Using Frame-of-Reference Training to Improve the Dispersion of Peer Ratings in Teams

Daniel M. Ferguson, Chad Lally, Hilda Ibriga Somnooma, Olivia Murch, Matthew W. Ohland
Engineering Education, Computer and Electrical Engineering, Statistics, Biomedical Engineering, Engineering Education
Purdue University
West Lafayette, IN
dfergus@purdue.edu, clally@purdue.edu, hibriga@purdue.edu, omurch@purdue.edu, ohland@purdue.edu

***Abstract*— An engineer's ability to work in teams is critical to their engineering career and often a significant factor in a corporate hiring process. Recognizing this need, and motivated to demonstrate outcomes required for accreditation by ABET, most U.S. undergraduate engineering programs include team-based courses in their curricula. Hundreds of these engineering programs use the CATME application (the Comprehensive Assessment of Team-Member Effectiveness), which asks students to assess the teamwork behaviors of their peers. Our goal was to determine how effectively engineering students rated their peers and whether a proven method for improving the use of a rating schema, Frame of Reference training, would improve engineering students' peer rating behavior. We discovered that engineering students do not materially change their peer ratings based on their teaming experiences and we verified that repeated Frame of Reference trainings have a significant impact on peer rating dispersion and potentially the quality of engineering students' peer ratings.**

Keywords—frame-of-reference, teamwork, peer ratings, dispersion

I. INTRODUCTION

According to a report in the Wall Street Journal [1] teamwork behavior is a key skill many companies, as diverse as NASA [2], look for when hiring new employees. Chen argues that many students entering into the workplace lack key teamwork skills [3]. In addition teamwork skills training has become more prevalent throughout college programs due to the addition of teamwork accreditation requirements in engineering [4], in business [5] and other accreditation bodies and employer recognition of the importance of teamwork skills [6,7]. Teamwork is defined as "cooperative or coordinated effort on the part of a group of persons acting together" [8]. Teamwork is not only important in schools and colleges, but also within a professional setting. Many employers throughout various industries require team based work and projects [3].

Our primary research goal was to determine how much engineering students were varying their ratings of themselves and their teammates in peer reviews and to what extent does frame-of-reference training on peer rating dimensions improve the quality of peer evaluations of teamwork behaviors by engineering students? Frame-of-Reference (FOR) training is an accepted and proven method for improving the use of rating schema documented in Organizational Behavior literature. Improved quality in peer reviews in this paper is defined as more dispersion in the ratings of teamwork dimensions.

CATME (Comprehensive Assessment of Team Member Effectiveness) is a web-based tool created for academic teamwork environments and is used to assist students in giving peer feedback to their team members [9]. CATME is constructed around five behavioral dimensions: Having Relevant Knowledge, Skills and Attributes (KSAs), Contributing to the Team's Work, Interacting With Teammates, Keeping the Team on Track, and Expecting Quality [10,11]. These dimensions are defined as follows:

- **Having** (H) relevant KSAs refers to the base knowledge of individual team members. It means having the required skills to solve the problems at hand, or an individual being willing to learn the skills he/she lacks.
- **Contributing** (C) to the Team's Work is being able to add value to your team's work/project. It includes completing your portion of the work in a timely fashion.
- **Interacting** (I) with teammates refers to the various ways individuals communicate with and show respect for their teammates. Encouraging every team member to give their opinion and ensuring their voice is heard are part of this.
- **Keeping** (K) refers to alerting the team to conditions that could affect the team's success.
- **Expecting** (E) quality is about both expressing the believe that the team can do a good job and encouraging the team to do its best.

Every aspect of the five teamwork dimensions is equally important to team success and a critical element in the peer reviews [10].

Peer reviews facilitate better learning outcomes in upper level education, encouraging students to continue their engagement with constructive team behavior in future team activities [12]. A behavior peer review is an evaluation of an individual's contribution to a work activity by their peers, either as students or as professionals in industry [13]. Peer reviews help to teach individuals how to act in teams and how to evaluate one another's performance. Peer reviews, as facilitated by the CATME system, potentially point out behaviors where an individual excels and areas where he/she may need improvement [14,15].

Frame-of-Reference is defined as a construct of values, views, or concepts that a group of people build to serve as a reference structure in order to understand, assess or evaluate an observed phenomenon [16]. CATME contains a FOR that students use to evaluate teamwork behavior by team members. CATME's FOR creates a basis of understanding expected teamwork behavior, better than expected teamwork behaviors and worse than expected teamwork behaviors. CATME's FOR therefore defines teamwork expectations and is a guide for evaluating the teamwork behavior of team members.

Accessing de-identified student peer review data in over 1,000 engineering students in 10 sections of a team-based First Year Engineering class we observed two problems in engineering student peer reviews:

1. Students often gave all team members on a CATME dimension the same rating.
2. Students often gave all team members the same rating across multiple CATME dimensions.

While it is clearly possible that an individual exhibits the same level of performance across all five CATME dimensions, it is not likely, and similarly it is unlikely that there are no differences in teamwork performance among all team members on any of the five CATME dimensions [17,18]. There are several possible explanations for student peer rating behavior, e.g. the Halo Effect, [19]. To potentially influence peer rating behavior by making it more distributed or diverse and therefore possibly more constructive, we offered FOR Training [20] as a corrective intervention regarding the observed student peer review rating behavior [21]. More effective rating was defined as having more diverse ratings across team members and across the five CATME team behavior dimensions than observed in our control groups who received no FOR training.

In this paper we discuss our research population, FOR training on the CATME schema, experimental procedures, analysis structure, analysis processes, findings and conclusions.

II. RESEARCH METHODS

A. Research Population

FOR materials were used in 6 of 16 sections of First-Year Engineering (FYE) students at a major Midwestern university in the fall of 2015. FYE students at this institution number over 1,600 including 23% women and 6% minorities. Each FYE section was composed of up to 120 students and 10 of the 16 FYE sections were reserved as the control group with over 1000 students with 6 additional FYE sections exposed to 2 different versions of the FOR training. All of the six experimental sections and the 600+ students included in the two experimental groups were exposed to the FOR materials as an in-class training activity since participation in peer reviews was a required FYE grading activity. Peer review individual results were a component of the individual's FYE course grade in all FYE sections. Five different instructors were the instructors of record for the six experimental FOR training sections.

B. FOR Training

FOR training consisted of an instructor explaining the five CATME peer rating dimensions using PPT slides, the instructor sharing word [or video illustrated] examples of using the peer rating framework, a class peer rating problem exercise using word descriptions of behavior, done on paper and reviewed by the instructor with the class as a whole, and finally an online CATME peer rating quiz done individually by the students. CATME teamwork behaviors were categorized as the five dimensions and also described as occurring at five performance levels running from well below expectations (=1) to well above expectations (=5). The online CATME rating quiz consisted of 10 statements to be categorized and level-rated (20 answers) and five summary level questions by category. The CATME rating quiz was completed only at the 1st of the FOR sessions and completed by all of the six experimental FYE sections.

C. Experimental Procedure

All six 120 engineering student FYE sections received one FOR introductory training session of approximately 25 minutes and three of those experimental FYE sections received three additional FOR trainings of 25 minutes per session. The second to fourth FOR trainings for these later three sections emphasized only three of the CATME teamwork dimensions: Contributing, Interacting and Having. One instructor conducted all FOR trainings in all six experimental sections. FYE students are required to do three CATME peer reviews of their teammates per semester and the CATME peer reviews are scheduled concurrently with the completion of three major team project assignments. The FYE team projects are designed to require substantial team interactions. All FOR training was positioned after the completion of the first peer review/team project assignment but before the second peer review/team project assignment was due to be submitted. This scheduling of the FOR training provided comparisons of experimental to control peer review data for three different peer reviews. Data from two peer reviews was collected after all the FOR training was conducted.

CATME peer reviews are conducted online and students receive peer review feedback in an online form. Peer review feedback is shown as pointers to word descriptions of behaviors similar to their own rating or better or worse than they were rated on average by their peers as shown in Figure 1 for the CATME dimension Interacting. No numbers are provided to students and behavior descriptions appropriate to improving their average ratings are cited. For each of the five CATME peer rating dimensions students see their own ratings of themselves, their average ratings by their teammates and the average ratings for all team members as shown in Figure 1. Students see only average ratings interpreted as pointer placements on feedback screens and do not see their actual numeric ratings by their teammates as all ratings are held confidential, although numeric rating data is provided to instructors.

This work is based on National Science Foundation Award 1431694.

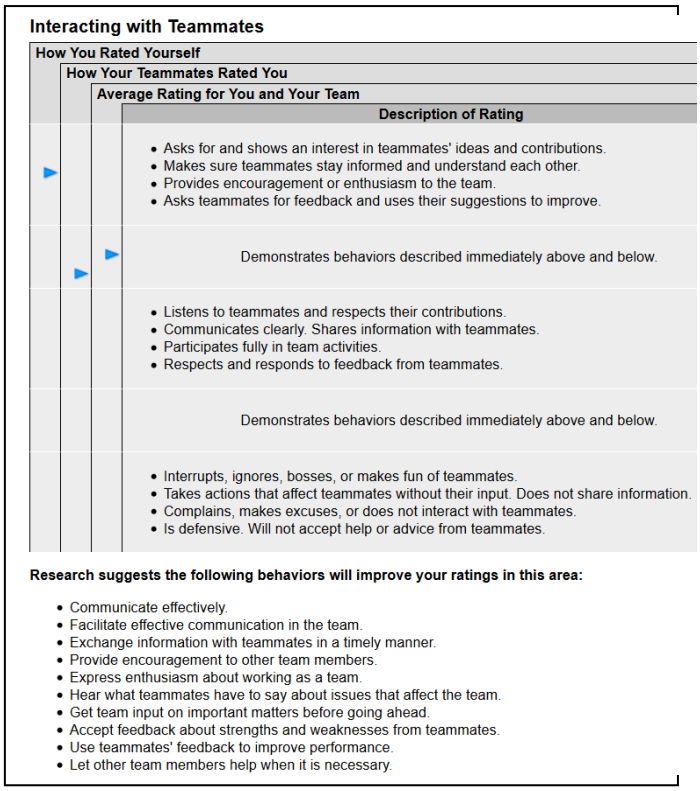

Fig. 1. Student feedback for the Interacting Dimension

			Rater 1					Rater 2		
Team ID	Section	Rater #	C	I	K	E	H	C	I	K
023-02	23	1	3	4	3	3	4	3	3	4
023-02	23	2	4	5	3	4	3	4	3	4
023-02	23	3	5	4	4	5	5	5	3	4
023-02	23	4	3	4	4	4	4	3	3	3

Fig. 2. Sample instructor raw data display: C = Contributing, I = Interacting, K = Keeping, E = Expecting, H = Having

D. Analysis Structure

For analysis of the peer review data a standard deviation of each student's ratings of themselves and each team member across the five CATME dimensions was calculated along with the standard deviation of a student's rating of themselves and each team member on each dimension. For a team of four students and the five CATME dimensions this is a matrix of 20 data points as shown in Figure 2.

Each FYE section has 30 teams and the standard deviations for each student for their ratings of team members on the five CATME dimensions were calculated and referred to as the dispersion factor for the CATME dimensions for that student. Figure 2 shows the ratings of one student for themselves in numeric form and their other team members. We calculated a standard deviation for each of the four rows in Figure 2 and then calculated an average of those four numbers and repeated the process for Rater 2, etc. The resulting dispersion matrix for all students across all the experimental sections in each of the two interventions was placed in a dispersion matrix. A similar calculation was done to create a dispersion matrix including all the students in the 10 control sections. This dispersion calculation procedure was repeated for each of the three FYE peer reviews for all 16 sections of approximately 1600 students and 400 teams. Similar calculations were done to create a mean absolute score for each student and the mean scores were placed in a matrix of mean scores for each student in the experimental and the control samples.

E. Data Analysis Processes

In this analysis we used a repeated measures ANOVA [22] with experimental sections (Control, 1-FOR, 4-FOR) between subject factor and peer review times (1, 2, 3) and within subject factor. An interaction term (experimental sections by peer review times) was also added to the model in order to enable pairwise comparison for all possible combinations of experimental sections and peer review times. The dependent variables were the standard deviations (dispersion factor) computed for each student. Pairwise comparisons of the mean dispersion factors for each experimental section at each peer review time were done and p-values were adjusted using Tukey's correction method. The data analysis was generated with PROC MIXED using SAS software Version 9. Copyright © [2002-2012] SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA. Table 1 summarizes the results of six selected ANOVA comparisons.

The first four columns of Table 1 summarize the six sets of repeated measures ANOVA [23] comparisons that were prepared using the Statistical Analysis Software, SAS24. Analysis was done across the experimental and control samples for peer reviews 1, 2, and 3. 1-FOR is a label that identifies samples where only one FOR training was conducted. 4-FOR is a label that indicates the samples where four FOR trainings were conducted. The comparisons of the control sample at the three peer review times were done to determine if there were overall changes in the peer reviews due to all the course and team activities that were occurring in all FYE sections. The comparisons of the single (1-FOR) and four (4-FOR) training session interventions to the control sample and the comparisons of the two interventions at times 1 to 2, 2 to 3 and 1 to 3 were done to see if any intervention differences could be identified. Finally the two interventions were compared at times 1, 2 and 3 to see if the peer rating behaviors in the two experimental groups differed significantly.

III. FINDINGS

Table 1 contains the ANOVA comparison results for the six different sets of comparisons. For each study sample we calculated the p value of the dispersion distribution comparisons and the change in the average standard deviations for all rating categories across all team members. This analysis step was done to determine if there had been a significant dispersion change in the combined peer and self-ratings. We then calculated the p values when comparing the actual average peer value rating matrices and show the mean changes for the actual ratings. This analysis step was done to determine both the direction and significance of the rating pattern changes.

TABLE I. ANOVA COMPARISON RESULTS FOR EXPERIMENTAL AND CONTROL SAMPLE DISTRIBUTIONS

Experimental sample	Peer Review Time	Session	Peer Review Time	p-value Δ Dispersion	Δ Standard Deviation	p-value Δ average rating	Δ Average rating
Control	1	Control	2	**< 0.0001**	-0.1177	**< 0.0001**	0.0455
Control	1	Control	3	**< 0.0001**	-0.1657	**< 0.0001**	0.0921
Control	2	Control	3	**< 0.0001**	-0.0480	0.1764	0.0466
1-FOR	1	1-FOR	2	**0.0013**	-0.0710	0.9996	-0.0190
1-FOR	1	1-FOR	3	**< 0.0001**	-0.1506	1	0.0035
1-FOR	2	1-FOR	3	**0.0002**	-0.0796	0.9986	0.0225
4-FOR	1	4-FOR	2	**0.0139**	0.0589	**< 0.0001**	-0.4754
4-FOR	1	4-FOR	3	0.9891	0.0162	**< 0.0001**	-0.3220
4-FOR	2	4-FOR	3	0.2139	-0.0427	**< 0.0001**	0.1534
Control	1	1-FOR	1	0.9915	-0.0171	0.9637	-0.0173
Control	2	1-FOR	2	0.8017	0.0296	0.3855	-0.0818
Control	3	1-FOR	3	1	-0.0020	0.0923	-0.1059
Control	1	4-FOR	1	0.5015	-0.0372	0.5322	0.0124
Control	2	4-FOR	2	**< 0.0001**	0.1394	**< 0.0001**	-0.5085
Control	3	4-FOR	3	**< 0.0001**	0.1447	**< 0.0001**	-0.4017
1-FOR	1	4-FOR	1	0.9932	-0.0201	0.9992	0.0297
1-FOR	2	4-FOR	2	**< 0.0001**	0.1098	**< 0.0001**	-0.4267
1-FOR	3	4-FOR	3	**< 0.0001**	0.1467	**< 0.0001**	-0.2958

For the Control to Control comparisons shown in Table 1 the dispersion is significantly different at a p value < 0.0001 for time comparisons 1:2 and 2:3. At the same time the mean standard deviation is increasing across 1:2 and 2:3 time comparisons. The mean peer rating is also slightly decreasing in each time comparison and the mean peer rating pattern change is significant in 1:2 and 1:3 time comparisons. These changes in rating patterns in the Control sample could be due to several factors, e.g., better teamwork functioning just due to more familiarity with team members as the semester advances. However, all such external influences on rating patterns should have similar effects on our experimental samples.

For the 1-FOR sample there are significant differences in the dispersion matrices at times 1:2, and 2:3. The dispersion differences for 1-FOR potentially mimic the peer review rating behavior occurring in the control sample. The average standard deviation changes in 1-FOR 1:2, and 2:3 comparisons are increasing as in the Control sample. There are no significant changes in the mean value peer rating patterns in 1-FOR for time comparisons 1:2, and 2:3.

Comparing 4-FORs at times 1:2 and 2:3 we see significant changes at times 1:2 ($p < 0.0014$) in the dispersion matrices but no significant changes in times 2:3. This result is consistent with other FOR research which suggests the effect of FOR training is potentially mitigated over time [25]. Possibly this is also the result of other team behavior influences. Comparing 4-FORs at times 1:2 and 2:3 also see significant changes ($p < 0.0001$) in the mean ratings patterns times 1:2 and 2:3 although the changes move in opposite directions from 1:2 to 2:3.

When we compare the Control sample to the 1- FOR training experimental group we see no significant differences at peer review times 1: 2, or 1:3 in either the dispersion matrix comparisons or in the average peer rating matrix comparisons as compared to the Control sample. The average mean peer rating in 1-FOR compared to Control does increase from peer rating 1 to peer rating 3.

Comparing the dispersion in peer review distributions at time 1:1 for the Control sample to 1-FOR and comparing the two Experimental samples (1-FOR:4-FORs) there are no significant differences in the dispersion matrices or in the mean matrices. This result was expected because no FOR training had taken place until after the time 1 peer review.

However for Control sample to 1-FOR comparisons and comparing the two Experimental samples (1-FOR:4-FORs) there are significant differences at peer review times 2 and 3. Further examining these same comparisons, we see significant

dispersion differences ($p < 0.0001$) between Control and 4-FORs at peer review times 2 and 3 with decreasing average standard deviations.. There are also similar significant dispersion differences ($p < 0.0001$) between 1-FOR and 4-FORs at times 2 and 3. Similarly there are significant mean differences ($p < 0.0001$) in the mean ratings patterns between these two experimental groups at times 2 and 3. Comparing the Control and 1-FOR samples to the 4-FORs sample the average means also are significantly decreased ($p < 0.0001$).

IV. CONCLUSIONS

Our first conclusion was peer ratings in engineering students tend to move towards the mean over multiple peer ratings. Our second conclusion was that FOR training works if sufficiently reinforced and if it occurs relatively close to the actual peer review [25-27]. We found that our single session FOR training was not sufficient to significantly increase the dispersion in the peer ratings and mimicked the control group in their peer ratings. However, our 4-FOR experimental group significantly changed their rating dispersions and average mean ratings consistent with the goal of the experiments. There is also substantial longstanding and recent evidence in support of the effectiveness of FOR training [28-31]. Therefore we concluded that the use of FOR training will improve the dispersion of engineering student peer ratings and the quality of their peer evaluations.

This research is important because Improving the quality of peer ratings makes it more likely that engineering students will modify their teamwork behavior in ways deemed important by their peers and thereby improve their teamwork behavior, especially if the peer feedback is based upon a good understanding of positive team work behavior-which using FOR with CATME provides.

V. FURTHER RESEARCH

We are continuing to examine the changes in self-ratings, the ratings of other team members without self-ratings and the pairwise ratings of team members one to another for dispersion changes. We also will be examining the under or overconfidence comparisons of self to peer ratings for dispersion differences. We intend to repeat this experiment in subsequent semesters and increase the type, timing and frequency of the FOR interventions that we use (e.g. requesting student review of training materials outside of class time). We also expect to extend the experiments to capstone courses, other STEM disciplines and to student cohorts outside of engineering who already make extensive use of the CATME system.

In this experiment we were looking for increased dispersion in the ratings in order to improve the potential of constructive feedback received by students. When conducting ratings for selecting students for awards (e.g., entrepreneurship pitch or business plan competitions) or capstone projects, less dispersion, that is, increased interrater reliability, meaning less ratings dispersion, may be desired [32-34].

VI. LIMITATIONS

This experiment was done with large samples but all participants were first year engineering students. Different results may be obtained with senior capstone STEM students or other non-STEM majors who complete peer reviews. Frame of Reference training has been tested with many different types of students and in many contexts and proven effective in increasing the quality of ratings whether you are looking for more dispersion in peer reviews or less dispersion (inter-rater reliability). While one instructor delivered all the FOR training in this experiment, different instructors may experience dispersion or absolute magnitude of peer ratings change results which differ from our research findings.

ACKNOWLEDGMENT

We were significantly aided in this research by the instructors who allowed us time to present the FOR training in their FYE classes, especially the 4-FOR sample, and by the University's statistics faculty who guided our analysis work.

REFERENCES

[1] Alsop, R. in Wall Street Journal (New York, 2002).

[2] Pellerin, C. J. How NASA builds teams: Mission critical soft skills for scientists, engineers, and project teams. (John Wiley & Sons, 2009).

[3] Chen, J. C. & Chen, J. Testing a new approach for learning teamwork knowledge and skills in technical education. Journal of Industrial Technology 20, 37-46 (2004).

[4] ABET Inc. (ed Applied Science Accreditation Commission) (ABET, Inc., Baltimore, MD, 2009).

[5] AACSB International. (AACSB International, 2013).

[6] Elrick, L. (Rasmussen College, 2015).

[7] Calloway. A Report on Recruiters' Perceptions of Undergraduate Business Schools and Students. (School of Business and Accountancy of Wake Forest University, 2004).

[8] Dictionary.com. (Dictionary.com website, 2016).
http://www.dictionary.com/browse/teamwork

[9] Layton, R. A., Loughry, M. L., Ohland, M. W. & Ricco, G. D. Design and validation of a web-based system for assigning members to teams using instructor-specified criteria. Advances in Engineering Education 2 1-28 (2010).

[10] Ohland, M.W., Loughry, M.L., Woehr, D.J., Finelli, C.J., Bullard, L.G., Felder, R.M., Layton, R.A., Pomeranz, H.R., & 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.

[11] Loughry, M. L., Ohland M.W. & Moore, D. D. Development of a Theory-Based Assessment of Team Member Effectiveness. Educational and Psychological Measurement 67, 505-524 (2007).

[12] Thomas, E. J. Improving teamwork in healthcare: current approaches and the path forward. BMJ quality & safety (2011).

[13] Dictionary.comUnabridged.peerreview, <http://www.dictionary.com/browse/peer-review?s=t> (2016).

[14] Loughry, M. L., Ohland, M. L. & Woehr, D. J. Assessing teamwork skills for assurance of learning using CATME Team Tools. Journal of Marketing Education 36, 5-19 (2014).

[15] Ohland, M. W., Layton, R. A., Loughry, M. L. & Yuhasz, A. G. Effects of behavioral anchors on peer evaluation reliability. Journal of Engineering Education 94, 319-326 (2005).

[16] Dictionary.com Unabridged. Vol., (Dictionary.com website, 2016).

[17] Poling, T., Woehr, D. J., Arciniega, L. M. & Gorman, A. in Annual Meeting for the Society for Industrial and Organizational Psychology (2005).

[18] Resick, C. J., Dickson, M. W., Mitchelson, J. K., Allison, L. K. & Clark, M. A. Team composition, cognition, and effectiveness: Examining mental model similarity and accuracy. Group Dynamics: Theory, Research, and Practice 14, 174-191 (2010).

[19] in Psychology 104: Social Psychology (Study.com, 2015).

[20] Woehr, D. J. & Huffcutt, A. I. Rater training for performance appraisal: A quantitative review. Journal of Occupational and Organizational Psychology 67, 189-205 (1994).

[21] Johnson, R., Penny, J. & Gordon, B. Assessing performance: Developing, scoring, and validating performance tasks. (Guilford, 2009).

[22] Statistics Solutions. <http://www.statisticssolutions.com/conduct-interpret-repeated-measures-anova/> (2016).

[23] Miller Jr, R. G. Beyond ANOVA: basics of applied statistics. (CRC Press, 1997).

[24] Freund, R.J., Littell, R.C., and Spector, P.C. (1986), *SAS System for Linear Models, 1986 Edition*, Cary, NC: SAS Institute Inc.

[25] Roch, S. G. & O'Sullivan, B. J. Frame of reference rater training issues: recall, time and behavior observation training. International Journal of Training and Development 7, 93-107 (2003).

[26] Sulsky, L. M. & Day, D. V. Frame-of-reference training and cognitive categorization: An empirical investigation of rater memory issues. Journal of Applied Psychology 77, 501 (1992).

[27] Sulsky, L. M. & Day, D. V. Effects of frame-of-reference training on rater accuracy under alternative time delays. Journal of Applied Psychology 79, 535 (1994).

[28] Roch, S. G., Woehr, D. J., Mishra, V. & Urszula, K. The importance of frame of reference training: An updated meta-analysis. Journal of Organizational and Occupational Psychology 85, 370-395 (2012).

[29] Woehr, D. J. Understanding frame-of-reference training: The impact of training on the recall of performance information. Journal of Applied Psychology 79, 525-534 (1994).

[30] Fehrmann, M. L., Woehr, D. J. & Arthur, W. J. The Angoff cutoff score method: The impact of frame-of-reference rater training. Educational and Psychological Measurement 51, 857-872 (1991).

[31] Dierdorff, E. C., Surface, E. A. & Brown, K. G. Frame-of-reference training effectiveness: Effects of goal orientation and self-efficacy on affective, cognitive, skill-based, and transfer outcomes. Journal of Applied Psychology 95, 1181-1191 (2010).

[32] Tinsley, H. E. & Weiss, D. J. Interrater reliability and agreement of subjective judgments. Journal of Counseling Psychology 22, 358 (1975).

[33] Shi, H., Ferguson D.M., Beagley J. & Huyck M. ,."Improving inter-rater reliability used to measure learning outcomes." Proceedings of 38th IEEE/ASEE Frontiers in Education Conference Vol. paper 1284 2 (Saratoga Springs, NY, 2008).

[34] Ferguson Daniel M., M. Govekar, and A. Stype, "Quality and Consistency in Idea Pitch, Research Proposal and Business Plan Competition Judging," *Proceedings, ASEE Annual Conference*, paper 1665, 17 pages (Louisville, KY, 2010).