AC 2008-897: EVALUATING DIFFERENT ASPECTS OF PEER INTERACTION USING AN ON-LINE INSTRUMENT

Alan Cheville, Oklahoma State University

James Duvall, Oklahoma State University

James Duvall is completing his BSEE degree at Oklahoma State University and expects to attend graduate school studying microwaves or photonics.

Evaluating Different Aspects of Peer Interaction Using an On-Line Instrument

Background and Context

As universities move towards integrating in-depth team-based design experiences there is an increasing need to train students in teamwork. Evaluating students' ability to function on multidisciplinary teams is a required element of ABET accreditation [1]. Design experiences in which students work on teams are common at many schools, particularly in capstone design courses. In fact capstone courses are the primary mechanism used by many universities for integrating teamwork into the engineering curriculum. In a 2005 survey 64% of capstone courses included teamwork [2]. The importance of teamwork in the development of engineering students is echoed by engineering employers. A recent survey asked employers of new engineering graduates to rate desirable characteristics; the three questions on teamwork were rated among the most important student attributes [3].

Another reason there is an increasing interest in getting engineering students to function on teams is the demonstrated impact of cooperative education techniques on student learning. Several meta-analyses [4, 5] of cooperative learning have shown a strong positive impact compared to more traditional, competitive approaches. Students, however, needed to be trained in teamwork in order to obtain maximal benefit from such cooperative learning approaches. As a result of the positive effect on students learning and engineering programs' interest in developing teamwork skills in their graduates, a wide variety of resources on team building are becoming available. Such resources include team development handbooks [6, 7], formative assessment for students working on teams [8], and summative assessment instruments [9]. Research has also been performed on effective methods of team formation [10] as well as the variability of team success [11].

Despite the number of resources that have been developed for teamwork in general and engineering teams in particular, it has been pointed out [12] that there is little agreement on the most effective way to measure team performance and individual accountability using peer evaluation. A large amount of work has been done on this problem, much is reviewed in [12] which shows the effectiveness of using "behavioral anchors" in peer evaluation. Several on-line peer evaluation systems are available [12, 13]. The problem of having students evaluate peers is exacerbated by the fact that it has recently been shown that low performers overestimate their abilities across multiple contexts [14, 15]. In evaluating peers in engineering courses another variable is where the teamwork experience falls on the spectrum of team projects. On one extreme of this spectrum are fully cooperative experiences in which the team works together towards a common goal. This extreme is defined by a single shared experience. On the other extreme are "divide-and-conquer" projects. Here a team assigns each individual separate tasks which each contribute to a shared team goal. Such approaches—an example is the jigsaw teaching technique—are defined by unique experiences for each individual. This approach is common to many capstone design courses.

This report looks at peer evaluation in a electrical engineering capstone design course required for all senior students at Oklahoma State University. In this course student teams use a block diagram approach for functional decomposition [16] of their projects. Multiple teams work on separate projects. The course consists of two projects, the first is designed to teach electronic fabrication techniques while the second is a more traditional design project. Students are required to create a block diagram of their project and each student takes responsibility for one or more functional components of their system. The team experience thus falls towards the divide-and-conquer end of the teamwork spectrum. A custom-designed electronic peer evaluation instrument was used during two projects to obtain peer evaluation of student performance. Five separate measures of team performance were made. This system is available for use by engineering faculty and further details, including the ability to use the system, are available [13].

Conceptual Framework

This work examines correlations between and within a peer evaluation instrument that measures three separate aspects of teamwork. Teamwork in educational settings has been examined from several perspectives [17]. In this study student responses on peer evaluation instruments are analyzed using constructs from motivational theory, particularly that of goal structures [18], in the context of a required capstone design course in an electrical engineering program at a comprehensive land grant university. As discussed in the introduction, teamwork is a vital part of most engineering capstone design courses. From the perspective of researching team function, capstone courses often impose unique conditions on students. Capstone design projects are often ill-defined which can result in potentially large time commitments. In most degree programs the time put into capstone courses takes away from the time devoted to other courses, extra-curricular activities and personal time. Student motivation will likey have a strong impact on how well students perform and how this performance is viewed by peers. An additional complicating factor is that capstone design courses commonly occur near the end of a baccalaureate program. Large demands on time are made at a time when students are preparing to graduate and extrinsic motivators are low; grades are not critical to students who have already found a job or been accepted to graduate school. For these reasons motivation is hypothesized to be critical to success and the peer evaluation is analyzed using constructs of motivation theory.

Following previous research on the role of goals in academic achievement this work assumes that students in the capstone design course set goals for themselves which they work to achieve. Achievement goals are generally divided into either mastery goals or performance goals [18]. Mastery goals are inwardly directed toward developing ones' abilities and have been positively correlated with self-development. Performance goals are outwardly directed towards portraying a positive image of oneself or inwardly directed at protecting an individual's self image. Mastery goals are directed towards objectively measurable performance while performance goals depend more on subjective evaluation. Achievement goal structures are not as simple as "mastery good, performance bad" in regards to academic achievement, however [18]. This paper further does not make an attempt to distinguish goals on the approach-avoid axis [18].

Motivation theory accepts that the goals students set for themselves may not be directly related to course objectives or learning outcomes. For example, other goal-motivation constructs reported in the literature include social or cognitive goals and it has been demonstrated [19] that such goals are often linked to academic goals. For example academic and social goals may be interconnected through a desire to please parents or instructors. Additionally goals may be classified as either intrinsic or extrinsic, i.e. set internally from one's own desires or by external factors such as class requirements. Intrinsic motivation is generally accepted as resulting in increased learning compared with extrinsic motivators.

Here it is hypothesized that *an individual's goals act as a lens to interpret and give valuation to the actions' of team members*. Thus how one evaluates peers depends on the goals an individual has for herself or himself. Given the large number of possible connected goals, this work focuses on two specific goal structures and examines connections between them and connections to other measures of performance in design classes. All goals are evaluated in the context of understanding peer performance in a team project with well-defined individual roles and outcomes. The first goal that this project attempts to measure are performance/social goals- how an individual and their actions are perceived by peers. Social goals are categorized here as intrinsic and performance oriented. The second goal is that of the effort or labor a student puts into the team design project. This goal is assumed to be related to mastery goals since effort is required for mastery. The third goal is designed to measure student effectiveness defined as the ratio of results and effort. Effectiveness is assumed to be related to self efficacy and is an intrinsic and mastery oriented goal. Determining effectiveness as the ratio of two separate measurements—effort and results—may additionally permit the determination of whether student perceive effort or results as contributing more to team design projects.

Research Questions

The overall research question addressed in this study is to understand what formats of peer evaluation instruments are more or less effective for measuring student performance on divide-and-conquer team projects. To answer this overall question four inter-related research questions were posed:

1) Are measures of team member *characteristics* (posited to relate to social and performance goals) or measures of team member *output* such as effort or results (posited to relate to self-efficacy and mastery goals) more predictive of performance in a design class. Similarly how closely related are team member characteristics to other measures of peer evaluation?
2) In measures of student output do students perceive the effort or the results of team members as being more important when evaluating their performance? How are measures of effort, results, and efficiency ($^{\text{results}}/_{\text{effort}}$) correlated with design performance?
3) How do students responses to peer evaluation instruments change over the duration of a capstone design course?
4) How are the scores given on a peer evaluation related to the scores received on a peer evaluation? Can responses given by a student predict how they are viewed by their peers?

These research questions have direct impact on practice in engineering education. Understanding whether students value performance goals or mastery goals in evaluating peer's work in capstone courses can lead to design of better peer evaluation instruments. If, as hypothesized, students evaluate peers based on their own goal structures, teaching a common goal structure to students could result in much more accurate peer evaluations. The same logic applies to the second research question. Understanding the relative importance students place on effort as opposed to results both permits improved evaluation instrument and gives insight on how to "calibrate" students by communicating common expectations. In a broad sense this research contributes to work on determining whether evaluations based on the theory of performance or mastery goals are more strongly correlated to other measures of team performance.

The third research question addresses whether student beliefs are changed by the design experience. Research on goals hypothesizes several models of the inter-relation between goals and actions, the simplest being the goals are reflect deeply help individual beliefs. In this case it would be possible for the evaluation instrument to measure static student beliefs more than peer's performance. However if responses change significantly over the duration of the capstone course then some change to goal structures is occurring. While not providing a definitive evaluation tool, this work can potentially lead to methods to gauge changes to an individual's ability to function of a team. Judging student efficacy on a team is also addressed by the fourth question. It has been recently shown that unskilled individuals are less aware of skill distinctions in others [14].

Methodology

Peer evaluations were given in a the first course of a two course capstone design sequence at Oklahoma State University. The first course teaches skills required for design and the design process through two design projects. The format of both courses is representative of capstone courses at other universities as determined by the overlap of structure, outcomes, and grading methods with other capstone course nationwide [2]. There are two design projects in the course and a classroom component that formally teaches the engineering design process [16]. The first project focuses on design skills in electronic simulation, design, fabrication, and testing techniques. The second design project involves design and fabrication of a prototype electronic system. The different outcomes and expectations of the two projects are a limitation of this study; it is not possible to determine the extent to which changes in peer evaluations result from experience or instruction. Teams were reformed after the first project; few students were on the same team for both projects.

Teams were formed by the instructor based on several criteria. The first criteria was to assign students to teams based on knowledge from elective courses taken as part of each student's "areas of specialization". The instructor determined the set of skills required to complete the second design project and assigned students to teams based on their declared area of specialization. For example if the design constraints for one project are to operate on batteries for an extended period at least one student from the power and energy area of specialization is assigned to the team. The second criteria was student schedules. Students filled out a survey at the start of the course listing blocks of time they could meet with peers and teams were assigned to maximize overlap of these times. No attempt was made to control for previous team experience. Most students have practiced teamwork in required courses before the capstone course, however students from a remote campus and transfer students may have significantly less teamwork experience.

An online, electronic peer-evaluation system was used to collect data on team performance and has been reported previously [20]. The peer evaluation uses behavioral anchors in questions due their reported efficacy in improving student responses [12]. Students were required to submit peer evaluations. Evaluations did not directly affect student grades in the Fall 2007 and later iterations of the capstone course, but did in previous semesters which may affect comparison of results between semesters [12]. The peer evaluation system is configured by the faculty member in charge of the class and a full peer evaluation consisted of five separate elements (modules) to obtain student feedback. The five elements of the peer evaluation include:

1) *Valuation*: a Likert (1-5) scale rating of team member attitude and value.
2) *Work*: a numeric reporting by peers on the perceived work put forth by team members on team tasks as a percentage of total work. Tasks included design, fabricating the project, debugging and error checking, contributing to written reports, and maintenance roles.
3) *Effectiveness*: two separate Likert (1-5) ratings are completed by students. One is designed to measure the perceived *Effort* and the other the *Results* of team members' effort. The ratio of these is used to define a team member's effectiveness. This module was newly developed in fall 2007 and was not included in previous semesters.
4) *Comment*: an open-ended text box which allows students to give anonymous feedback to team members.
5) *Overall*: an overall effectiveness rating given as a percentage of the expected contribution put forth by each team member.

Except on the numeric reporting of total work students did not rate themselves. The overall effectiveness rating system provided a limited number of points to ensure that students did not simply rate all their team members at the highest possible rating value [21]. The five peer evaluation elements used in each category are given as an appendix at the end of this paper.

During both projects student teams performed required peer evaluations at nominally three week intervals. Due to the length of the peer evaluation instrument full evaluations were only performed at the end of each project. Interim evaluations only requested the overall rating and comments. Once all students submit an evaluation and the evaluations are approved by the instructor, students are given anonymous feedback on team performance as their mean score on each of the sections of the peer evaluation as well as the mean and standard deviation of the scores for the entire class. Thus students receive feedback on their performance at regular intervals. It is not known how many students viewed this feedback or how seriously it was considered by students, but anecdotal evidence indicates at least some students do view and reflect on peer evaluation scores.

To determine student's "design ability" three other measures were compared to peer evaluation. The first measure is a "block diagram test" given to individual students in which they are asked to describe their design project through an engineering block diagram [22]. The block diagram test assesses how well students understand and can articulate the design process. The second measure is a reflective statement done at the end of the course which is graded using a rubric. The reflective statement is designed to elicit responses on changes in student's perceptions of engineering design through their own experiences. The third measure of design ability is a concept inventory given to all students was drawn from one being created nationally [23]. The questions on the concept inventory were chosen by at least two faculty from each sub-area of electrical engineering to represent conceptual knowledge that (a) all students should know without review, and (b) represent fundamental concepts that are taught to all students in the program. Thus the inventory measures how well student understand the concepts underlying electrical engineering.

To determine how ratings on different sections of the peer evaluation compared with each other and with scores on the other measures of engineering design correlation coefficients were calculated. The level of significance was set to $p < 0.05$. Due to the relatively small sample size the data was also analyzed for correlations with $p < 0.1$. While such correlations are reported, they are identified in the paper and not considered significant. Peer evaluation scores were also correlated with the three independent measures of design ability outlined above. At the time of publication student comments are still be numerically scored on multiple factors using a rubric and will be reported at the meeting.

To answer the fourth research question—comparing scores given to scores received—the mean and standard deviation of the scores given by each student was extracted from peer evaluation data and correlated with scores given and measures of design ability. The electronic peer evaluation system allows scores given to be easily separated from scores received by students. This analysis was undertaken to see if the relationship between skill and awareness reported by Kruger and Dunning [14] extended to peer evaluation.

This study suffers from several fairly significant limitations. The capstone design course in which measures were taken is still evolving and changes to the course goals, content, and grading mean the results between separate semesters are not directly comparable. The changes affect the academic aspects of the course- team make up and difficulty level of the design projects is consistent between semesters. The initial sample size for the semester for which most results are presented was N = 21 students. Students who did not complete the course were dropped from the sample. This is a relatively small sample size and thus it is difficult to draw significant conclusions from the data. Since the capstone design course is required for all students the sample is assumed to represent a valid cross section of engineering seniors. However the same course was offered at a branch campus that primarily serves non-traditional students that is located seventy miles from the main campus. While unlikely, self-selection of students cannot be ruled out. The measurements were done in a course at a comprehensive land grant university that emphasized a "divide-and-conquer" approach to an engineering design project and thus may not be applicable to more cooperative learning situations. The sample was predominantly traditional students and had minimal ethnic or gender diversity.

Findings

This section compares student responses given and received between different elements of the peer evaluation instrument and compares the peer evaluation to other measures of design ability. Since this study essentially was designed to be a coarse grained mapping of the "peer evaluation parameter space" a large number of comparisons are performed. For this reason the analysis of the data is described separately from the conclusions drawn from the data. The reader more interested in the answers to the research questions may wish to skip ahead several pages.

Correlations were made on data from the two complete peer evaluations performed at the end of each of the two projects. Results from the two projects were not compared since peer evaluations for each project are considered independent. In reporting scores the naming conventions from the previous section are used to identify different elements of the peer evaluation instrument. Mean scores given to students are reported in italics, for example *Overall*. Scores given by students are identified by the subscript G and scores received by the subscript R, for example $\mathit{Overall}_R$ corresponds to the mean score received on the overall evaluation section of the evaluation. The range of scores given or received was determined by the standard deviations and are reported as $s(\mathit{Overall}_G)$, i.e. the standard deviation of the overall scores given by the students in the sample.

There were few significant correlations between scores students gave or received for the different peer evaluation formats on the first design project. This likely means that students did not understand the format of the evaluation or lacked experience in working on teams. As far as the mean scores students received the only significant correlation was between the overall score received by students, $\mathit{Overall}_R$, with $\mathit{Valuation}_R$ at $r = 0.69$. Thus on the first evaluation students tended to rate team members on their personal attitude and perceived value to the team. Comparing received scores with scores given by students $\mathit{Overall}_R$ was correlated with $\mathit{Valuation}_G$ at $r = -0.52$. Thus on the first evaluation students who gave lower ratings *to* peers received higher ratings *from* peers. There were several instances of positive correlations between standard deviations of scores given on different parts of the evaluation and negative correlations between the mean and standard deviations of scores given. Considering both means and standard deviations for valuation, work, effectiveness and overall scores there were few significant correlations as shown in the left side of Table 1 below.

Comparing peer evaluation scores on the first project to other measures of design ability, the scores are not significantly correlated. If the level of significance is reduced to $p < 0.10$, student scores on the block diagram test are correlated with $\mathit{Valuation}_R$ at $r = 0.38$ and scores on the concept inventory are correlated with $\mathit{Overall}_R$ at $r = 0.37$. These correlations may become significant if larger samples were available.

Significant Correlations P#1

	Received	Given
Received	5%	7%
Given	7%	33%

Significant Correlations P#2

	Received	Given
Received	32%	24%
Given	24%	39%

Table 1: Comparison of the number of statistically significant correlations ($p < 0.05$) on different elements of the peer evaluation for the first design project (left) and the second design project (right).

In summary, at the end of the first project student seem to rate peers based primarily on personal attributes. Students who give lower average scores, hypothesized to be more discerning, receive higher scores. This lends some support to the conclusions of [14] when applied to peer evaluation in engineering capstone courses. Similarly the range of scores given by students is larger for those who give lower mean scores supporting the hypothesis of increased discernment among more competent students. Scores on the peer evaluation do not significantly correlate with other measures of design ability.

Three peer evaluations were performed in the second project, but only at the end of the project did students complete the full evaluation, discussed in this paper. The overall correlations between given and received scores are shown on the right half of Table 1, and show a much greater number of correlations between separate evaluation measures on the peer evaluation. The format of the final peer evaluation in the first and second projects were identical (see Appendix). Due to the large number of significant correlations the data was analyzed in terms of correlations between scores received on the peer evaluation, scores given on the peer evaluation, and the correlation between given and received scores.

From the mean scores received by students on the peer evaluation, $Valuation_R$ and $Work_R$ were both positively correlated with the overall score $Overall_R$ at $r = 0.56$ and 0.79 respectively. There was no significant correlation between $Valuation_R$ and $Work_R$ ($r = 0.16$, $p = 0.5$) indicating students likely saw personal attributes and effort as being disconnected. Although the individual measures $Effort_R$ and $Results_R$ used to determine $Effectiveness_R$ were strongly correlated with all other measures, $Effectiveness_R$ was not. $Effort_R$ and $Results_R$ were highly correlated at $r = 0.92$, $p < 0.001$. Interestingly $Effectiveness_R$ was negatively correlated with $Valuation_R$ at $r = -0.51$. As on the peer evaluation for the first project, the standard deviation of scores students received on different elements of the peer evaluation were positively correlated against other standard deviations, but negatively correlated with mean scores received.

Comparing scores given by students on different elements of the peer evaluation the same general trends hold. Mean scores given are positively correlated between $Valuation_G$, $Effort_G$, and $Results_G$ meaning that if students give high scores to a team member on one section of the peer evaluation they are likely to give high scores on other sections as well. Note that it is not possible to evaluate $Work_G$ due to the fact that peer evaluation enforces the sum of all work to sum to 100% to make scores consistent between team members. Standard deviation of scores given were generally negatively correlated with the mean scores given; i.e. students who give a wide range of scores generally gave lower average scores. Compared to scores received there are fewer significant correlations between $Valuation_G$ and more correlations between $s(Valuation_G)$ and $s(Work_G)$ with other sections of the peer evaluation. If one assumes that the personal attributes a student evaluates to give a valuation score are shared by all students in the sample then $Valuation_R$ and $Valuation_G$ would be expected to show similar correlations with other measures of peers. These results may indicate that there may be great variation in the way different individuals judge personal attributes that contribute to success in design projects.

The evidence to support the view that different individuals judged the personal attributes of peers differently is mixed. $Valuation_R$ and $Valuation_G$ are correlated at $r = -0.46$ which indicate some similarity, but this correlation is relatively weak compared to other correlations between elements of the peer evaluation instrument. Note that the negative factor indicates students who give high ratings on personal attributes tend to receive low ratings. In comparing other scores received to scores given there are clear differences. The score received on peers' judgment of work performed, $Work_R$, and overall value to the team, $Overall_R$, is negatively correlated with average scores given and positively correlated with the standard deviation of scores given on other elements. Beyond the correlation between valuation scores given and received mentioned earlier there are no correlations between $Valuation_R$ and scores given. However the valuation score given, $Valuation_G$, is negatively correlated with $Work_R$ $Effort_R$, $Results_R$, and $Overall_R$ and positively correlated with s($Work_R$). This was an unexpected result and a hypothesis that may explain these results is presented later in the paper.

Unlike peer evaluations on the first project there are several statistically significant correlations between the scores students *received* on the peer evaluation and other measures of "design ability". The ratings students received on $Effort_R$ and $Results_R$ were positively correlated with certain questions on the block diagram test [22] with coefficients of r = 0.45 to r = 0.55. Not surprisingly, the questions on the block diagram test that were correlated with peer evaluation scores were those that involved aspects of the design project that involved communication within the design team or focused on general or interconnection aspects of the design project. There were no significant correlations on test questions dealing with aspects of the project that focused on individual work. Neither the concept inventory or reflective statements were correlated with scores received on the peer evaluation in Fall, 2007 but positive correlations were observed in two previous iterations of the course.

There are also correlations between independent measures of design ability and the scores *given* by students on peer evaluations. The valuation score given, $Valuation_G$, is negatively correlated with questions on the block diagram test designed to gauge knowledge of the first steps in the design process. Similarly $Valuation_G$ is negatively correlated (r = -0.46) with scores on the concept inventory. Positive correlations are found between block diagram test scores and the scores determined for effectiveness of peers, $Effectiveness_G$, even though the scores used to calculate effectiveness—$Effort_G$ and $Results_G$—are not significantly correlated with test scores. While not conclusive, these results lend support to the hypothesis that student discernment of peers is related to ability to perform design. Students who give peers high ratings on personal attributes are less discerning while those who can distinguish results from effort tend to do better on tests which measure knowledge of a team's design project.

<u>Conclusions from Findings</u>

Question 1: "Are measures team member *characteristics* or measures of team member *output* such as effort or results more predictive of performance in a design class?"

Analysis of peer evaluation data provided a fairly clear answer to the first research question within the limitations of the study. Which element of the peer evaluation has more predictive power depended on the amount of experience students had in working in teams on capstone projects. In the first project students rated each other on personal attributes as measured by $Valuation_R$. However there was no statistically significant correlation between any element of the peer evaluation and other measures of design ability. In the second evaluation students appeared to judge each other more on the amount of work they did, $Work_R$ and $Effort_R$, as well as how peers judged the results of effort, $Results_R$. Since $Results_R$ and $Effort_R$ have nearly identical correlation coefficients it may be that students have difficulty distinguishing between the effort put in by team members and the results they obtained. It may also be that students ability in teamwork improved over the course of the semester and the scores reflect this improvement. For the related question of **"How closely related are team member characteristics to other measures of peer evaluation?"**, the answer is there was no relation between these on the first project and a moderate to strong relation on the second project.

Question 2: "In measures of student output do students perceive the effort or results of team members as being more important when evaluating their performance?"
The answer to the question is unclear from the data since both effort and results were strongly and nearly equally correlated for the second design project. It is hypothesized that many students were unable to distinguish between effort and results. Whether this is due to poor design of the peer evaluation instrument or the fact that students are novices at evaluating peers needs to be determined. However since the perceived work done by students, $Work_R$, was highly correlated with the overall rating students received, $Overall_R$, it is likely work and effort are more visible to peers than the results they produce. This is a sensible assumption for the type of "divide-and-conquer" design projects student teams completed. While a student's physical presence is observable, the results they generate are only visible when individual systems are integrated.
For the question **"How are measures of effort, results, and efficiency ($^{results}/_{effort}$) correlated with design performance?"** the answer is both effort and results have nearly identical correlations with the written test of understanding of the design process and project. This lends additional support to the assumption that students have difficulty distinguishing between effort and results. More interestingly how students rated peers on their effectiveness, $Effectiveness_G$, were correlated with the scores they received on the block diagram test. While not at all conclusive, it may be that students who better understand the design process and the overall team project are more discerning at identifying the effectiveness of team members.

Question 3: "How do students' responses to peer evaluation instruments change over the duration of a capstone design course?"
There was an increase in the importance of work, effort, and results on peer ratings for the second project in comparison to the first project where personal attributes were more valued. Since teams were rearranged after the first project it is assumed this shift in importance of different elements is not due to students being more familiar with team members. However there may be an effect due to increased team cohesion since the second project was two weeks longer. This study suggests that measures of how students value personal attributes of team members play a larger role early in a design project but the work and effort contributed by students becomes more important as students become more experienced. It is important to note that how students perceive attitudes and values of their peers still play a large role in the overall peer evaluation even late in a design project. These factors begin to be outweighed by other measures as students become more experienced. A likely scenario is that students who are novices at engineering design simply lack a frame of reference to evaluate their peers and thus use personal attributes. As students become more expert in the design process they are better able to judge other characteristics of peers that help the team succeed such as hard work, reliability, and technical knowledge.

Question 4: "How are the scores given on a peer evaluation related to those received on a peer evaluation? Can responses given by a student predict how they are viewed by their peers?"
Across multiple elements of the peer evaluation there are statistically significant correlations between the scores given by a student and those received by the student. The valuation score given, $Valuation_G$, is negatively correlated with the scores a student receives on all other elements of the peer evaluation except for their efficiency calculated from the effort and results metrics. *Students who rate peers lower on personal attributes are themselves rated higher by*

their peers. There are no other clear indications of correlations of one category of scores given to scores students receive on other elements of the peer evaluation. However the scores students receive on peer ratings of the work they have done, $Work_R$, are negatively correlated with the average scores they give and positively correlated with the range of scores given on all other elements of the peer evaluation except $Efficiency_G$. *Students who are reported by peers as doing a large amount of work on the design project give lower scores on average and the range of scores they give to peers is larger.*

These findings match those of Kruger and Dunning [14] who looked at why incompetent individuals consistently overrated their own abilities and even estimated their abilities higher than more competent peers. Across multiple domains of knowledge the above referenced study found that incompetent individuals were less adept at using available information to judge their own or others' performances. Top performers more easily calibrated their own evaluations when given feedback, but incompetent individuals were less able to recognize they weren't adept at judging others. Recall that for the capstone course reported here full peer evaluations were performed for two projects. There were very few significant correlations between measures of team function or design ability on the first project. Kruger and Dunning's work indicates that initial evaluations can suffer from false consensus effects resulting in competent students overestimating performance of peers. Given feedback, competent students self calibrate while less competent students do not. Thus the higher correlations on the second project may indicate that at least some students—likely the more competent ones—become more adept at evaluating peers. The negative coefficients between scores given and scores received may be due to the fact that less competent students (lower rated by peers) consistently overestimate others' ability.

Kruger and Dunning note that incompetent behavior is rarely self-correcting. One possible reason given for this observation is that incompetent students are less able to make social comparisons with competent individuals. From the theoretical framework on goals adopted here it is hypothesized that perhaps students who do not learn from experience do so because of a goal orientation that does not fit what is being evaluated. Students who focus mainly on social goals may substantially ignore elements of the design project that don't involve human interaction. This work does not provide any definitive answers to this question since students' goal structures were not measured in the capstone design class. It is worth noting that students who gave high valuation ratings to peers tended to perform more poorly on the written test over the team design project.

While it the conclusions of this study of senior engineering students have implications for how to perform meaningful peer evaluations it is important to reiterate the limitations of the study. Results have been presented for one semester of data for a sample of N = 21 engineering students. Very similar, but not identical results have been obtained over three semesters with similar sample sizes. Only one semester of data has been discussed since changes to the course make it questionable to directly compare results between semesters. The measurements were done in a course at a comprehensive land grant univeristy that emphasized a "divide-and-conqure" approach to an engineering design project and thus may not be applicable to more cooperative learning situations. The sample was predominantly traditional students, male, and had little ethnic diversity.

Impact on Practice

The results of this study have several potential impacts on practice. Given the importance of team projects in capstone design courses and the need to measure teamwork as an element of a comprehensive ABET CQI process it is important to improve the reliability of peer evaluation. Developing reliable proxy measures of individual student's performance on team design projects can help instructors better assess individual student performance in team-based design courses.

The first recommendation to come out of this study is that measures of how peers value behaviors and attitudes is likely a better indicator of team function for novices than it is for students who have had prior experience in a design course. As students gained more experience they tended to rate peers more on effort and results. Instructors may wish to weight evaluations of effort more heavily than measures focused on personal attributes in final peer evaluations. Note that attribute measures may play an important role early in the design process to help calibrate behavior and thus should not be neglected.

The element of the peer evaluation tool used to measure results is highly correlated with that of effort. More work is definitely needed to develop metrics which can accurately let peers evaluate the results generated by individual students in a team project. There seems to be a consistent over-emphasis by students on the amount of work done rather than the quality and timeliness of results. If this observation is borne out by further studies, peer evaluations my unknowingly positively reinforce novice behaviors and penalize expert behaviors. A main difference between novices and experts is the time required to accomplish a given task. This effect could be mitigated if engineering educators developed formal methods for teaching students to recognize the difference between effort and results.

Besides rating students based on the evaluations they *receive* from peers, the scores students *give* also provide useful information for peer evaluation. Although the conclusions are tenuous due to the small sample size, the mean of scores *received* by students tends to be highly correlated while the range (standard deviation) of scores *given* are more highly correlated with other peer measures than the mean score given. Thus examining the range of scores given to peers may help instructors determine the most competent individuals on a team. The range of scores given is hypothesized to be related to how well students are able to discern the work done by peers.

This work also shows that students' ability to evaluate peers likely improves over the course of design courses when students are given feedback on team performance. One recommendation is that regular peer evaluations be conducted and rapid feedback given to students. Mapping changes in *how* students evaluate peers can be used to determine how students ability to work on multidisciplinary teams changes as a result of capstone design courses. Thus peer evaluations that measure different aspects of team experiences could be used in program evaluation when given regularly.

Finally in line with [14] it may be worth giving stronger consideration to the evaluations more competent students when using peer evaluations to determine grades on team design projects. While this is a potentially inflammatory suggestion with some negative consequences, research indicates that more competent individuals are both more adept at rating peers and are better able to learn from experience. Hence the most competent individuals will become better evaluators while less competent individuals learn more slowly.

The authors acknowledge support from the National Science Foundation through award NSF0530588. James Duvall would like to acknowledge support from a National Science Foundation REU award (NSF0631565) which enabled him to develop the peer evaluation software used in this study. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Bibliography

1. "Criteria for Accrediting Engineering Programs," (Accreditation Board for Engineering and Technology, Baltimore, MD, 2004).
2. S. Howe, and J. Wilbarger, "2005 NATIONAL SURVEY OF ENGINEERING CAPSTONE DESIGN COURSES," in *ASEE Annual Meeting*, (Chicago, 2006).
3. D. C. Davis, S. W. Beyerlein, and I. T. Davis, "Development and Use of an Engineer Profile," in *ASEE Annual Conference and Exposition*, (Portland, 2005).
4. M. Prince, "Does Active Learning Work? A review of the research," J. Eng. Educ. **93,** 223-246 (2004).
5. K. A. Smith, S. D. Sheppard, D. W. Johnson, and R. T. Johnson, "Pedagogies of Engagement: Classroom-Based Practices," J. Eng. Educ. **94,** 87-101 (2005).
6. D. Carr, Herman, K. Zarotney, Miller, and A. Wakefield, *The Team Learning Assistant Workbook* (McGraw Hill, Boston, 2005).
7. K. A. Smith, *Teamwork and Project Management* (McGraw Hill, Boston, 2007).
8. J. McGoutry, and K. DeMeuse, *The Team Developer: An Assessment and Skill Building Program* (J. Wiley & Co., New York, 2000).
9. D. B. Kaufman, R. M. Felder, and H. Fuller, "Accounting for Individual Effort in Cooperatire Learning Teams," Journal of Engineering Education **89,** 133-140 (2000).
10. S. G. Adams, "Building Successful Student Teams in the Engineering Classroom," Journal of STEM Education **4,** 27-32 (2003).
11. K. L. Tonso, "Teams that Work: Campus Culture, Engineer Identity, and Social Interactions," J. Eng. Educ. **95,** 25-37 (2006).
12. M. W. Ohland, R. A. Layton, M. L. Loughry, and A. G. Yuhasz, "Effects of Behavior Anchors on Peer Evaluation Reliability," Journal of Engineering Education **94,** 319-326 (2005).
13. J. Duvall, and A. Cheville, "Web-Based Configurable Peer Evaluation System," (2007).
14. J. Kruger, and D. Dunning, "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," Journal of Personality and Social Psychology **77,** 1121-1134 (1999).
15. M. Lejk, and M. Wyvill, "The Effect of the Inclusion of Self-assessment with Peer Assessment of Contributions to a Group Project: A Quantitative Study of Secret and Agreed Assessments," Assessment & Evaluation in Higher Education **26,** 551-561 (2001).
16. R. M. Ford, and C. S. Coulston, *Design for Electrical and Computer Engineers* (McGraw-Hill, New York, 2007).
17. N. M. Webb, and A. S. Palincsar, "Group Processes in the Classroom," in *Handbook of Educational Psychology*, D. C. Berliner, and R. C. Calfee, eds. (Simon and Schuster MacMillan, New York, 1996).
18. P. R. Pintrich, "A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts," Journal of Educational Psychology **95,** 667-686 (2003).
19. K. Wentzel, "What is it that I'm trying to achieve? Classroom goals from a content perspective.," Contemporary Educational Psychology **25,** 105-115 (2000).
20. A. Cheville, C. Co, and B. Turner, "Improving team performance in a capstone design course using the jigsaw technique and electronic peer evaluation," in *ASEE Ann. Conf. and Exhib.*, (Honolulu, 2007).
21. L. K. Michaelsen, "Team Learning: A Comprehensive Approach to Harnessing the Power of Small Groups in Higher Education," To Improve the Academy **11,** 5-54 (1992).

22. A. Cheville, C. Co, and B. Turner, "Communication as a proxy measure for student "design ability" in capstone courses," in *ASEE Ann. Conf and Exhib.*, (Honolulu, 2007).
23. K. E. Wage, J. R. Buck, C. H. G. Wright, and W. T. B., "The Signals and System Concept Inventory," IEEE Trans. Educ **48,** 448-461 (2005).

Appendix: Peer Evaluation Instrument Elements

Please rate each of your team members on their personal attributes such contributed to the performance of your team. Their rating should be based on **all** the work done as a team, both in and out of class. You may use non-integer values (i.e. 3.5).

Your scores should have an average in the range of 3.0. Ratings of 1 and 5 should be reserved for very special cases. Ratings that are very high or very low may require that you later justify them to the instructor or TA.
Rate each team member based on the following scale:
1 = unacceptable performance, I would fire this person
2 = could use improvement, bottom 20% of people I have worked with
3 = meets expectations, middle 50% of people I have worked with
4 = exceeds expectations, top 20% of people I have worked with
5 = outstanding, a rare individual

	Team Member 1	**Team Member 2**	**Team Member 3**
Honesty: What this person honest? Did they stand up for their beliefs? Did they take personal responsibility when they made a mistake.			
Self Awareness: Was this person aware of the limits of their knowledge and competence. Could they accurately self evaluate their ability at any task?			
Maturity: Was this person mature in dealing with others on the team or were they immaature? Did they contribute to creating a professional environment?			
Attitude: Did this person have a positive attitude that helped the team function? Were they pleasant to work with?			
Responsible: Was this person responsible in getting their work done correctly and on time. Could you rely on this person?			

Figure 1: Valuation element of peer evaluation, a Likert (1-5) scale rating of team member attitude and values.

Please provide the percentage of work each person contributed to each of the tasks listed below. All columns must add to 100%! If all team members contributed equally, put the same percentage in each column.

A well functioning team will have different team members be responsible for different tasks and the values below should reflect this distribution of effort. In this section of the evaluation you are asked to assign the relative percentage of work done by all member on the team, including yourself. Each column represents part of the team grade for the class and each column should add up to 100%. If any column does not add up to 100% when you click the submit button, your scores will not be accepted, and you will have to return and correct the values. Use integer numbers only or you may have to start from the beginning!

The numbers you enter here will be used to calculate a suggested rating for each of your team members. You do not need to use the number suggested, but should your rating differ greatly from the values you enter, you may be asked to justify them to the instructor or TA.

	Individually learning and practicing the skills they needed to support the team project.	Helping research, define, and understand the project.	Building or designing your portion of the project. Doing the work required.	Writing the reports and documentation.	Other tasks such as getting parts, cleaning up the bench, meeting with the TA's, etc.
TeamMember1					
TeamMember2					
TeamMember3					
TeamMember4					

Figure 2: Work element of peer evaluation, a numeric reporting by peers on the perceived work put forth by team members on team tasks as a percentage of total work.

Overall Rating
Please provide an overall rating for each team member.

- A score of 100% represents an average team member who did all the work expected of them.
- Scores below 100% reflect team members who contributed less than expected to the project.
- Scores over 100% are for team members who contributed more than their share.

You have 100 points per team-mate to assign. You do not have to use all of your points if you feel team members did not deserve a score of 100 or greater. **You have: 300 points to assign to your teammates.**

Team Member	**Overall Rating**
Team Member 1	
Team Member 2	
Team Member 3	

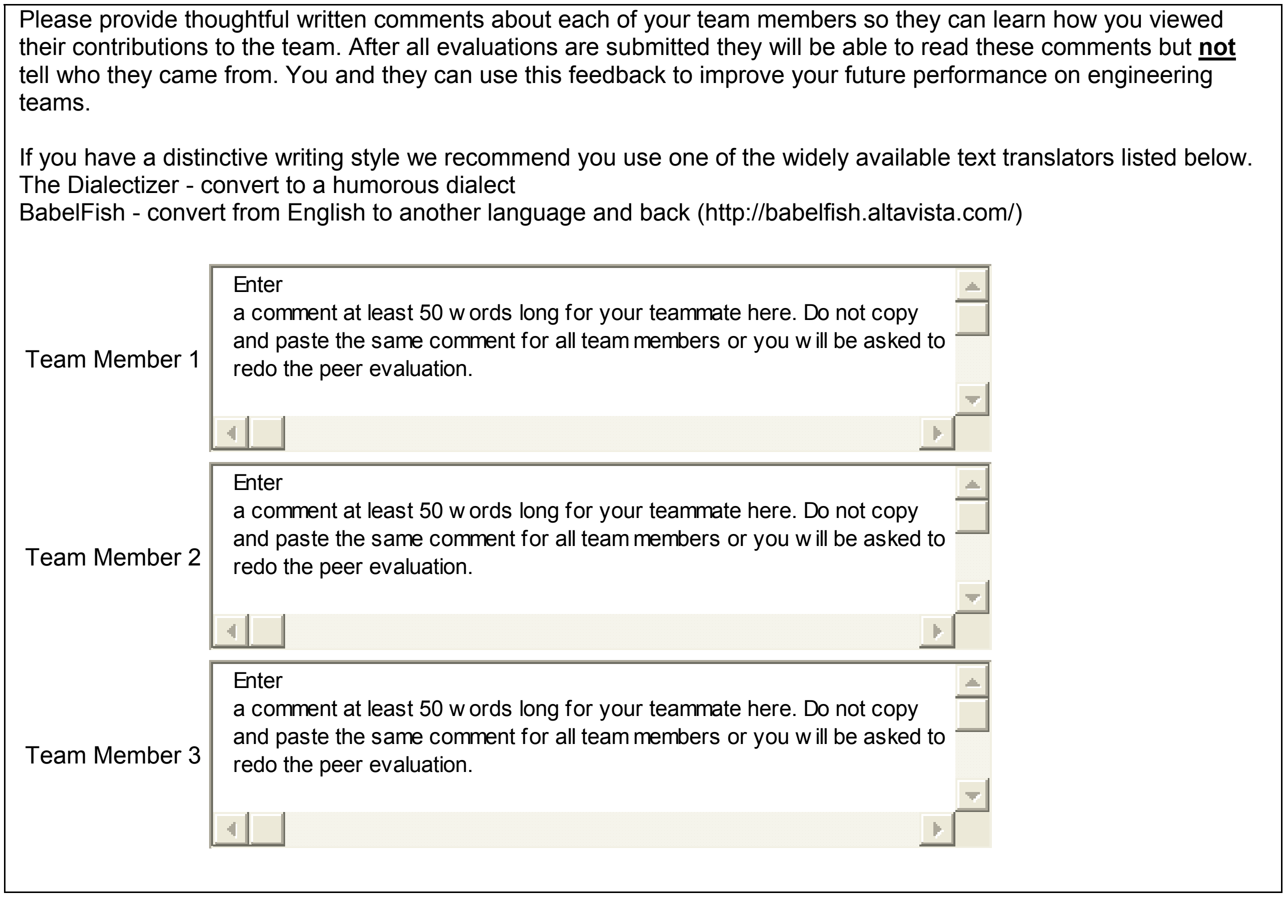
Figure 3: The overall effectiveness rating given as a percentage of the expected contribution put forth by each team member.

Please rate each of your team members on how much they contributed to your team. This rating should be based on **all** the work done as a team, both in and out of class. You may use non-integer values (i.e. 3.5). Your scores should have an average in the range of 3.0. Ratings of 1 and 5 should be reserved for very special cases. Ratings that are very high or very low may require that you later justify them to the instructor or TA.

Rate each team member based on the following scale:
1 = Among the very lowest in the team or the class.
2 = Not terrible, but clearly below most others in the class or on the team. In the lowest 20% of those I've worked with.
3 = In the middle, in the middle 50% of people I have worked with
4 = Consistently better than most class or team mates. In the top 20% of people I have worked with
5 = One of the best in the class or on the team. A rare individual

	TeamMember1	TeamMember2	TeamMember3
Hours Worked: How often was this person working in the lab? Did you always seen them when you came in?			
Additional Work: How willing was this person to take on extra work? Did your team always turn to them when something needed to get done?			
Doing Others' Work: Was he or she the fall-back person? Did they always seem to be given or help with tasks others couldn't or wouldn't do?			
Perfectionism: How often did this person insist on quality? Did they need to do every job perfectly?			
Completing Tasks: How good was this person at getting their tasks **done**. Did they always finish what they started?			
Quality Work: Was the work this person did right? Did their tasks work and get done right the first time?			
Promptness: Did this person get done when they said they would? Could you count on them to meet the deadlines set by the team?			
Value of Advice: Did the suggestions and advice this person give during the project turn out to be right more often than not? Did they know what they were talking about?			

Figure 4: Effectiveness element of peer evaluation consisting of two separate Likert (1-5) ratings. The upper four questions are designed to measure the perceived Effort and the bottom four the Results of team members' effort.



Please provide thoughtful written comments about each of your team members so they can learn how you viewed their contributions to the team. After all evaluations are submitted they will be able to read these comments but **<u>not</u>** tell who they came from. You and they can use this feedback to improve your future performance on engineering teams.

If you have a distinctive writing style we recommend you use one of the widely available text translators listed below.
The Dialectizer - convert to a humorous dialect
BabelFish - convert from English to another language and back (http://babelfish.altavista.com/)

Team Member 1

Enter
a comment at least 50 words long for your teammate here. Do not copy and paste the same comment for all team members or you will be asked to redo the peer evaluation.

Team Member 2

Enter
a comment at least 50 words long for your teammate here. Do not copy and paste the same comment for all team members or you will be asked to redo the peer evaluation.

Team Member 3

Enter
a comment at least 50 words long for your teammate here. Do not copy and paste the same comment for all team members or you will be asked to redo the peer evaluation.

Figure 5: The open-ended comment element of the peer evaluation which allows students to give anonymous, unstructured feedback to team members.