How Competent are Freshman Engineering Students in Constructively Rating Their Peers in a Team Context?

Dr. Benjamin Emery Mertz, Arizona State University

Dr. Benjamin Mertz received his Ph. D. in Aerospace Engineering from the University of Notre Dame in 2010 and B.S. in Mechanical Engineering from Rose-Hulman Institute of Technology in 2005. He is currently a part of a lecturer team at Arizona State University that focuses on the first-year engineering experience, including developing and teaching the Introduction to Engineering course. He also teaches Thermo-Fluids and High Speed Aerodynamics for the Mechanical and Aerospace Engineering Department at ASU. His interests include student pathways and motivations into engineering and developing lab-based curriculum. Recently, he has developed an interest in non-traditional modes of content delivery including online classes and flipped classrooms.

Dr. Daniel M. Ferguson, Purdue University, West Lafayette

Daniel M. Ferguson is CATME Managing Director and a research associate at Purdue University. Prior to coming to Purdue he was Assistant Professor of Entrepreneurship at Ohio Northern University. Before assuming that position he was Associate Director of the Inter-Professional Studies Program [IPRO] and Senior Lecturer at Illinois Institute of Technology and involved in research in service learning, assessment processes and interventions aimed at improving learning objective attainment. Prior to his University assignments he was the Founder and CEO of The EDI Group, Ltd. and The EDI Group Canada, Ltd, independent professional services companies specializing in B2B electronic commerce and electronic data interchange. The EDI Group companies conducted syndicated market research, offered educational seminars and conferences and published The Journal of Electronic Commerce. He was also a Vice President at the First National Bank of Chicago [now J.P. Morgan Chase], where he founded and managed the bank's market leading professional Cash Management Consulting Group, initiated the bank's non-credit service product management organization and profit center profitability programs and was instrumental in the breakthrough EDI/EFT payment system implemented by General Motors. Dr. Ferguson is a graduate of Notre Dame, Stanford and Purdue Universities, a special edition editor of the Journal of Engineering Entrepreneurship and a member of Tau Beta Pi.

Mohd Iramul Hoque, CATME

How competent are freshman engineering students in constructively rating their peers in a team context?

Abstract

In this evidence-based practice paper, we investigated how well freshman engineering students are able to constructively rate their peers in a team context and evaluated interventions that may help them improve their rating abilities. Teamwork is an important professional skill and can have significant benefits during a student's academic career, and peer feedback is one mechanism to help students improve their teamwork performance. A web-based tool called CATME (Comprehensive Assessment of Team Member Effectiveness) was used to measure the effect of two different interventions on students' ability to provide good feedback on teamwork performance to their peers: Frame-of-Reference training (FOR) and Rater Error Training (RET). This paper is meant to address how much freshman engineering students' peer evaluation ratings of themselves and their teammates vary with and without these interventions. The effect of the interventions was analyzed using statistical analysis to identify differences in rating patterns between the three peer evaluations. The third peer evaluation was used to evaluate the longevity of the intervention effect. The results showed that the interventions did improve how students rate each other by decreasing rater variance between each peer evaluations and increasing the target variance, meaning that after the interventions the students were better able to differentiate between teammates and between the different CATME teamwork performance dimensions. The effects of the interventions did seem to persist through the third peer evaluation where no additional intervention was applied.

Introduction

Teamwork is a skill that has been integrated into many different courses in a variety of different academic disciplines. This is largely due to the importance that companies place on teamwork skills in potential new employees [1-2] as well as the incorporation of teamwork requirements for accreditation across a variety of disciplines [2-5]. Studies have also shown learning and other benefits of working in teams while in college [6-7]. However, Chen argues that many students lack these important teamwork skills when they enter the workforce [8]. Since about half of U.S. companies, [9] and 81% of Fortune 500 companies [10] have company rely heavily on team-based structures, students who are deficient in their teamwork skills are ill-equipped to function within these companies upon graduation.

Teamwork is defined as "cooperative or coordinated effort on the part of a group of persons acting together" [11]. One way to improve teamwork is to provide opportunities for behavioral peer review that is defined as an evaluation of an individual's contribution to a work activity by their peers [12]. Peer reviews of teamwork performance provide a valuable professional skill of

evaluating other's performance, which also helps teach students how to act within a team through a self-evaluation process [13]. Peer reviews have also been shown to improve learning outcomes in upper level education, which promotes more constructive team behaviors in future team experiences [14].

A web-based tool called CATME (Comprehensive Assessment of Team Member Effectiveness) is a tool for academic teamwork environments and was used in this paper to assist students in giving peer feedback to their team members [15]. CATME is constructed around five behavioral dimensions: Having relevant Knowledge, Skills, and Attributes (KSA's), Contributing to the Team's Work, Interacting with Teammates, Keeping the Team on Track and Expecting Quality [16-17]. 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 belief that the team can do a good job and encouraging the team to do its best.

Since each of these five dimensions is equally important to the success of the team, these form the basis for the peer reviews [16].

This paper is meant to address how much freshman engineering students' peer evaluation ratings of themselves and their teammates vary. We then want to evaluate how specific interventions change the rating patterns of students in their peer evaluations. This work builds on the work of Ferguson et. al who employed a Frame-of-Reference (FOR) Training to their first-year engineering class and saw improvements in rating dispersion and potentially the quality of the feedback given [18]. They identified two significant and recurring problems in the peer reviews of the students in their study:

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

The quality or accuracy of the ratings is quantified by evaluating the diversity of the rating across team members and across the five CATME dimensions (with more diversity being better). While it is possible that team members legitimately deserve similar ratings, this is highly unlikely [19-20] and is more probably explained by documented peer rating behaviors like the Halo Effect [21]. While Ferguson et. al did show improvement in the amount of dispersion in the peer ratings of their students, this was limited to a single institution and a single type of intervention. In this paper, two different kinds of interventions (Frame-of-Reference Training and Rater Error Training) were implemented at different points within four sections of the Introduction to Engineering course at a large public research institution. Four sections taught by a different instructor who used CATME for peer evaluations but did not implement the interventions were used as a control group. The peer reviews were analyzed and the dispersion in the ratings was calculated for each of the three peer reviews. One of the interventions preceded each peer review. The goal of this study was to:

1. Demonstrate improved peer ratings using the FOR training in a different context
2. Determine the extent of the effectiveness of the new interventions in improving the feedback students give each other
3. Examine the cumulative effects of using multiple interventions within a course and the longevity of these effects (if any).

In this paper, we will discuss our research population, the different interventions implemented, experimental procedures, analysis structure and processes, findings, and conclusions.

Research Methods

Research Population

The interventions were implemented in 4 of 32 sections of the Introduction to Engineering class at a major southwestern, public, tier 1 research institution during the Fall 2017 semester. These sections include Aerospace, Mechanical, Chemical, and Electrical Engineering students. Each section consists of approximately 40 students. While there were approximately 1,300 first year engineering students in these sections, at this particular institution, only approximately 150 students were in the treatment sections. The students in the treatment group were students that were enrolled in the sections taught by one of the investigators. An additional section (approximately 40 students), which was taught by a different instructor (not one of the investigators), was used as a control group in which no rater training was given. The demographics of the population were not included in the analysis and so no direct comparison can be made between the populations of these different sections. Due to withdraws from the course or non-participation of some students, some teams had to be removed from the analysis. This will be discussed in more detail in the Findings and Limitations sections of this paper.

Training

The FOR training was given to students during the first week of class. This was done via a set of PowerPoint slides describing the five CATME dimensions and showing examples of the rating scales used to evaluate these dimensions ranging from well below expectations (= 1) to well above expectations (= 5). Examples of behaviors that would warrant each of the rating levels were described to the students and an in-class activity was done with the students in which word descriptions of behaviors were given to students and they were asked to rate the behaviors on the five CATME dimensions scales. The FOR training was used for approximately 25 minutes of the first lab time (which is 2 hours and 50 minutes long) with other team activities and lecture topics discussed in the remainder of the class.

The Rater Error Training was done before the second peer evaluation (near the midway point of the course). This consisted of a set of PowerPoint slides that were discussed in lecture. The slides included descriptions of common rating problems including giving everyone on the team the same scores across all dimensions, giving the same teammate the same scores across all dimensions, bimodal ratings (giving one teammate all 1's and others all 5's), etc. This lecture also included a discussion of what information you are trying to give your teammates when rating them and how the results of the evaluations can be interpreted in order to improve team performance. General comments were also made regarding what the rating patterns looked like in the class without identifying individuals or teams that used poor rating patterns. The goal was to help students reflect on their own ratings and encourage them to improve the feedback they were giving to their teammates.

Experimental Procedure

Students were assigned into four-person teams during the first week of the Introduction to Engineering class. The CATME Team Formation survey was used to match people based on a variety of different criteria including availability, major, skills, etc. The students stayed in these teams for the entire semester as they worked through the design process on a semester-long project. During the project various team deliverables were completed and the peer evaluations were placed after some of the more intensive assignments where they would have to work together closely with their teammates.

As stated before, the FOR training was done during the first week of classes. The first peer evaluation was then given during the fifth week of the class after they had gone through the process of identifying the problem based on customer requirements. They were also reminded of the FOR training when this first peer evaluation was assigned.

The Rater Error Training was given during the 50 minute class time dedicated towards "Addressing Negative Feedback". This lecture topic is meant to help students synthesize the feedback they were given from their customers during their proposal and help them iterate on their design. The proposal consisted of both a team oral presentation and report and constituted the largest assignments up to this point in the class and required the most team collaboration to complete as well. These assignments occur at about the halfway point of the class. The Rater Error portion of this lecture again takes about 25 minutes of the class and the peer evaluations are assigned to the students at the end of that lecture.

The final peer evaluation occurred at the end of the course. No further instruction on giving peer evaluations was given prior to the evaluation. The purpose of this evaluation, for the sake of this study, is to determine if students would revert to undesirable rating practices if no intervention was administered.

To encourage participation in these activities, points were awarded to students who completed the surveys. The total amount of points a student could earn for doing the evaluations amounted to 2% of the final grade.

The control group did the same three peer evaluations at similar times within the course. While the students in the control group did receive instruction about teamwork in general, they were not trained specifically on the CATME dimensions. The extent of the training in the control sections was limited only to the first week or two of the course.

The CATME peer reviews were completed online and the results of the reviews are given back to the students in an online form. The online feedback consists of pointers to word descriptions of behaviors similar to their own behaviors on each of the five CATME dimensions. The average ratings given to them by their teammates, along with the average team rating and their self-rating are included on the report for each dimension. No numbers are given as a part of the results in order to help maintain confidentiality of the raters. Numerical data is provided to the instructors along with written comments so that the instructor can address specific concerns with the teams offline. An example of the results the students see is shown in Figure 1. An example of the numerical data that is provided to the instructor for a single team is shown in Figure 2.

Contributing to the Team's Work

How You Rated Yourself	How Your Teammates Rated You	Average Rating for You and Your Team	Description of Rating
			• Does more or higher-quality work than expected. • Makes important contributions that improve the team's work. • Helps teammates who are having difficulty completing their work.
►		►	Demonstrates behaviors described immediately above and below.
	►		• Completes a fair share of the team's work with acceptable quality. • Keeps commitments and completes assignments on time. • Helps teammates who are having difficulty when it is easy or important.
			Demonstrates behaviors described immediately above and below.
			• Does not do a fair share of the team's work. Delivers sloppy or incomplete work. • Misses deadlines. Is late, unprepared, or absent for team meetings. • Does not assist teammates. Quits if the work becomes difficult.

Research suggests the following behaviors will improve your ratings in this area:

- Do a fair share of the team's work.
- Fulfill your responsibilities to the team.
- Come to team meetings prepared.
- Complete your work in a timely manner.
- Do work that is complete and accurate.
- Make important contributions to the team's final product.
- Keep trying when faced with difficult situations.
- Offer to help teammates when it is appropriate.

Figure 1: Student Feedback for Contributing to the Team's Work Dimension.

		Rater 1					Rater 2					Rater 3				
Team ID	Rater #	C	I	K	E	H	C	I	K	E	H	C	I	K	E	H
01	1	4	4	4	4	4	4	3	4	4	4	4	4	3	4	3
01	2	5	4	3	4	5	4	3	4	4	4	4	3	4	4	3
01	3	2	4	3	3	4	3	3	3	3	4	2	4	3	4	4

Figure 2: Sample Instructor Raw Data Display C = Contributing, I = Interacting, K = Keeping, E = Expecting, H = Having

The numerical results from the ratings in the three evaluations were downloaded from the CATME program and all identifiable information was removed for analysis. It was also at this point that those who did not complete the evaluations (whether due to withdrawing from the course or non-participation) were removed from the data sets. This does limit how much information could be included in the analysis and does limit the extent of the conclusions that can be drawn from the results. This will be discussed further in the limitations section of this paper.

Analysis Structure

Mean ratings, and standard deviations were calculated for each student's self and teammates' ratings across the CATME dimensions and for each CATME dimension. Dispersion factors were then calculated according to Ferguson et al. [18].

Three different analysis models were employed to understand the ratings. A convergence analysis was performed to see how much the self-ratings of students corresponded to the ratings their teammates gave them. It is expected that as the students get better at rating themselves and others that the difference in the mean scores and their standard deviations would both decrease.

The second analysis model was a repeated measure ANOVA [22-23] analysis that was used to determine if the changes in mean and standard deviation were significantly different for each intervention as compared to the other interventions and the control sections.

The third analysis model is a Social Relations Model (SRM) [24] that is used to determine how rater variance and target variance changes with time. The key idea in SRM is that individual ratings can be divided into key variance components. These variance components are believed to act together to produce the rating observed in the data collected. The distinct variances are:

- the rater variance (person doing the rating)
- the target variance (person being rated)
- dyadic variance (this captures the relationship/interaction between rater and target)
- the team variance (the team in which the rater and target belong)

However, in the analysis for this paper we do not include the team variance.

The rater effect/variance measures the tendency of an individual rater to rate his/her teammates consistently. A larger rater effect points to the fact that, on average, raters tends to rate their entire team members the same way. In our case, a large rater variance is not desired in the intervention groups because it would indicate less variability in ratings. The target effect/variance measures the tendency of all team members to rate the same individual consistently. Hence, a higher target effect means that all their team members rated individuals similarly. A higher target variance with respect to the other variances shows rating that is more reliable. The dyadic (relationship) effect/variance captures the variability in a person's ratings depending on the team member being rated. A large dyadic effect means that raters do discriminate between team members when rating and do not give all teammates the same rating. This is a good thing as that means they are treating each teammate as an individual and not just giving ratings to individuals based on an averaged team performance.

Findings

Sample results of the convergence analysis can be seen in Table 1. In this table, we are looking at the differences between how students rated themselves and how others rated them. In the control group, there was a significant difference between self and peer mean ratings in the first peer review, but not in the last peer review. This means that there was convergence in the student's abilities to rate each other that naturally occurs without intervention throughout the semester. Convergence was also seen in the intervention group between the first and the second peer evaluations. The third evaluation showed convergence compared to the first evaluation, but there was not a significant change between the second and the third evaluation. This is encouraging since this means that the observed improvements are maintained even after the interventions are applied without additional interventions. It also means that without interventions, the students did not significantly improve their ratings (from the second to the third rating).

Table 1: Convergence Analysis Sample Data

Intervention Sample	Peer Review Time	Difference self to peer mean ratings	P-Value
Control	1	-0.6500	0.0078
Control	3	0.05444	0.8204
Experimental	1	-0.5711	0.0006
Experimental	2	-0.00303	0.9903
Experimental	3	-0.04428	0.7826

Sample results from the repeated measure ANOVA analysis can be seen in Table 2. This analysis shows that the dispersion values decrease significantly from peer evaluation 1 to peer evaluation 3 in both the control and in the experimental groups. This is consistent with the interpretation from Table 1. However, it can also be seen that the mean rating in the experimental groups was significantly higher ($p = 0.0481$) than the mean in the control group. This may indicate that the experimental group had a different perception on how to use the rating scale which may indicate the effectiveness of the FOR training.

Table 2: Repeated Measure ANOVA Analysis Sample Data

Intervention Sample	Peer Review Time	Intervention Sample	Peer Review Time	STDev Difference Dispersion	P-Value Difference Dispersion	Difference Mean Rating	P-Value Difference Mean
Control	3	Control	1	-0.5292	<0.0001	0.1978	0.1377
Experimental	3	Experimental	1	-0.5196	0.0045	0.4233	0.0346
Experimental	1	Control	1	-0.1284	0.7996	0.1596	0.6887
Experimental	3	Control	3	-0.1188	0.8718	0.3852	0.0481

The results from the SRM analysis are shown in Table 3 for the experimental group and Table 4 for the control group. Due to withdraws or non-participating teammates the number of teams used in the first peer evaluation was 18 comprising of 72 individuals, the second evaluation only had 16 complete teams comprising of 64 individuals, and the third analysis included 13 teams made up of 52 individuals. The coefficients in the tables are unstandardized. The results do show that as time progressed, the rater variance declined, that is the students gave more consistent evaluations after the interventions (from 58% to 38%). It was also seen that the number of instances where a student rated everyone on their team the same across all five dimensions decreased. The target variance also increased from 15% to 30% that shows that the students did a better job differentiating across the different CATME dimensions. While promising, because so many teams had to be removed from the analysis, more work needs to be done to show that this data is representative of the entire class.

Table 3: Differences in Variance Components for Peer Ratings Across Interventions for Intervention Groups, % = Percent of Variance Component

Variance Component	*Internal Review 1*			*Internal Review 2*			*Internal Review 3*		
	Estimate	SE	%	Estimate	SE	%	Estimate	SE	%
Rater	0.255	0.056	**57.9**	0.255	0.066	**49.5**	0.276	0.078	**38.1**
Target	0.065	0.026	**14.7**	0.104	0.040	**20.2**	0.218	0.079	**30.1**
Relationship	0.120	0.021	**27.4**	0.156	0.018	**30.3**	0.231	0.041	**31.9**
Actor effect reliability	0.848			0.820			0.782		
Target effect reliability	0.586			0.650			0.739		

In the control group, after all of the teams were removed that had missing data due to student withdraws or students not completing the evaluation, only 5 teams consisting of 20 students were left for this analysis. This may be too few teams to accurately understand rating behavior and is something that should be addressed in future work before more conclusions can be drawn. This analysis did show an increase in the target variance and a decrease in the rater variance similar to the experimental group, but no direct comparison can be made until a larger control sample is analyzed.

Limitations

There are some important limitations of this study that must be addressed. We are currently working to obtain and analyze a larger control group to better understand the differences between the control group and the experimental group. In addition, this particular control group only did two peer evaluations that we had access to rather than the three evaluations performed in the experimental section. This was not the plan during the planning stages of this study, but this

ended up being the only control data available during analysis. As mentioned before, even though the data sets were larger for the experimental group, there were still many teams that needed to be removed for the SRM analysis. This means more work does need to be done to ensure that this sample population represents the overall class population, although there have been no identifiable differences between the teams included in the study and those that have not at this point. Finally, the extremely small sample population for the control group makes the results of the SRM analysis very questionable. A larger control sample needs to be obtained before any conclusions can be drawn between experimental and control groups based on the SRM analysis.

Table 4: Differences in Variance Components for Peer Ratings Across Interventions for Control Groups, % = Percent of Variance Component

Variance Component	*Control Review 1*			*Control Review 3*		
	Estimate	SE	%	Estimate	SE	%
Rater	0.250	0.080	**48.4**	0.183	0.108	**21.6**
Target	0.200	0.128	**38.7**	0.433	0.238	**51.0**
Relationship	0.067	0.025	**12.9**	0.233	0.071	**27.5**
Actor effect reliability	0.909			0.662		
Target effect reliability	0.889			0.822		

Conclusion

Based on the analysis of first-year students' peer and self-ratings with and without FOR and Rater Error Training (RET), the interventions do seem to have a positive effect on the way that students rate their peers. The effects do improve over time indicating that the cumulative effect of the interventions does seem to improve the ratings and that the students are able to rate more effectively even without reminders once the trainings have occurred. This analysis further demonstrates the effectiveness of these training in different contexts by showing that similar effects can be demonstrated across universities [18] (those in current literature and the one in this study). While more work needs to be done to make definitive claims of the benefits of these trainings over the control group, FOR and RET seem to be one way of effectively helping students improve the way they rate one another.

References

[1] Alsop, R., (2002). "Playing well with others". The Wall Street Journal, September 9, 2002, R11.

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

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

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

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

[6] Oakley, B.A., Hanna, D.M., Kuzymn, Z., Felder, R.M. (2007). "Best practices involving teamwork in the classroom: Results from a survey of 6435 engineering student respondents". IEEE Trans. Ed. 50(3): 266-272 Aug 2007.

[7] Smith, K.A., Sheppard, S.D., Johnson, D.W., Johnson, R.T. (2005). "Pedagogies of engagement: Classroom-based practices". J.Eng. Ed. 94(1): 87-101 Jan 2005.

[8] 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).

[9] Devine, D.J., Clayton, L.D., Philips, J.L., Dunford, B.B., & Melner, S.B. (1999). "Teams in organizations: Prevalence, characteristics, and effectiveness". Small Group Research, 30(6), 678-711.

[10] Lawler, E.E., Mohrman, S.A., Benson, G. (2001). "Organizing for high performance: Employee Involvement", TQM, Reengineering, and Knowledge Management in the Fortune 1000. San Francisco, CA: Jossey-Bass.

[11] Dictionary.com. (Dictionary.com website). [Online]. Available: http://www.dictionary.com/browse/teamwork [Accessed Jan. 1, 2018]

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

[13] Thomas, G., Martin, D., & Pleasants, K. (2011). "Using self-and peer-assessment to enhance students' future-learning in higher education". Journal of University Teaching & Learning Practice, 8(1): 5

[14] Brutus, S., & Donia, M. B. (2010). "Improving the effectiveness of students in groups with a

centralized peer evaluation system". Academy of Management Learning & Education, 9(4): 652-662.

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

[16] 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.

[17] 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, 505-524.

[18] Ferguson. D. M., Lally, C., Ibriga, H., Murch, O., & Ohland, M. (2016). "Using frame-of-reference training to improve the dispersion of peer ratings in teams". 1-6. 10.1109/FIE.2016.7757706.

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

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

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

[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] Christensen, P. N., & Kashy, D. A. (2012). "Using the social relations model to understand interpersonal perception and behavior". In H. Cooper (Ed.), APA Handbook of Research Methods in Psychology, Vol. 3: 425-437. Washington, D.C.: APA.