Student Perceptions on the Impact of Formative Peer Team Member Effectiveness Evaluation in an Introductory Design Course

Prof. Nathan Mentzer, Purdue University, West Lafayette

Nathan Mentzer is an assistant professor in the College of Technology with a joint appointment in the College of Education at Purdue University. Hired as a part of the strategic P12 STEM initiative, he prepares Engineering/Technology candidates for teacher licensure. Dr. Mentzer's educational efforts in pedagogical content knowledge are guided by a research theme centered in student learning of engineering design thinking on the secondary level. Nathan was a former middle and high school technology educator in Montana prior to pursuing a doctoral degree. He was a National Center for Engineering and Technology Education (NCETE) Fellow at Utah State University while pursuing a Ph.D. in Curriculum and Instruction. After graduation he completed a one year appointment with the Center as a postdoctoral researcher.

Mr. Andrew Jackson, Purdue University, West Lafayette

Andrew Jackson is currently pursuing a Master of Science in Technology Leadership and Innovation in Purdue University's College of Technology. His previous middle school teaching experience informs his role as a graduate teaching assistant for an introductory course in design thinking. His research interests are engineering self-efficacy, creativity, and decision making.

Dr. Kevin Andrew Richards, Northern Illinois University

K. Andrew R. Richards is currently a visiting assistant professor at Northern Illinois University. Prior to his current post, Richards was a post-doctoral research associate with the Center for Instructional Excellence at Purdue University, USA. His post-doctoral position focused on the evaluation of a large-scale course transformation project that sought to increase active learning and student-centered pedagogies in university-level teaching. Prior to post-doctoral studies, Richards completed his Master's degree and PhD at Purdue University, and Bachelor's degree at Springfield College (MA). His research centers on teacher socialization in physical education with a particular focus on role stressors that arise from combining teaching and extracurricular roles, such as athletic coaching. He has co-authored empirical articles using occupational socialization theory, and has also written research reviews on the topics of socialization theory and role theory published in Quest, the Kinesiology Review, and Research Quarterly for Exercise and Sport. Richards has served as a peer reviewer for numerous professional conferences and journals. Highlights include reviewing for the Journal of Physical Education, Recreation, and Dance, the Journal of Teaching in Physical Education, the European Physical Education Review, and Measurement in Physical Education and Exercise Science. He is currently serving as the editor of the Advocacy in Action section of the journal Strategies.

Dr. Angelika N Zissimopoulos, Purdue University, West Lafayette

Dr. Zissimopoulos holds a PhD in Biomedical Engineering From Northwestern University. She is currently an Instructional Developer at the Center for Instructional Excellence at Purdue University and works extensively with course design and faculty development. She also holds a courtesy appointment with the College of Technology.

Dr. Dawn Laux, Purdue University

Dawn Laux is a Clinical Assistant Professor in the Department of Computer and Information Technology (CIT) at Purdue University. She has been with the University since 2007 and is responsible for teaching database fundamentals courses and introductory technology courses. Dawn has 10 years of industrial experience in the information technology field, and her research area of interest includes technology readiness, the social impacts of technology, and increasing interest in the field of computing.

©American Society for Engineering Education, 2015

Student Perceptions on the Impact of Formative Peer Team Member Effectiveness Evaluation in an Introductory Design Course

Teamwork and Collaboration are among the three primary competencies needed for graduates to be successful in the workplace according to the Committee on Defining Deeper Learning and 21st Century Skills organized by the National Research Council.[1] The committee reviewed eight thematic reports and subsequently presented a framework intended to inform curriculum programs of these desired skills. Organizations and criteria governing the accreditation of various higher education disciplines also address the need for teamwork skills. Several student outcomes in the ABET Engineering Accreditation Commission[2] specifically relate to collaborative work:

> 3(d): an ability to function on multidisciplinary teams
> 3(g): an ability to communicate effectively (p. 3)

These skills should be acquired through the duration of a student's undergraduate experience and programs are required to "identify, collect, and prepare data to evaluate the attainment" of these outcomes (p. 2). Literature building on these criteria has reiterated that engineers will work in interdisciplinary teams and be required to communicate effectively as the world becomes more globalized.[3] Generally, "scholars are recognizing that teamwork is among the most essential learning outcomes for college students in all disciplines, both because it facilitates other learning and because employers in most fields value teamwork."[4]

Empirical evidence from employers supports claims that teamwork is a valued skill.[1] A 2013 report on job outlook indicated that the "ability to work in a team structure" was the most important candidate quality sought by recruiters. Candidate "ability to verbally communicate" was also ranked as being highly important.[5] According to an American Management Association (AMA) survey, 93% of managers and executives felt that collaboration and team building were important for growing their organization and 74.6% believed that collaboration, communication, critical thinking and creativity skills will become more important in the future.[6] Despite these remarks however, nearly one fifth of the recent graduates hired were assessed at below average collaboration skill and the number of managers admitting their employees were below average in collaboration skills (12.4%) and communication skills (13.2%) represents an increase from previous studies. Feedback from recruiters and employers demonstrates the importance of embedding "soft skills," like teamwork and communication, into the educational experiences of students in order to prepare them for future demands.[6]

Teamwork Skills Can Be Developed Through Collaborative Learning Experiences

Although there are varying names and strategies, generally collaborative learning represents an instructional method where students work together to accomplish structured tasks. Kaufman, Felder, and Fuller[7] described five conditions that need to be met in the design of collaborative learning experiences: "positive interdependence, individual accountability, face-to-face interaction, appropriate use of collaborative skills, and regular self-assessment of team functioning" (p. 133). Through these conditions and the underlying mechanisms of group interactions, cooperative learning helps build teamwork and communication skills.[8] The

increased frequency of collaborative instructional strategy can be seen as a response to the present demands for teamwork skills in new employees.[9]

Prior work has identified several elements of collaborative learning that lead to the development of skills needed for success in the workplace. By its nature, collaborative work requires and builds interpersonal and communication skills;[7, 10] provides exposure to different views, ideas, and perspectives;[10] leads to opportunities for negotiation;[11] and supports questioning among team members;[7, 12] among other benefits. Through social interactions with other learners, students have an opportunity to learn through reflection on their own experience and benefit from hearing the experiences of others.[13] Learner-learner interactions present an opportunity to learn both content and these "group behavior or group leadership skills" (p. 462)[14]. According to Verzat, Byrne, and Fayolle[15] "in the case of teamwork, doing it rather than listening about how important it is, is likely to have a more direct impact on student understanding" (p. 359). Burdett[9] surveyed 344 senior business students about group learning experiences. Students reported that through their group experiences they gained competency with teamwork. Comments from the same students showed that sharing views, building relationships, and improving learning about group processes were highlights of group experiences. Verzat et al.[15] conducted a group based design challenge with first year engineering students, followed by a debrief about the team experience. Following participation in the collaborative activities, student reflection highlighted learning of teamwork skills including: listening to others, trusting others, avoiding conflict, and interpersonal communication as main lessons learned.

Collaborative Learning Has Some Challenges

While experience working with teams is beneficial, unfortunately simply asking students to work together effectively does not work.[8] "Based on both antidotal [sic] experiences with teams and empirical or theoretical support, it is known that teams are not easily implemented, that the creation of a team of skilled members does not ensure success, and that teamwork does not just happen" (p. 556)[16]. Some of the challenges from teamwork may come from the initial learning curve that occurs as students deal with the complexities of group dynamics.[9, 17] This learning curve includes figuring out how to communicate, establishing meeting times that accommodate varying schedules, and making decisions. Students with strong tendencies toward individual work may show a decrease in enjoyment, enthusiasm, and perceived learning when placed in a group.[18]

Aside from overcoming the challenges of group dynamics, Fellenz[10] summarized "accurately and fairly assessing individual performance, conflict within groups, and free riding of individual members" (p. 571) as several main concerns of group work. These are regularly raised as impediments for including group activities in the classroom. Instructors may be unaware of how to deal with these issues and they further complicate the group learning environment.[19] These concerns are described and the case is made for a solution which addresses these issues.

Fair Assessment of Individual Performance

Fair assessment represents a great challenge to group work. Within competitive grading environments the learning goals of cooperative exercises are undermined because students may "resent others benefiting from the results of their hard work" (p. 185)[9]. For this reason, Felder

and Brent[8] recommend against using group activities in norm-referenced classrooms where students are competing with each other for course ranking. Even with standardized grading procedures (i.e., criterion-referenced), student frustrations about the imbalance of group member efforts are justified when there is no individualized assessment. When grading methods do not include individual effort as a complement to group effort it may decrease student willingness to participate.[20]

In order to structure individual responsibilities, groups may divide up labor on the project, each completing a part of it to provide balance in work effort. An unfortunate outcome of this process is the lack of "appreciation or understanding of what their colleagues have done. The individual pieces get cobbled together, and the combined output is submitted" (p. 109)[20]. Without a structured method for reporting back and teaching on individual efforts within the team, complete learning goals are only obtained when considering the team collectively rather than individually.

Conflict Within Groups

The collection of diverse individuals on a team may lead to conflict or friction among team members which, in turn, results in production blocking. Deeter-Schmelz, Kennedy, and Ramsey[21] reported that even with highly competent team members, a team that lacks cohesion will suffer in performance and goal achievement. Students who experience problems may be uncomfortable speaking out or may wish to endure the issue to avoid confrontation.[19] Conflict represents a frequent cause for group failure.[22] Several authors recommend a technique called *active listening* where team members take turns sharing their perspective and restating the other's perspective to reach understanding.[8, 19] While active listening may be beneficial for resolving issues, nascent learners may need further structure to facilitate the needed discussion.

Free Riding of Individual Members

Free riding, also called social loafing, is frequently identified as a problem of group work. The situation occurs when a team member is unwilling or unable to contribute effectively to the team, takes advantage of team members, and forces others to do more than their share of the work.[20] Team members may have different expectations for success on the project leaving those with higher goals to carry the weight of the project and make up for the poor contribution of those with lower goals. Free riding is more likely to occur as team size increases[21] and when individual contributions are subsumed by the overall group grade[10].

Peer Feedback is Beneficial in Addressing Challenges with Collaborative Learning

When asked to improve teamwork experiences, students identified improving assessment procedures and improving communications within the group as top priorities.[9] Salas et al.[16] identified *mutual performance monitoring* as an important element leading to effective teams. The authors referred to the previous definition that well performing teams are good at "keep[ing] track of fellow team members' work while carrying out their own…to ensure that everything is running as expected and...to ensure that they are following procedures correctly" (p. 23)[23]. This suggests that within-group peer evaluations can contribute significantly to teamwork; peer feedback provides an established time for reflection on the process and contribution of team

members and can help ensure things are running smoothly. Exposure to measurement instruments also alert students to the indicators of effective teamwork and may help inform their subsequent goals and behavior.[24] When peer feedback is structured with opportunities to provide and receive feedback it becomes a form of double-loop communication that effectively involves both parties in the learning process.[13]

How should peer feedback be administered within the classroom?

Recommendations drawn from literature include: 1) assessing team citizenship, 2) provide debriefs following peer evaluations to close the feedback loop, 3) administer evaluations iteratively, and 4) thoroughly integrate peer assessment in the curriculum. First, peer evaluations should be based on team citizenship, teamwork skills and responsibility of performance rather than academic ability.[7, 19] If all team members contribute and work cooperatively they should receive the team grade; in the case of free riders they will be penalized by their insufficient contributions. Second, peer feedback should conclude with a session of 'straight talk' regarding student ratings, praise, and feedback. Holding team conversations with the previously conducted peer evaluation as a starting point can also facilitate conflict resolution including addressing issues of free riding. Third, peer evaluation should be a type of formative assessment. In promoting learning, scholars have made the case that most effective learning occurs when students have an opportunity to practice, receive feedback, and then try again.[25-27] This knowledge reinforces the decision to administer peer evaluation in a formative way, allowing students to repeatedly receive feedback and try again within their groups. With each new attempt in using teamwork knowledge and skills, students receive peer feedback that can identify areas for improvement and motivate future effort. This framework also presents a clear opportunity for renewal if team members struggled due to conflict or free riding. Fourth, assessment should be integrated into the curriculum and be perceived as a learning opportunity. According to Brew[28]:

> Assessment and learning must increasingly be viewed as one and the same activity; assessment must become an integral part of the learning process. . . . When teachers share with their students the process of assessment - giving up control, sharing power and leading students to take on the authority to assess themselves - the professional judgment of both is enhanced. Assessment becomes not something done to students. It becomes an activity done with students. (p. 169)

Many researchers agree that peer evaluation may be used as a multiplier to transform group grades into individual grades.[7, 8, 29, 30] Although the calculation methods vary, the principle is that the whole of ratings on an individual's performance are ultimately used as an indicator of their contribution.

Quantitative analysis of peer evaluations has demonstrated its usefulness. Chapman and van Auken[31] found that alleviating work and grade equity concerns while assessing group projects improved student attitudes toward group work. Students who had more positive experiences with group work also reported that peer evaluations had been used much more frequently than those who had negative group experiences (64.2% versus 49.6%). Consequently the authors recommend using various assessment techniques including peer evaluations.

Research conducted in two consecutive cooperative learning courses in a university supported the ability of peer feedback to address concerns in group learning activities.[7] Peer feedback ratings were administered and used to weight those final homework grades which had been

completed as groups. In the face of many concerns, the results showed that peer feedback provided an equitable evaluation measure for each team's cooperation. "Contrary to expectations, inflated self-ratings proved to be less common than deflated self-ratings" (p. 136). There were insignificant differences between male and female evaluations, as well as non-minority and minority peer evaluations suggesting that any perceived biases didn't influence the peer evaluations. The researchers also believe that instances of social-loafing were reduced because of the "knowledge that there will be a penalty for those who try it" (p. 139) in the form of lower ratings.[7]

Dominick, Reilly, and McGourty[32] found that using peer feedback after group decision tasks led to positive increases in a variety of team skills including collaboration, communication, and decision making compared to teams that didn't use feedback. Others have found similar increases in team effectiveness and ratings in repeated measures.[29] Brutus et al.[30] found that repeated attempts of peer feedback helped students become more confident and specific in their evaluations of performance, which is an important professional skill. As they put it, this is "actually enabling students, from one class to the next, to obtain feedback of increasing quality to guide their development. This is likely to translate into better functioning teams in undergraduate classes" (p. 28).

Contributions of This Work

Although other work has recommended and evaluated the use of peer feedback in courses, these studies generally have focused on the benefit of peer evaluation rather than changes across repeated administrators and feedback cycles.[7, 9, 22, 33] Some studies have used repeated measures of peer evaluation, however these spanned multiple projects or years rather than being embedded within the same course and project.[10, 29, 30, 32, 34, 35] These have also been quantitative in nature and have not triangulated data with student reflections, which might clarify the underlying causes of the documented benefits. The present study utilized repeated measures within a single project allowing peer evaluation ratings to be formative in nature, informing team interactions for the duration of the project. While the trends observed over time match the expectations based on previous research, the longitudinal data obtained for the present study extends that conducted by previous researchers in the number of peer evaluations used. In order to help understand the trends in peer evaluation, the present study gathered student reflections on the process. A clear understanding of how students perceive the process is requisite for effective use of the strategy and therefore, the research questions driving this inquiry are:

1. How do the results of peer evaluation of change over time in an extended collaborative project?
2. How do students describe the influence of peer evaluations on their experience working in teams?

Classroom research provides the benefit of being able to work with the same participants over the course of a project or term. In this way, the researchers create a research study that can evaluate trends over time.[13] The administration of peer evaluation surveys in the present study was structured to provide a repeated measures longitudinal (spanning eight weeks) investigation. The instructors administered peer evaluation surveys five times during a single project creating multiple change points and the opportunity to evaluate changes in student contribution after repeated feedback. Brutus and Donia[29] expressed interest in this type of study by stating “it

would be interesting to investigate the impact of this exercise [peer evaluation] over a longer period of time and establish how the improvements detected in this study evolve…over repeated uses of the system" (p. 658).

Another unique contribution of this work is the effort to provide a student perspective on peer evaluation through qualitative analysis. "Researchers [should not treat] collaboration as a 'black box', but zoom in the collaborative interactions in order to gain better understanding of the underlying mechanisms" (p. 12).[11] Several researchers emphasized the importance of understanding student perspectives on peer evaluations.[24, 33, 36] The qualitative responses obtained in this study create an opportunity to hear the students' voice and perceptions on peer feedback experiences in the course; in that way it is an indicator of how well these assessment opportunities are being integrated in the course. Brutus et al.[30] stated that "one of the main limitations of [their] study is that it does not specify what, in the PES [peer evaluation system] experience, underlies the detected effects. Questions remain as to which component(s) of students' educational experience actually contributed to their increase in confidence with observation" (p. 28). While previous studies have been able to demonstrate significant effects through repeated uses of peer feedback during team projects, this study aims to explore the underlying mechanisms that lead to those changes from the student perspective.

Methods

Context

This research study was situated in a required College of Technology course titled, "Design Thinking." Students in this course will engaged in critical analysis of real-world problems and global challenges. They demonstrated the ability to recognize opportunity and to take initiative in developing solutions applying the principles of human-centered design. Students practiced communicating and working on teams. Problems and solutions were examined from societal, cultural, and ethical perspectives.

The course incorporated a flipped and blended approach which was the result of a faculty course redesign supported by Purdue University's Center for Instructional Excellence three years prior to this study. The course was administered in 25 sections annually with a capacity of 40 students per section. Each section met for 50 minutes per week during the 16 week semester and had a significant work expectation prior to class accounting for the blend of face to face and distance interactions via Blackboard. Class time was primarily spent in small and large group interaction while work prior to class was the primary content delivery mechanism through readings, videos and application assignments.

Conceptually, students in the course experienced four cycles. In the first four weeks, students learned about the process of design and practiced applying design using a Human Centered Design approach modeled after Stanford's d.School approach and IDEO design firm methods. During the second four weeks, students engaged in design in small teams focused on a local crosswalk. Student investigation included observations, interviews, literature review for known issues at crosswalks related to pedestrian safety followed by an evaluation of existing safety improvements and brainstorming potential innovative safety mechanisms that might improve

pedestrian safety. Students evaluated their alternative solutions and presented their ideas to the class in written and oral form.

The individual, pair, and limited small group work in the first half of the course set the stage for the second half of the course where students established final project groups and identified a problem to work on. Group formation and topic selection were by student choice. Students identified potential problems to work on and posted their ideas around the classroom on white boards. Students then had time to walk around the room looking for a problem they would like to work on and begin to congregate around that problem which allows them to simultaneously identify a team and topic. During the final project, in the final eight weeks of class, students worked on problem definition, benchmarking, decision making, prototyping, feedback and refinement, and technical communication via written form and oral presentation culminating in a public presentation in the College's main building lobby. Student groups remained intact for the duration of the study, including the 5 instances of peer evaluation and the qualitative reflection.

The Comprehensive Assessment of Team Member Effectiveness (CATME)[37] peer evaluation tool was chosen for the course and administered with the final project groups paralleling the submissions of five key elements of the final project.[*] The CATME instrument is a behaviorally anchored rating scale that describes behaviors typical of various levels of performance.[38] The development, reliability, and validity of the instrument are described in Loughry, Ohland, and Moore[39] and Ohland et al.[38]; more information is also available through www.catme.org. Raters select the category of behaviors that most closely matches the actual behavior of each student (including themselves) on their team. Five scales of teamwork were included: Contributing to the Team's Work, Interacting with Teammates, Keeping the Team on Track, Expecting Quality, and Having Relevant Knowledge Skills and Abilities. The CATME interface asks students to rate themselves and their peers by selecting one of five behavioral descriptions per metric selected by the instructor. To consider Interacting with Teammates, for example (abbreviated for publication – more detail in the actual survey), does your peer: ask for or show an interest in teammates' ideas and contributions, do they simply respect and respond to feedback from teammates, or do they interrupt, ignore, boss or make fun of teammates?[37] The instrument quantifies these behavioral ratings such that high quality interactions receive a 5, average interactions receive a 3, and poor interactions receive a 1.

In preparation, students experienced the CATME evaluation process during the first half of the term prior to working in the final project group. Following this practice experience, the purpose of CATME was discussed as well as how to interpret the CATME results. Of the five CATME surveys administered in the final project, only the last three had impacts on student grades. CATME provides the instructor with a series of scores representing student contribution as indicated by team member evaluations. The scores are consolidated into one number ranging from 0-1.05 which was used to modify the group's score for each individual's grade on that project element. In a team where each agreed that team members contributed equally, all students received a score of 1.00 which was multiplied by the group's score for the deliverable resulting in each student receiving the same score. However, in cases where one or more students did not

[*] CATME also has a Team-Maker tool for team formation that was not used for this context; instructors only used the peer evaluation tool contained in the suite.

contribute equally, they would receive a higher or lower CATME contribution score which would adjust their individual scores after multiplying it by the group deliverable score.

At the conclusion of the term, students were asked to reflect on the semester in class. The reflection included a review of the learning objectives and an overview of the course in a large group format. Students were also asked in a whole group discussion format to identify elements of the course that were successful as well as what issues could be improved. Then after identifying opportunities to improve the course, students were asked to help the instructor prioritize them and identify potential solutions to high priority issues. After the discussion, students were presented with a series of individual written reflection questions based on improvements or changes made in previous semesters to provide indicators of success. One of the questions related to this study, "Please describe the impact peer evaluation (CATME) had on your contribution to the team's work." Students were provided with approximately 10 minutes to respond and were informed that their response would not be graded and would be used to improve future offerings of the course.

Sample

The sample represented by this study includes a subset of sections offered during the Spring of 2013 and Fall of 2013. Seventeen sections were chosen randomly out of 25 offered during 2013 representing four instructors, three were faculty and one was a graduate teaching assistant who was a licensed public school teacher. All students in the sections were included in the study; this was possible because the data analyzed for this study were gathered as a normal educational practice (IRB Exempt) in the course and analyzed after grades were issued for the term.

Five hundred and sixty one students were enrolled in the 17 sections. Two forms of data collected in this study were quantitative CATME scores over five repeated administrations and a qualitative response to an open ended reflection question which asked about the impact peer evaluation had on a student's contribution. Four hundred and fifteen students participated in this study because they responded to the voluntary reflection question at the end of the term. Of those participants, four hundred twelve completed at least one CATME survey and two hundred and eighty-three students completed all five CATME surveys. Students were predominantly male (84%). Most were freshmen (58%), some were sophomores (24%), juniors (14%) and seniors (4%). Most were white (73%) and domestic (90%).

Procedures

To address the first research question, investigating change over time, this study used a repeated measures ANOVA using five CATME survey scores representing team member effectiveness in the final project team. The second question was addressed using a qualitative analysis of student response provided insight to understand quantitative trends discovered in the repeated measures analysis. Analysis is discussed here in three sections. First, given that not all students completed all five iterations of CATME, a pre-analysis manipulation check was conducted to compare students who completed in all five CATME surveys (n=283) to those of missed one or more surveys (n=132). Demographic variables and SAT scores were compared in order to examine similarities and differences among the student two groups. Fifty-four students did not complete SAT or ACT scores and were omitted from this comparison. ACT scores were converted to SAT

equivalent scores are per a table published by the College Board.[40] Non-significant tests would indicate that there were no differences between groups.

Second, a repeated measures ANOVA was conducted to examine changes in CATME ratings over the five iterations. The assumption of sphericity was evaluated. If the assumption was violated, a Greenhouse-Geisser adjustment was made to the test degrees of freedom. Data from the 283 students who completed all five CATME surveys were divided into three categories across five measures, each ranging from 1-5:

> Self - Represents student's self-reported effectiveness and contribution in their team.
> Peer - Represents an average of the student's peer team member ratings of the student's effectiveness and contribution. This number does not include the self-rating.
> Of others - Represents each student's average rating of their peers, indicating a holistic representation student perception of their team mates.

Each category (Self, Peer, Of Others) was analyzed for changes across time separately using a repeated measures ANOVA.

Third, students' qualitative responses (n=415) were analyzed following a phenomenological approach to qualitative inquiry. The purpose of this approach is to understand the meaning of an experience or phenomenon for individuals involved.[41] Responses to the question, "Please describe the impact peer evaluation (CATME) had on your contribution to the team's work" were imported into NVIVO software for analysis and coded by four researchers over five iterations. This coding process was conducted using a combination of inductive analysis and the constant comparative method.[42] Through inductive analysis, themes emerged from the data analysis processes rather than being imposed on it through an a priori coding structure. The constant comparative method allowed the themes to be shaped and reshaped throughout the data analysis process before arriving at a final set of themes that best described the data.

In the first phase of coding, a random subset of 60 responses (~15%) were selected representing equal numbers from each of the instructor's sections from Fall and Spring sections. Four researchers participated in the analysis process. Each researcher reviewed the same 60 responses independently. Then each researcher individually used an emergent coding strategy to identify and categorize "meaning units" (p. 150)[41], or themes, in the data. The four researchers collaboratively discussed their themes and began to converge on a set of themes that represented the collective review. Theme definitions began to form and were documented. This approach was repeated twice more, each with 60 additional responses.

The remaining data (n=235 or ~55%) were divided into two halves. Two researchers applied the coding scheme to one half of the remaining data set and the other pair of researchers coded the other remaining half. Each researcher individually summarized the themes into meaning statements explaining the students' experience with peer evaluation. The two pairs convened to negotiate the summary to consensus and then the group of four met to establish an understanding of the data resulting from the coding. Data previously coded in the first three iterations (n=180 or ~45%) were re-coded individually by each of the four researchers using the finalized coding document and discussed collaboratively to verify the coding scheme.

Trustworthiness

While reliability and validity are essential for quantitative work, the believability and accuracy of qualitative work are described in terms of trustworthiness.[41] The rigor of qualitative investigation is, therefore, based on the procedures employed to show verification in the results. Evident in the previously described methodology are several steps taken to enhance the trustworthiness of the researchers' interpretation of qualitative results.[41] These are:

- *Prolonged engagement and persistent observation* – Members of the research team were involved in the course as instructors. This provided opportunities for regular observation and involvement with the student teams. Additionally, the regular involvement in class allowed for an understanding of the context for peer evaluation (i.e., researchers had an understanding of the assignments, course materials, and activities that might be referenced in student reflections).
- *Researcher Triangulation* – The research team worked collaboratively throughout the study, investigating student responses and discussing findings. This divergent and convergent process allowed for a collective review of the findings and a "check of the research process." The debrief results and evolution of themes were well documented.
- "*Negative case analysis*" – Convergence on themes for this report was based on review of all cases student experience. Coding included negative cases where students presented evidence contrary to the common experience. This information was used to refine the hypothesis as the research advanced.

These procedures support the trustworthiness of this work. In addition, they meet the standard recommended by Creswell[41] that "qualitative researchers engage in at least two [strategies out of eight presented] in any given study" (p. 203).

Analysis

Pre-analysis Checks

Students who completed all five CATME surveys were compared to those who completed less than five. These two groups of students were included in the qualitative analysis and were compared here to identify similarities on gender, class rank, ethnicity, international status and SAT scores. T-tests and chi-square tests where used to compare individual students who completed all five CATME iterations (n = 283) to those who completed some number less than five (n = 132). Results are summarized in Tables 1 and 2 below:

Table 1. ***Chi-Square Tests to Compare Demographic Variables.***

Demographic Variable		Completed All CATME Surveys		Pearson χ^2
		No	Yes	
Gender	Male	115 (87.1%)	234 (82.7%)	$\chi^2(1) = 1.32$, $p = .250$
	Female	17 (12.9%)	49 (17.3%)	
Class Rank	Freshman	79 (59.8%)	162 (57.2%)	$\chi^2(3) = .746$, $p = .862$
	Sophomore	28 (21.2%)	69 (24.4%)	
	Junior	20 (15.2%)	39 (13.8%)	
	Senior	5 (3.8%)	13 (4.6%)	
Ethnicity	White	97 (73.5%)	204 (72.1%)	$\chi^2(1) = .089$, $p = .776$
	Other	35 (26.5%)	79 (27.9%)	
International Status*	Domestic	125 (94.7%)	249 (88.0%)	$\chi^2(1) = 4.55$, $p = .033$
	International	7 (5.3%)	34 (12.0%)	

Note. n = 283 for completed surveys, n = 132 for non-completed surveys, *$p < .05$.

Table 2. ***Independent-Samples T-tests to Compare SAT Composite and SAT Writing Scores.***

Dependent Variable	Completed All CATME Surveys				*t*	*df*	*p*	*d*
	No		No					
	M	*SD*	*M*	*SD*				
SAT Composite	1106.12	129.93	1131.83	137.13	-1.71	359	.088	.18
SAT Writing*	511.74	78.47	526.58	79.67	-1.68	359	.094	.18

Note. n = 283 for completed surveys, n = 132 for non-completed surveys, *$p < .05$.

To summarize, the pre-analysis check indicates that there are not substantive differences between the completers (i.e., completed all five iterations of CATME) and non-completers (i.e., completed less than five iterations of CATME). Relative to demographic variables, the only significant difference was that there was a higher percentage of international students who completed CATME than did not complete it. T-tests to examine differences based on SAT performance found that there were no significant differences between completers and non-completers relative to SAT Composite or SAT Writing scores. Similar to the chi-square tests, these results indicate that the completers and non-completers are relatively similar and supports a collective analysis of student responses.

Changes Over Repeated CATME Experiences

For "Self" rating, after applying the Greenhouse-Geisser adjustment to compensate for the violation of sphericity, there was a significant, linear trend over time, $F(3.38, 953.72) = 12.31$, $p < .001$. This significant difference was associated with a small effect size, $\eta^2 = .042$. There was an increase in self-rating from across the five measures as shown in Figure 1.

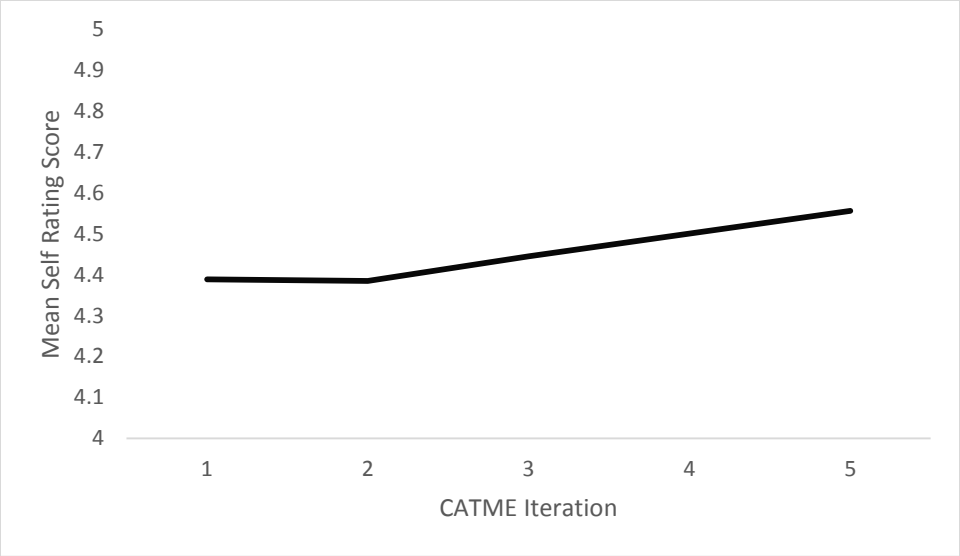

Figure 1. **Changes in self-ratings over five CATME iterations**

For "Peer" ratings (ratings made about a student by peers in their team), after applying the Greenhouse-Geisser adjustment to compensate for the violation of sphericity, there was a significant, linear trend over time, $F(3.48, 981.19) = 13.18$, $p < .001$. This significant difference was associated with a small effect size, $\eta^2 = .045$. There was an increase in peer-rating across the five time points as shown in Figure 2.

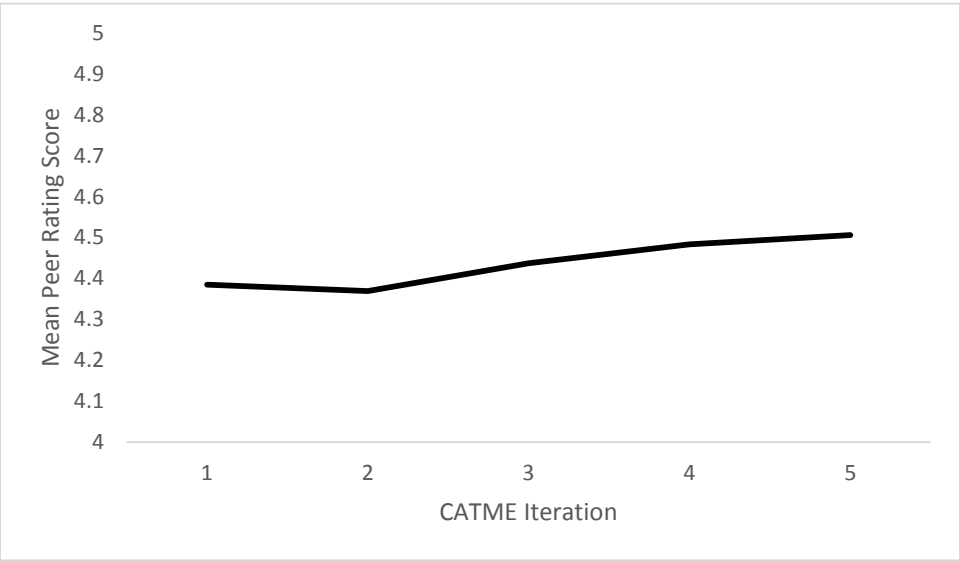

Figure 2. **Changes in ratings from others over five CATME iterations**

For "Of Others" ratings (ratings by a student about their team mates), after applying the Greenhouse-Geisser adjustment to compensate for the violation of sphericity, there was a significant, linear trend over time, $F(3.23, 910.60) = 7.88$, $p < .001$. This significant difference was associated with a small effect size, $\eta 2 = .027$. There is an increase in ratings of others across the five time points as shown in Figure 3.

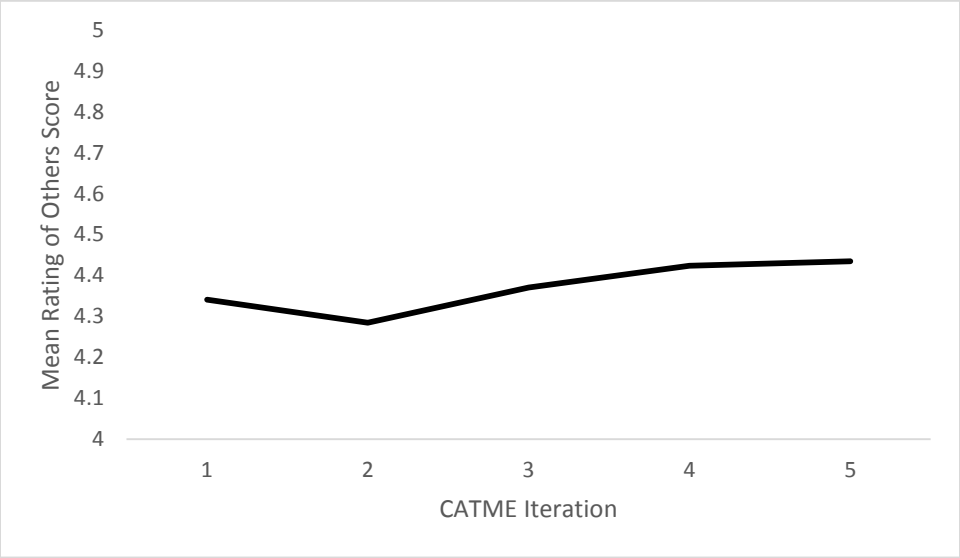

Figure 3. **Changes in ratings of others over five CATME iterations**

Qualitative Results

Quantitative data revealed an increasing pattern in peer evaluation: student ratings of themselves, from peers, and of other team members grew through prolonged work in the team. This growth was of interest and qualitative analysis was used to further explore the phenomenon. The research team made meaning from the coded themes by considering them individual pieces of a larger mosaic and distilled four distinct elements of the student experience which were:

1. Students felt a sense of motivation (social pressures) because they were held accountable
2. Students seek external validation for their efforts
3. Students intend to change behaviors if peers are dissatisfied with their contribution
4. Students believed peer evaluation improved overall team function

This summary of meaning was the overwhelming emergent theme, but did not reflect every student's response. There were counter examples, but of the 415 messages coded, the vast majority collectively articulated one or more of these four experiences.

Theme 1. Students felt a sense of motivation (social pressures) because they were held accountable.

The students indicated that they felt motivated to be part of a team and did not want to let their teammates down, as articulated by a student: "I felt a sense of responsibility when doing my work I might not have had otherwise." They appeared to be more concerned with their peers' evaluations than their instructor's evaluation indicating that grades were less of a concern then

how their peer's perceived their contribution. Some students articulated the "social pressures" they feel to contribute to group work. While this may be the case regardless of the use of peer evaluation, CATME provides them with tangible feedback, which legitimates and intensifies this social pressure. One student commented, "Peer evaluation made sure that I was always doing my best on my work." CATME gave students who were genuinely concerned with meeting their peer's expectations some concrete feedback on their performance, which made it so they don't have to guess about what their classmates think of them. While most students articulated a sense of motivation and generally attributed it to social pressures, some students did share that they would have been motivated regardless. Less than 1% of the students expressed negative concerns about peer evaluation and how it may have actually decreased their sense of motivation citing concerns of fairness and bias in the process.

Theme 2. Students seek external validation for their efforts.

While students generally reported being motivated, they also sought validation or confirmation that their efforts were recognized and sufficient. Students used the information from peer evaluation as a barometer (confidence, validation, confirmation) for how they were doing individually, and how their group was functioning as a unit. They were concerned with how others perceived their effort. Students discussed their peers' evaluation more often than they discussed the impact it had on their grade. Students commonly articulated a feeling of "doing what I'm supposed to be doing" from the perspective of their peers. The data seemed to suggest that students considered their peer feedback more so than they did their self-reflection. CATME peer evaluation allowed students to communicate in a non-conformational way allowing for this open anonymous dialog between teammates about contribution. This may be particularly helpful for those students who are not comfortable discussing performance openly with their teammates as indicated by another student:

> It helped me see the way my team thought I was doing. It helps you realize the truth rather than what you think you're doing. It can also help you feel better about yourself if you team says you did better than you think you did.

Another student commented:

> After completing each project and seeing the catme results for the project help me feel that I was helping my team out. I was always ranking myself low because I felt that I did not contribute much toward the project, but that was not how my teammates felt which made me feel good. It made me want to continue to try hard at the projects.

Theme 3. Students articulated an intention to change behaviors if peers are dissatisfied with their contribution.

One student shared, "Catme helped me see what I needed to do to improve how I worked with my group so I could be a better teammate." Students discussed intent for making adjustments in response to the feedback they received. Few, however, actually articulated implementing those changes. Students indicated that peer feedback was taken seriously and students would improve if their teammates suggested it was needed. A student stated, "It allowed us to see what our team members thought of our work which was good because we knew if we needed to change something and we were then able to make our work better." Another student summarized how peer evaluation provided feedback on the team functioning and how that impacted their contributions:

> Peer evaluation helped in finding out the strengths and weaknesses in a team. It brought out the qualities that the team lacked and needed to do better. Peer evaluation gives an honest opinion of people on how they are doing in a team and what must be done to further do well in a team. It gives a rating of how and where the team is going. Peer evaluation builds confidence in people to do well in their team. One does not want to get marked badly in an assignment and hence this pressure keeps him/her going well in a team.

Theme 4. Students believed peer evaluation improved overall team function.

Students articulated that they were either in a functional team or that CATME helped their team become functional, in one student's words:

> The peer evaluation has motivated the entire group to do well, since peer evaluation affects final gradings. Overall I think catme boosted our entire group's behavior and contribution. Once noticing the importance and significance of peer evaluation, we strived for better performance and never missed meetings etc.

Students generally appreciated the process of evaluation and found it helpful for their team in some way as described by a student, "CATME had a good impact on our teamwork. Based on that evaluation, everyone knew what was expected and what was discouraged". A few students mentioned that it was their first experience in a successful team and that in previous teams they were always forced to do all the work because their peers would not contribute. The feedback from teammates was mentioned by students, but also there was a sense of importance of having a voice in reflecting on the team's work. Students appreciated being able to articulate their perspective on how the team was functioning such that CATME served as a venue for them to share (not only receive) feedback, as one student commented: "The CATME surveys were very helpful in getting our team in sync without causing disturbances."

Discussion

The purpose of this study was to identify if students' self, peer and evaluation of others changed across five repeated measures of peer team member effectiveness evaluation in an eight week final project. After discovering significant increases, qualitative data were analyzed to determine causes from the student's perspective and better understand why changes may have occurred. This study was set in a design course which positioned students in a team, problem based learning environment. In the final project, teams of four to five students worked for the second half of the semester (about eight weeks) on a cumulative design project. Typically deliverables were submitted weekly by the team. On five occasions after key elements of the design project were completed, each student was asked to rate themselves and their peers. The evaluation was considered formative as students received feedback and continued to work in the team prior to the next evaluation cycle. Data suggest that students rate themselves significantly higher across these five repeated measures indicating they believe they are a more effective teammate as time passed in the course. This finding is congruent with literature suggesting team skills can be learned.[7-10] This improvement may represent actual improvement of contribution, although it was not measured in this study. Alternatively, it may represent an increase in student self-confidence as students may be humble initially and discover through the feedback mechanism that they are a valued member of the group. Kaufman et al.[7] suggested that self-evaluation deflation may be a concern in that students feel they could have done better while their peers recognize the

significant contributions they have made. The use of repeated peer evaluation and feedback cycles was not investigated as a potential intervention in the Kaufman study, but findings in the current study suggest it may have a positive influence on providing students will the confidence to recognize the contributions they have made. It is also possible that students recognize that rating themselves higher has a positive impact on their grade and therefore it is academically valuable to bias their self-ratings, although the Kaufman study indicated that concern was generally unfounded.[7]

Peers' ratings of each student also show a significant increase over time indicating that team members generally report the students are improving in their effectiveness and contribution. This improvement over time corresponds with improvement reported by the student and provides triangulated evidence that students may be learning from their previous contribution, reflection, feedback and making changes in their behavior accordingly. Students are also reporting significant growth among their peers as they feel their teammates were contributing significantly more with each iteration of feedback and peer evaluation.

The qualitative data support and provide a potential explanation for the quantitative data. Improvements in student self-evaluation scores are consistent with students stating increases in confidence after peers validate their efforts. Also, students indicated that they would make changes to their behavior if the feedback indicated it was necessary. They may have followed through on those changes, which explain the increases in self and peer scores. The sheer existence of peer evaluation as a process may have motivated students to improve over time as they wanted to please their peers. The desire to make their peers happy may have led to more functional teams based on the existence of the instrument. The team members who claimed that peer evaluation had no effect on their team's work may have not realized that their peers were performing because of the evaluation process and, while the benefits may have been real, the connection was not apparent. Since the impact of peer evaluation depends on individual perception and response to the feedback, the research team continues to support the inclusion of peer evaluation in team projects.

Limitations

Although there is correspondence between student responses and the observed trends in quantitative data, the research team recognizes several limitations which warrant future research in this vein. One limitation relates to the nature of the research, which was not experimental. While data support conclusions made about the benefit of formative peer evaluation, teams may have had other mechanisms that influence their work which were not captured in the study. Additionally, the process of working together for more time, and subsequently becoming more comfortable as a team, may explain a portion of the growth in teamwork. We are unable to discriminate between these outside effects and the effect of peer evaluation.

In the Design Thinking course, we acknowledge that the "average" team was functional, as illustrated by the initially high ratings among teams. The research team is unable to draw conclusions about the impact of peer evaluation on nonfunctioning teams. However, this evidence supports the instructional design of the course which seems to be working; to a certain extent teams wouldn't be as functional in a poor quality environment. It also demonstrates that

including peer evaluation in a course may draw out incrementally better teamwork, helping students be even more prepared as job candidates.

Recommendations

The research team recognizes the dip in ratings after the first evaluation, which presents an opportunity for further research. As mentioned, the teams were generally functional and an investigation into the change in pattern between the first and second evaluation and evaluations two through five may be insightful. Members of the research team hypothesize that it may be related to calibration with team members or growing in comfort and openness with the team as they work together more. Also, in response to the limitations of the study, future work could frame an investigation into how students experience peer feedback as an experimental study. An experimental design might also use a performance indicator to show whether peer evaluation actually had an impact on performance or if the benefit was perceptual. Finally, multilevel modeling methods could be considered because the data are nested within teams and classes. This analysis could help evaluate the conditions under which CATME has an impact on team performance. However, the researchers will continue the use of peer evaluation as the present evidence seems to indicate a positive effect on students in a diverse classroom setting.

Based on the data and analysis presented here, the research team recommends the use of peer evaluation multiple times during a long term project. The data here suggest that students improve over repeated administrations of peer evaluation. This is congruent with learning theories related to feedback wherein students make an attempt, reflect (during the peer evaluation process), receive feedback and attempt again. This process can be applied as an overlay to existing coursework in a variety of collegiate and other educational settings. This study was unable to determine an optimum number of repeated peer contribution evaluations, but the trajectory of improvement in the five repeated measures shows a generally positive slope indicating that additional evaluations may result in higher scores. Alternatively, it is possible the team is maturing over time and the instrument simply measures the maturation process, but does not actually cause the change. However, this alternative explanation is in conflict with the qualitative responses students provided. Students indicated they are very concerned with social pressures to perform, will improve if their peers communicate the need, and that their sense of value is validated by their peers.

References

1. Pellegrino, J. W., & Hilton, M. L. (Eds.). (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press.
2. ABET Engineering Accreditation Commission. (2013). *Criteria for accrediting engineering programs*. Baltimore: Accreditation Board for Engineering and Technology (ABET). Retrieved from: http://www.abet.org/eac-criteria-2014-2015/
3. National Academy of Engineering. (2004). *The engineer of 2020: Visions of engineering in the new century*. Washington, DC: National Academies Press.
4. Loughry, M. L., Ohland, M. W., & Woehr, D. J. (2014). Assessing teamwork skills for assurance of learning using CATME team tools. *Journal of Marketing Education, 36*(1), 5-19. doi: 10.1177/0273475313499023

5. National Association of Colleges and Employers. (2013, October 2). Job outlook: The candidate skills/qualities employers want. Retrieved January 21, 2015, from https://www.naceweb.org/s10022013/job-outlook-skills-quality.aspx
6. American Management Association. (2012). AMA 2012 critical skills survey. Retrieved January 21, 2015, from http://www.amanet.org/uploaded/2012-Critical-Skills-Survey.pdf
7. Kaufman, D. B., Felder, R. M., & Fuller, H. (2000). Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education, 89*(2), 133-140. doi: 10.1002/j.2168-9830.2000.tb00507.x
8. Felder, R. M., & Brent, R. (2001). Effective strategies for cooperative learning. *Journal of Cooperation & Collaboration in College Teaching, 10*(2), 69-75.
9. Burdett, J. (2003). Making groups work: University students' perceptions. *International Education Journal, 4*(3), 177-191.
10. Fellenz, M. R. (2006). Toward fairness in assessing student groupwork: A protocol for peer evaluation of individual contributions. *Journal of management education : a publication of the Organizational Behavior Teaching Society., 30*(4), 570-591. doi: 10.1177/1052562906286713
11. Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1-19). Oxford: Elsevier.
12. Raelin, J. (2006). Does action learning promote collaborative leadership? *Academy of Management Learning & Education, 5*(2), 152-168. doi: 10.5465/AMLE.2006.21253780
13. Loyd, D. L., Kern, M. C., & Thompson, L. (2005). Classroom research: Bridging the ivory divide. *Academy of Management Learning & Education, 4*(1), 8-21. doi: 10.5465/AMLE.2005.16132533
14. Kraiger, K. (2008). Transforming our models of learning and development: Web-based instruction as enabler of third-generation instruction. *Industrial And Organizational Psychology-Perspectives On Science And Practi, 1*(4), 454-467. doi: 10.1111/j.1754-9434.2008.00086.x
15. Verzat, C., Byrne, J., & Fayolle, A. (2009). Tangling with spaghetti: Pedagogical lessons from games. *Academy of Management Learning & Education, 8*(3), 356-369. doi: 10.5465/AMLE.2009.44287936
16. Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? *Small group research, 36*(5), 555-599. doi: 10.1177/1046496405277134
17. Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive and individualistic learning* (4th ed.). Massachusetts: Allyn and Bacon.
18. Shankar, P. G., & Seow, J. L. (2010). The association between accounting students' lone wolf tendencies and their perceptions, preferences and performance outcomes in team projects. *Journal of Accounting Education, 28*(2), 75-84. doi: 10.1016/j.jaccedu.2011.03.004
19. Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of student centered learning, 2*(1), 9-34.
20. McCorkle, D. E., Reardon, J., Alexander, J. F., Kling, N. D., Harris, R. C., & Iyer, R. V. (1999). Undergraduate marketing students, group projects, and teamwork: The good, the bad, and the ugly? *Journal of Marketing Education, 21*(2), 106-117. doi: 10.1177/0273475399212004
21. Deeter-Schmelz, D. R., Kennedy, K. N., & Ramsey, R. P. (2002). Enriching our understanding of student team effectiveness. *Journal of Marketing Education, 24*(2), 114-124.
22. Hernandez, S. A. (2002). Team learning in a marketing principles course: Cooperative structures that facilitate active learning and higher level thinking. *Journal of Marketing Education, 24*(1), 73-85. doi: 10.1177/0273475302241009
23. McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. *Team effectiveness and decision making in organizations*, 9-45.
24. Thomas, G., Martin, D., & Pleasants, K. (2011). Using self-and peer-assessment to enhance students' future-learning in higher education. *Journal of university teaching and learning practice, 8*(1).
25. Fink, L. D. (2003). A self-directed guide to designing courses for significant learning. *University of Oklahoma*, 27.
26. Hughes, R. L., & Jones, S. K. (2011). Developing and assessing college student teamwork skills. *New Directions for Institutional Research, 2011*(149), 53-64. doi: 10.1002/ir.380
27. Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
28. Brew, A. (1999). Towards autonomous assessment: Using self-assessment and peer assessment. In S. Brown & A. Glasner (Eds.), *Assessment matters in higher education: Choosing and using diverse approaches*. Buckingham, UK: Open University Press.

29. Brutus, S., & Donia, M. B. L. (2010). Improving the effectiveness of students in groups with a centralized peer evaluation system. *Academy of Management Learning & Education, 9*(4), 652-662. doi: 10.5465/AMLE.2010.56659882
30. Brutus, S., Donia, M. B. L., & Ronen, S. (2013). Can business students learn to evaluate better? Evidence from repeated exposure to a peer-evaluation system. *Academy of Management Learning & Education, 12*(1), 18-31. doi: 10.5465/amle.2010.0204
31. Chapman, K. J., & van Auken, S. (2001). Creating positive group project experiences: An examination of the role of the instructor on students' perceptions of group projects. *Journal of Marketing Education, 23*(2), 117-127.
32. Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The effects of peer feedback on team member behavior. *Group & Organization Management, 22*(4), 508-520.
33. Cassidy, S. (2007). Assessing 'inexperienced' students' ability to self-assess: Exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education, 32*(3), 313-330. doi: 10.1080/02602930600896704
34. Mayo, M., Kakarika, M., Pastor, J. C., & Brutus, S. (2012). Aligning or inflating your leadership self-image? A longitudinal study of responses to peer feedback in mba teams. *Academy of Management Learning & Education, 11*(4), 631-652. doi: 10.5465/amle.2010.0069
35. Moriarty, P., & Buckley, F. (2003). Increasing team emotional intelligence through process. *Journal of European industrial training, 27*(2-4), 98-110.
36. Davies, P. (2003). Closing the communications loop on the computerized peer-assessment of essays. *Research in Learning Technology, 11*(1).
37. Ohland, M. W., Bullard, L. G., Felder, R. M., Finelli, C. J., Layton, R. A., Loughry, M. L., . . . Woehr, D. J. (2005). CATME. West Lafayette, IN.
38. Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Finelli, C. J., Layton, R. A., . . . Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning & Education, 11*(4), 609-630. doi: DOI 10.5465/amle.2010.0177
39. Loughry, M. L., Ohland, M. W., & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement, 67*(3), 505-524. doi: 10.1177/0013164406292085
40. The College Board. (2009). ACT and SAT concordance tables. *RN-40*.
41. Creswell, J. W. (1998). *Qualitative inquiry and research design : Choosing among five traditions*. Thousand Oaks, Calif.: Sage Publications.
42. Patton, M. Q. (2015). *Qualitative research and evalaution methods* (4th ed.). Thousand Oaks, CA: Sage.