, 2002, p. 9). Anecdotal evidence suggests there is frustration with implementing and using peer feedback to effectively change behavior (improve individual performance) and evaluate individual contributions to team performance (Topping, 1998). Thus, to achieve what Gueldenzoph and May (2002) suggest, an instrument that can be easily and effectively implemented by the faculty is needed.

PEER EVALUATION DESIGN

DeNisi et.al. (1983) clearly defines peer ratings “as the set of evaluations obtained by having each member rate every other member of a work group, using a specific set of

rating scales” (p. 457). Thus, I will define peer evaluation as the process of using peer ratings to generate information regarding individual performance pertaining to a member of a work group (team). The information generated from these peer evaluations can either be used for formative or summative purposes. Formative peer evaluation is designed to contribute to student learning by providing information about student learning while summative peer evaluations are primary used to determine the extent to which an individual who is part of a work group has achieved certain curricular objectives (Yorke, 2003). Most evaluations can be used for both purposes depending on when they are used in the course. For example, if peer evaluations are used early during the task/project of a particular work group and feedback is given, it is often for formative purposes (to change or encourage behaviors). In contrast, peer evaluations administered at the conclusion of group task/project are more often used for summative purposes (to assign credit for individual contributions).

From the above discussion it is apparent that these two purposes (formative and summative) will drive the design of peer evaluation tool. It is also clear that there will be two parts to the design of a peer evaluation tool: data collection and data feedback. Both these design elements will be discussed in the following sections.

PEER EVALUATION TOOL DATA COLLECTION

The peer evaluation tool being developed was designed to collect data for both formative and summative purposes. Thus, the collection of both qualitative and quantitative information was desirable. Given the difficulty the authors had with previous peer evaluation data collection methods, there was a strong motivation to design for ease of administration. A review of the peer evaluation literature was conducted to identify the best methods for soliciting evaluation information from students and it was decided to include as many of these methods as was feasible. Paper and pencil peer evaluations were ruled out as being too difficult to administer and experience with computer based peer evaluation tools such as *Team Developer* (McGourty & DeMeuse, 2001) was not positive. While these computer-based tools were easy to administer, they were insufficient at collecting and feeding back qualitative information. Thus, it was determined that a new tool would need to be developed.

Excel was chosen as the platform on which to build the peer evaluation tool primarily because of its availability to student respondents and its ease of programming. The design was based on ideas commonly used in business simulation games with students completing forms and Excel processing those forms by extracting and aggregating the data. The student forms contained a variety of instruments designed to collect both quantitative and qualitative input from students. It was decided to include several different commonly used methods for collecting these data, including open-ended questions, summated type questions, forced

ended question is designed to generate responses regarding behaviors that the participant would like to encourage in each team member: The second question is: "If you were this team member, what would you identify as your most significant contributions to the team (This information may be released to team members without disclosing your identity)." In both cases, it is made clear that the information will anonymously be fed directly back to the individual being evaluated.

The next rating instrument used is an allocation ranking. Allocation ranking is often used for summative evaluation because it provides information on the percentage contribution to the team's deliverable. However, it can also provide valuable formative feedback regarding how one is performing relative to others on the team. Specifically, the evaluator is challenged with the following problem: "You have been given \$100 to pay your team for their contribution to the team's project. Please allocate a portion of these funds to each team member (consider both performance and effort)." Unlike the forced ranking there is nothing to direct how the allocation is assigned and the evaluator is free to allocate these funds equally to all team members (Figure 3).

The peer evaluation concludes with two additional open-ended questions. The first is a self evaluation question, allowing the evaluator to state their level of contribution to the team. The question asks; "Briefly describe YOUR significant contributions to the team's performance. (This information may be released back to the entire team)." Data collected from this question is intended to allow the evaluator to contend that they were in fact a significant contributor to the team. Feeding this information back to the entire team provides some accountability to the

response because the evaluator knows that those evaluating him/her will be reading this response. This question is especially effective if the evaluations are done several times during the team's existence.

The last question is simply an opportunity for the evaluator to express confidential information that they would like to communicate to the instructor. It also allows the evaluator to be less guarded in their response and perhaps express opinions not captured by the previous questions. Using Excel allows extensive error checking to be programmed into the peer evaluation form. Students are restricted from submitting forms that are incomplete due to missing responses. Excel also allows some evaluation of the responses in terms of word counts and variance, making it possible to do some analysis of the quality of peer feedback being generated. Although this capability is recognized by the authors, more research is necessary to both strengthen the analysis and assess validity.

PEER EVALUATION TOOL DATA FEEDBACK

Ultimately, the purpose of collecting the peer evaluation data is to generate data that can be used for formative and summative purposes. Once the students complete and submit the peer evaluation forms (this is done electronically through Vista Blackboard), a Visual Basic enabled Excel spreadsheet is used to process the individual forms into a master spreadsheet containing all the relevant information. This master spreadsheet is then processed into individual feedback forms and a summary form for the instructor. The instructor form contains a quantitative summary of the feedback given to each individual on each team while the individual feedback forms contain the both

Allocation ranking

Figure 3

Team Member	Allocation								
Student 68				\$100	Total \$ available to allocate				
Student 63				\$0	Total \$ allocated				
Student 49				\$100	Total \$ remaining				

Quantitative Student Feedback

Figure 4

	Performance (judged by group)	Team player	Quantity of Contribution	Quality of Contribution	Effect on Team Performance	Pay Grade (% of average pay)
	Adequate	A-	B+	B+	B+	96%

aggregated quantitative and qualitative information.

While there is extensive research on peer evaluation there is relatively little research on what information can be effectively fed back to teams and is useful for improving individual performance. I aggregated the quantitative items and presented qualitative information without revealing the specific rater (or author or team member) in order to maintain anonymity. Rather than providing a numerical representation of the aggregated information (e.g., averages), I converted these averages to a letter grade. Students understand the meaning of letter grades; thus receiving a C is more meaningful than receiving a 3 (out of 5). An example of the quantitative feedback a student may receive is shown in Figure 4.

The student feedback in Figure 4 provides aggregated information regarding their peers' perception of the student's contribution to the team. It is intended to report

on a number of different dimensions and to motivate behavioral changes. The last column, Pay Grade, gives an indication of where they would be ranked relative to others on the team. In the case of the student shown in Figure 4, a pay grade of less than 100% indicates that this student would be ranked below the average student on the team. This number could be used for summative evaluation purposes to adjust the grade of this student downward. Additional qualitative feedback is provided to offer guidance on what behaviors might be changed to improve or maintain the quantitative scores. The two linked qualitative feedback items provide information on both the student's greatest contribution and where their performance could most be improved. An example of this feedback is shown in Figure 5.

In addition to the qualitative summary, I also provide a compilation of the qualitative self-reported feedback to the

Qualitative Student Feedback Figure 5

Greatest Contribution									
. Matt's most significant contributions are making phone calls to competitors and helping with excel. Matt helps keep the team from hating eachother.. Matt is very outgoing and uses humor to express ideas and motivate the team to complete quality work. . Keeping the group positive and enthusiastic during meetings.. I still believe that the greatest asset you bring to the team is your charisma.. Calling SunButter and finding out useful information.. . .									
Performance Improvement									
. to increase contribution to the team I would try to be a bit more school focused and not party life focused. But you still do a lot of work.. Try to stay on task more often.. Sometimes gets the group off task. Should not be afraid to ask for help when needed.. If I were this team member I might try to sacrifice more outside of class activities for bizblock. Matt definitely does make up for this by telling us all ahead of time about availability and by suggesting meeting times that accommodate more of the group.. You were not as involved lately, however I understand the financials do not play to your strengths.. Contribute more by doing work outside of class, like revising drafts.. . .									

Self Reported Team Feedback Figure 6

Student 1									
My significant contributions to the team's performance includes providing valuable opinions during discussions, as well as weighing pros and cons of ideas that the team comes up with. I go to group meetings ready to work and in a good mood to keep meetings productive. I was involved when the team wrote the mission statement and operating agreement.									
Student 2									
I believe that I am able to provide valuable input and ideas when we are discussing our business plan. I work well with other people and computer skills I think can benefit the team. I also have access to free printing in Phoenix.									

student regarding their individual contribution to the team. (Figure 6) This feedback has two purposes: to allow students to communicate what they believe they did to contribute to the team's performance and to keep this feedback honest (especially in subsequent peer evaluations). This feedback also provides an excellent tool for the instructor to use in counseling teams with a poor performing team member. Conversely, it allows a poor performing student to defend or explain their performance by discussing the contribution they reported to have made.

STUDENT PERCEPTION OF DESIGN

A preliminary study of the peer evaluation design was conducted during the fall of 2009 in a class of 70 undergraduate students. After they had completed the first peer evaluation of their teammates, they were given a survey in which they could evaluate the peer evaluation tool. A total of 69 surveys were returned and used in the evaluation. Two preliminary summated questions were asked to solicit the students' attitudes toward peer evaluation. The first question was: "I consider peer evaluations to be useful in motivating my teammates to IMPROVE their performance." Results showed that 61% of the students agreed with this statement while only 7% disagreed. These results indicate that most students believed that the peer evaluation would be helpful even before receiving any feedback on their performance.

The second question was: "Overall, I was able to communicate all that I desired regarding my teammates' performance." This question was designed to capture the students' perception regarding the capability of the evaluation to comprehensively convey their feedback. Similar results were observed with 62% of the students agreeing with this statement and only 9% disagreeing. These results were encouraging in that most of the students identified value in peer evaluation and thought that the peer

evaluation tool captured most of what they desired to communicate. Thus, it appears that the students would not be predisposed to respond negatively, because of their negative opinion of peer evaluations in general, when evaluating the specific questions in the peer evaluation

Four semantic differential scales were constructed to gather feedback regarding each of the peer rating instruments:

- Easy 1 2 3 4 5 6 7 Difficult
- Clear 1 2 3 4 5 6 7 Unclear
- Useful 1 2 3 4 5 6 7 Useless
- Biased 1 2 3 4 5 6 7 Unbiased

The data collected indicated that the students did not have difficult time understanding and completing the peer evaluation. All rating instruments were judged easy and clear at the 99% confidence level (Table 1). Although rating instruments also appeared useful, I cannot make this claim with equal confidence for all rating instruments. Both the forced ranking and the allocation ranking could not be confirmed at the 99% confidence level for usefulness. It is also interesting to note that the forced ranking was judged to be biased while the similar conclusion could not be confirmed for the allocation ranking. Finally, as intuition would suggest, the qualitative self evaluation question was considered to be biased.

Similar confidence intervals were constructed between rating scales to evaluate which of the scales were easier, clearer, most useful, and most biased. Table 2 clearly shows that the summated rating scales are the easiest to complete while the ranking scales (forced and allocated) are the most difficult to complete. Additionally the ranking scales were viewed as the least useful of the rating methods used by the students. It is unclear why the ranking scales were judged less useful (and less clear) but I can hypothesize that the information collected from these scales would be considered summative rather than formative, and thus, the students might be hesitant to mark these with the same candor as the

Evaluation Confidence Intervals – Difference from neutral (paired t-test)

Table 1

		Summated Questions	Forced Ranking	Qualitative: Performance Improvement	Qualitative: Significant Contributions	Allocation Ranking	Qualitative: Self Evaluation	
Ease of Different Peer Feedback Instruments	mean	2.07	3.38	3.32	2.96	3.36	2.68	
	significance	0.000	0.006	0.001	0.000	0.003	0.000	
Clarity of Different Peer Feedback Instruments	mean	2.04	2.80	2.54	2.35	2.61	2.01	
	significance	0.000	0.000	0.000	0.000	0.000	0.000	
Usefulness of Different Peer Feedback Instruments	mean	2.55	3.71	2.78	2.64	3.51	3.01	
	significance	0.000	0.183	0.000	0.000	0.020	0.000	
Bias of Different Peer Feedback Instruments	mean	4.22	3.33	3.68	3.68	3.74	3.13	
	significance	0.327	0.002	0.113	0.113	0.187	0.000	
	Key:	99.90%	confidence indicated by BOLD					
		99.00%	confidence indicated by BOLD					
		95.00%	confidence indicated by <i>Italics</i>					
		<95.00%	confidence indicated by plain text (these values are not significant)					

Confidence Intervals – Between rating scales (paired t-test)

Table 2

		Key:		99.90%	confidence indicated by BOLD			
				99.00%	confidence indicated by BOLD			
				<i>95.00%</i>	confidence indicated by <i>Italics</i>			
				<95.00%	confidence indicated by plain text (these values are not significant)			
	Mean Score	Summated Questions	Forced Ranking	Qualitative: Performance Improvement	Qualitative: Significant Contributions	Allocation Ranking	Qualitative: Self Evaluation	
Ease of Different Peer Feedback Instruments								
Summated Questions	2.07	1	0.000	0.000	0.000	0.000	0.006	0.006
Forced Ranking	3.38		1	0.823	0.120	0.956	0.012	
Qualitative: Performance Improvement	2.54			1	0.024	0.876	0.003	
Qualitative: Significant Contributions	2.96				1	0.145	0.175	
Allocation Ranking	3.36					1	0.006	
Qualitative: Self Evaluation	2.68							1
Clarity of Different Peer Feedback Instruments								
Summated Questions	2.04	1	0.000	0.012	0.109	0.003	0.872	
Forced Ranking	2.80		1	0.205	0.034	0.303	0.000	
Qualitative: Performance Improvement	2.54			1	0.102	0.708	0.001	
Qualitative: Significant Contributions	2.35				1	0.208	0.020	
Allocation Ranking	2.61					1	0.007	
Qualitative: Self Evaluation	2.01							1
Usefulness of Different Peer Feedback Instruments								
Summated Questions	2.55	1	0.000	0.217	0.634	0.000	0.029	
Forced Ranking	3.71		1	0.000	0.000	0.340	0.004	
Qualitative: Performance Improvement	2.54			1	0.402	0.007	0.275	
Qualitative: Significant Contributions	2.64				1	0.001	0.016	
Allocation Ranking	3.51					1	0.028	
Qualitative: Self Evaluation	3.01							1
Bias of Different Peer Feedback Instruments								
Summated Questions	4.22	1	0.001	0.020	0.016	0.051	0.000	
Forced Ranking	3.33		1	0.048	0.085	0.053	0.286	
Qualitative: Performance Improvement	2.54			1	1.000	0.743	0.013	
Qualitative: Significant Contributions	3.68				1	0.747	0.007	
Allocation Ranking	3.74					1	0.007	
Qualitative: Self Evaluation	3.13							1

other scales. As expected, the students seem to judge the qualitative self evaluation as the most biased, indicating they may be giving somewhat biased feedback (or believe that others will be giving biased feedback) to this question.

STUDENT PERCEPTION OF FEEDBACK

After the students received feedback and had several days to process the feedback, a follow-up survey was conducted to assess the value of the feedback they received. A total of 66 surveys were returned (in the undergraduate class of 70) and used in the evaluation. Two preliminary summated questions were asked to solicit the students' attitudes toward peer feedback. The first question asked: "I like to receive performance feedback from my teammates." Results showed that 79% of the students agreed with this statement while only 7% disagreed. The second question

asked: "Continuing to provide performance feedback to our team members will likely increase our team's performance." Similar responses were received to this question with 76% of students agreeing with this statement while only 7% disagreed. These results indicate that a majority of students see the value in receiving peer feedback and believe that it will increase team performance. Thus, the students do not appear to be predisposed to reject peer feedback and will more likely give a fair evaluation of the specific feedback given by my peer evaluation tool.

Three semantic differential scales were constructed to gather data regarding each of the aggregated feedback instruments:

- Helpful 1 2 3 4 5 6 7 Unhelpful
- Clear 1 2 3 4 5 6 7 Unclear
- Biased 1 2 3 4 5 6 7 Unbiased

From the data collected it is evident that the students viewed all the quantitative feedback as helpful and clear but somewhat biased (Table 3). Exploring these numbers a bit further reveals that Pay Grade was deemed the least helpful and clear. Perhaps the reason for this outcome is that students did not receive a clear explanation of what this number meant. Because the peer evaluation tool is being designed for use in different classes with different instructors, the guidance given with the feedback will likely vary. To avoid introducing any instructor bias in this beta test, there was limited discussion with the students regarding the feedback. This initial study was designed to collect reaction feedback without providing any verbal explanation of the feedback and to explore how much guidance might be necessary. I am considering developing a handout that would be distributed with the peer feedback report that will explain the meaning of each quantitative measure. I will use the results of continued research to guide what level of explanation is needed to assure the feedback is clear.

To identify if there was significant difference between each of the feedback instruments, I looked at the paired t-test statistics between measures (Table 4). From these data it is clear that, with the exception for the feedback on Pay Grade, there is no significant difference between the helpfulness, clarity, or bias of any of the quantitative feedback instruments. This result provides additional confirmation that further explanation of this measure may be needed for the recipient to fairly evaluate this measure.

Four semantic differential scales were constructed to gather data regarding each of the qualitative feedback questions:

- Useful 1 2 3 4 5 6 7
Useless
- Appropriate 1 2 3 4 5 6 7
Inappropriate
- Constructive 1 2 3 4 5 6 7
Destructive
- Fair 1 2 3 4 5 6 7
Unfair

The data collected shows the students viewed all the qualitative feedback as useful, appropriate, constructive, and fair (Table 5). Self reported feedback was the only measure that I cannot say with a high level of confidence was useful. This outcome may have occurred because the respondents did not have time to process these data and find them useful. Self reported feedback would most likely be used if the team was having a problem, and they chose to use these data as part of a discussion. Self reported feedback is unlikely to provide the individual information that would help a student to change their behavior, and thus, it would likely be judged less valuable than the other qualitative feedback.

This conclusion is also supported by looking at the between rating scale significance (Table 5). Here the self reported feedback is clearly less useful, appropriate, constructive, and fair than the other feedback. Again, some explanation of the purpose of this information may alter the response to this survey question.

**Quantitative Feedback Confidence Intervals – Difference from neutral
(paired t-test)
Table 3**

		Performance (judged by group)	Team Player	Quantity of Contribution	Quality of Contribution	Effect on Team Performance	Pay Grade (% of average pay)
Helpfulness of Different Peer Feedback	mean	3.00	2.86	2.88	2.76	2.82	3.82
	significance	0.000	0.000	0.000	0.000	0.000	0.473
Clearness of Different Peer Feedback	mean	3.11	2.68	2.64	2.59	2.65	3.55
	significance	0.000	0.000	0.000	0.000	0.000	0.039
Bias of Different Peer Feedback	mean	3.30	3.45	3.32	3.35	3.53	3.38
	significance	0.001	<i>0.011</i>	0.002	0.003	<i>0.037</i>	0.003
	Key:	99.90%	confidence indicated by BOLD				
		99.00%	confidence indicated by BOLD				
		95.00%	confidence indicated by <i>Italics</i>				
		<95.00%	confidence indicated by plain text (these values are not significant)				

Table 4
Quantitative Feedback Confidence Intervals – Between rating scales (paired t-test)

	Key:	99.90%	confidence indicated by BOLD				
		99.00%	confidence indicated by BOLD				
		95.00%	confidence indicated by <i>Italics</i>				
		<95.00%	confidence indicated by plain text (these values are not significant)				
Helpfulness of Different Peer Feedback	Mean Score	Performance (judged by group)	Team Player	Quantity of Contribution	Quality of Contribution	Effect on Team Performance	Pay Grade (% of average pay)
Performance (judged by group)	3.00	1	0.384	0.503	0.197	0.364	0.001
Team Player	2.86		1	0.924	0.518	0.770	0.000
Quantity of Contribution	2.88			1	0.280	0.677	0.000
Quality of Contribution	2.76				1	0.686	0.000
Effect on Team Performance	2.82					1	0.000
Pay Grade (% of average pay)	3.82						1
Clearness of Different Peer Feedback	Mean Score	Performance (judged by group)	Team Player	Quantity of Contribution	Quality of Contribution	Effect on Team Performance	Pay Grade (% of average pay)
Performance (judged by group)	3.11	1	0.019	0.020	0.010	0.049	0.086
Team Player	2.68		1	0.748	0.531	0.843	0.000
Quantity of Contribution	2.64			1	0.699	0.916	0.000
Quality of Contribution	2.59				1	0.636	0.000
Effect on Team Performance	2.65					1	0.000
Pay Grade (% of average pay)	3.55						1
Bias of Different Peer Feedback	Mean Score	Performance (judged by group)	Team Player	Quantity of Contribution	Quality of Contribution	Effect on Team Performance	Pay Grade (% of average pay)
Performance (judged by group)	3.30	1	0.385	0.503	0.778	0.257	0.715
Team Player	3.45		1	0.924	0.518	0.653	0.691
Quantity of Contribution	3.32			1	0.853	0.240	0.760
Quality of Contribution	3.35				1	0.292	0.875
Effect on Team Performance	3.53					1	0.463
Pay Grade (% of average pay)	3.38						1

Table 5
Qualitative Feedback Confidence Intervals – Difference from neutral (paired t-test)

		Greatest Contribution Feedback	Performance Improvement Feedback	Self Report Feedback		
Usefulness of Different Peer Qualitative Feedback	mean	2.95	2.80	3.59		
	significance	0.000	0.000	0.092		
Appropriateness of Different Peer Qualitative Feedback	mean	2.73	2.83	3.39		
	significance	0.000	0.000	0.007		
Constructiveness of Different Peer Qualitative Feedback	mean	2.95	2.86	3.42		
	significance	0.000	0.000	0.008		
Fairness of Different Peer Qualitative Feedback	mean	2.92	2.86	3.41		
	significance	0.000	0.000	0.007		
	Key:	99.90%	confidence indicated by BOLD confidence			
		99.00%	confidence indicated by BOLD			
		95.00%	confidence indicated by <i>Italics</i>			
		<95.00%	confidence indicated by plain text (these values are not significant)			

Qualitative Feedback Confidence Intervals – Between rating scales (paired t-test)
Table 6

	Key:	99.90%	confidence indicated by BOLD confidence		
		99.00%	confidence indicated by BOLD		
		<i>95.00%</i>	confidence indicated by <i>Italics</i>		
		<95.00%	confidence indicated by plain text (these values are not significant)		
			Greatest Contribution Feedback	Performance Improvement Feedback	Self Report Feedback
Usefulness of Different Peer Qualitative Feedback	Mean Score				
Greatest Contribution Feedback	2.95		1	0.395	0.006
Performance Improvement Feedback	2.80			1	0.000
Self Report Feedback	3.59				1
			Greatest Contribution Feedback	Performance Improvement Feedback	Self Report Feedback
Appropriateness of Different Peer Qualitative Feedback	Mean Score				
Greatest Contribution Feedback	2.73		1	0.481	0.000
Performance Improvement Feedback	2.83			1	0.002
Self Report Feedback	3.39				1
			Greatest Contribution Feedback	Performance Improvement Feedback	Self Report Feedback
Constructiveness of Different Peer Qualitative Feedback	Mean Score				
Greatest Contribution Feedback	2.95		1	0.544	<i>0.012</i>
Performance Improvement Feedback	2.86			1	0.005
Self Report Feedback	3.42				1
			Greatest Contribution Feedback	Performance Improvement Feedback	Self Report Feedback
Fairness of Different Peer Qualitative Feedback	Mean Score				
Greatest Contribution Feedback	2.92		1	0.718	0.004
Performance Improvement Feedback	2.86			1	0.003
Self Report Feedback	3.41				1

DISCUSSION

Our motivation was to develop an easy-to-use, valid, peer evaluation instrument that could be integrated into team-based classes. I also wanted to offer a tool for the faculty to use to encourage improved team performance over the duration of a team experience. Although this paper does not provide evidence that behavioral change occurred, data to support this claim is in the process of being collected. Because the data collection and processing is automated in Excel, the tool can be used multiple times in a single semester. During beta testing, four formative evaluations were given with a fifth evaluation used for summative purposes. The data presented in this paper was focused exclusively on the design of both the data collection instruments (input forms) and the feedback types given to team members.

The data presented shows the students were generally favorable to both the peer evaluation input instruments and the feedback generated. The less positive feedback on the forced and allocation ranking scales suggests the possibility

of further research. Although ranking scales provide some useful formative feedback they may be viewed as summative and generate an unintended leniency bias (raters giving a more generous rating to low performers because of fear that the data may be used for evaluation rather than developmental purposes). Recognizing that leniency bias may be an issue, the faculty involved saw the ranking information as a valuable early warning of future problems on the team. The forced ranking instrument was seen as the most effective early warning because it required at least one member of the team being identified as the lowest performer by a majority of the team. Being identified as the lowest performer over several evaluation periods provides a signal that there will likely be an issue with the final summative evaluation. If grades will be adjusted based on the final summative evaluation, this forced ranking instrument on the formative evaluations may signal to the lowest performer that they are at risk while they still have time to correct the behaviors that are placing them in this category.

Another observation that was made in the data collected but was not mentioned in the analysis was that students who

were negative on the first two summated feedback questions (e.g., I like to receive performance feedback from my teammates and Continuing to provide performance feedback to our team members will likely increase our team's performance) were generally more negative on all other questions regarding the feedback received (Table 7). This can be hypothesized as a Halo effect. Similarly, those that were more positive on the summated feedback questions were generally more positive when evaluating the types of feedback received. Unfortunately, there was not enough data to draw any strong conclusions from these data, and a larger sample will be collected. Additionally, future research will be conducted to look at changes in perceptions longitudinally and attempt to understand if these perceptions are the result of prior bad experiences and whether perceptions change with the introduction of better peer evaluation tools.

Finally, it was hypothesized that having access to feedback regarding individual contribution to team performance would greatly aid the instructor's effectiveness in counseling students. This hypothesis was confirmed by anecdotal evidence collected during the beta testing of the peer evaluation tool and prior experience of the authors. The future challenge is to link the peer evaluation tool with a facilitated process that is effective in changing individual team member behavior. The peer evaluation tool is only a part of the process, with the instructor playing a significant role in preparing the student prior to completing the peer evaluation and prior to receiving the feedback, and in constructively processing the feedback to modify future behavior.

The peer-to-peer assessment and feedback tool presented in this paper achieves the first objective identified: to develop a peer evaluation instrument that can be easily and effectively implemented in team-based learning environments. Although additional development may lead to significant refinements of the tool, its utility in the current form has been validated by the data collected. Now I can turn my attention to the second objective: to develop a process that can be used to teach students how to give, receive, and implement feedback from peer evaluations.

FUTURE RESEARCH

There are a number of related areas of research that can be facilitated by a peer evaluation tool that is easy to implement. In several sections of this paper I have highlighted the concern with ranking type measures, specifically, the forced ranking of peers. Issues surrounding peer ranking have been highlighted in the social justice literature (Roch, Sternburgh, & Caputo, 2007), in the general management literature (Gary, 2001), in the psychology literature (Blume, Baldwin, & Rubin, 2009; King & Hunter, 1980; Saal, Downey, & Lahey, 1980), by the Society for Human Resource Management (Olsen & Davis, 2006), and even in the popular business press (McGregor, 2006). While controversial, relatively little empirical evidence has been collected to enlighten the conversation. The peer assessment and feedback tool presented in this paper offers the opportunity to collect

**Halo effect
Table 7**

I like to receive performance feedback from my teammates			
	Count	Average rating on all other scales *	
Strongly Agree	23	2.38	
Agree	29	3.21	
Neutral	8	3.56	
Disagree	4	3.40	
Strongly Disagree	1	4.83	
Continuing to provide performance feedback to our team members will likely increase our team's performance			
	Count	Average rating on all other scales *	
Strongly Agree	20	2.36	
Agree	30	2.93	
Neutral	11	3.67	
Disagree	4	4.48	
Strongly Disagree	1	4.83	
* This is the average rating for both qualitative and quantitative feedback. It is based on a seven point rating scale where 1 is extremely positive and 7 is extremely negative.			

longitudinal data on these ranking type measures and measure both formative behavioral impacts and summative validity.

There also appears to be a gap in the current literature regarding how to assess the quality of feedback given on peer evaluation. Anecdotal evidence suggests that there is a broad disparity in the quality of feedback given by different raters. This disparity is evident in quantitative assessments where some raters will give all peers equal rating, and in qualitative assessments where some raters provide little or no constructive feedback. In an academic setting it is desirable to educate students to provide useful, constructive, and thoughtful feedback. To provide this education, the ideal peer evaluation tool would also be capable of generating information on the quality of the feedback given by each rater. This information could then be given back to the rater to encourage better feedback quality (more thoughtful responses) on future evaluations.

The peer evaluation tool was also designed to collect self evaluation information. There is an extensive literature on self evaluation versus peer evaluation. The problems of self-assessment have been recognized for decades (Atwater, Ostroff, Yammarino, & Fleenor, 1998). Some research reports students are accurate in their self-assessment (Brown, 1995), while other studies strongly indicate that self-ratings are generally less accurate, more biased, and hence, less reliable than are peer ratings (Atwater & Yammarino, 1992; Fleenor, McCauley, & Brutus, 1996; Harris & Schaubroeck, 1988; Mabe & West, 1982; VanVelsor, Taylor, & Leslie, 1993). These discrepancies lead to the need to continue researching the causes of deflated or inflated self-ratings. Preliminary work indicates that data collected from the current peer evaluation tool will be helpful in this ongoing research effort.

The data collected from peer evaluations could also be used to research demographic, psychographic, and geographic differences between raters and ratees. There is also an existing literature looking at the biases or perceived biases in peer evaluation. Given the ease in which data can be collected from a peer-to-peer evaluation tool, I can conceive of a number of studies isolating specific variables of interest. For example it would be relatively easy to combine commonly used psychological profiles (i.e., Myers Briggs personality test) with the peer evaluation tool and test hypotheses regarding both the self and peer ratings.

REFERENCES

- AACSB. (2005, January 1). Eligibility procedures and accreditation standards for business accreditation. Retrieved June 23, 2005, from <http://www.aacsb.edu/accreditation/business/AACSBSTANDARDS-Jan05-Final.pdf>
- Acona, D. G. (1990). Outward Bound: Strategies for team survival in an organization. *Academy of Management Journal*, 33(June), 334-365.
- Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: Does it really matter? *Personnel Psychology*, 51, 577-598.
- Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions. *Personnel Psychology*, 45, 141-164.
- Blume, B. D., Baldwin, T. T., & Rubin, R. S. (2009). Reactions to different types of forced distribution performance evaluation systems. *Journal of Business Psychology*, 24, 77-91.
- Brown, R. W. (1995). *Autorating: getting individual marks from team marks and enhancing teamwork*. Paper presented at the 1995 Frontiers in Education Conference, Atlanta, GA, USA.
- Cederblom, D., & Lounsbury, J. W. (1980). An investigation of user acceptance of peer evaluations. *Personnel Psychology*, 33(3), 567-579.
- Church, A. H., & Bracken, D. W. (1997). Advancing the state of the art of 360-degree feedback. *Group & Organizational Management*, 22(2), 149-161.
- DeNisi, A. S., Randolph, W. A., & Blencoe, A. G. (1983). Potential Problems with Peer Ratings. *The Academy of Management Journal*, 26(3), 457-464.
- Feder, D. B., Battenhausen, K. L., & Davis, W. (1999). Peer reviews: Employees dual roles as raters and recipients. *Group & Organizational Management*, 24, 92-120.
- Fleenor, J. W., McCauley, C., & Brutus, S. (1996). Self-other rating agreement and leader effectiveness. *Leadership Quarterly*, 7, 487-506.
- Gary, L. (2001). The controversial practice of forced ranking. *Management Update: A newsletter from Harvard Business School Publishing*, October.
- Gentry, J., Howard, B., Vaughan, M. J., Cudworth, A., & Vik, G. (2003). Improving the effectiveness of peer evaluations. *Developments in Business Simulation and Experiential Learning*, 30, 109-112.
- Gueldenzoph, L. E., & May, G. L. (2002). Collaborative peer evaluation: Best practices for group member assessments *Business Communication Quarterly*, 65(1), 9-20.
- Hansen, R. S. (2006). Benefits and problems with student teams: Suggestions for improving team projects. *Journal of Education for Business*, 82(1), 11-19.
- Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52(1), 123-142.
- Harris, M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41, 43-62.
- King, L. M., & Hunter, J. E. (1980). halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology*, 65(5), 507-516.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: a review and meta-analysis. *Journal of Applied Psychology*, 67, 280-290.

- McGourty, J., & DeMeuse, K. P. (2001). *The Team Developer: An assessment and skill building program*. Hoboken: John Wiley & Sons.
- McGregor, J. (2006). The struggle to measure performance. *Business Week*.
- Morahan-Martin, J. (1996). Should peers' evaluations be used in class projects? Questions regarding reliability, leniency, and acceptance. *Psychological Reports*, 78(3), 1243-1251.
- Nilsen, D., & Campbell, D. P. (1993). Self-observer rating discrepancies: Once an overrater, always an overrater? *Human Resource Management*, 32, 265-281.
- Olsen, C. A., & Davis, G. M. (2006). *Pros and cons of forced ranking and other relative performance ranking systems*: Society for Human Resource Management.
- Peiperl, M. A. (1999). Conditions for the success of peer evaluation. *International Journal of Human Resource Management*, 10(3), 429-458.
- Price, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, 93(3), 223-231.
- Roch, S. G., Sternburgh, A. M., & Caputo, P. M. (2007). Absolute vs relative performance rating formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment*, 15(3), 302-316.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88(2), 413-428.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.
- VanVelsor, E., Taylor, S., & Leslie, J. B. (1993). An examination of relationships among self-perception accuracy, self-awareness, gender and leader effectiveness. *Human Resource Management*, 32(2/3), 249-263.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.