35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Evaluation of FCS Self and Peer-Assessment Approach Based on Cooperative and Engineering Design Learning

Dean Cvetkovic *(Member, IEEE)*

Abstract— **The Cooperative Learning in Engineering Design curriculum can be enhanced with structured and timely self and peer assessment teaching methodologies which can easily be applied to any Biomedical Engineering curriculum. A study was designed and implemented to evaluate the effectiveness of this structured and timely self and peer assessment on student team-based projects. In comparing the 'peer-blind' and 'face-to-face' Fair Contribution Scoring (FCS) methods, both had advantages and disadvantages. The 'peer-blind' self and peer assessment method would cause high discrepancy between self and team ratings. But the 'face-to-face' method on the other hand did not have the discrepancy issue and had actually proved to be a more accurate and effective, indicating team cohesiveness and good cooperative learning.**

I. Introduction

The largest problem that is often encountered in Cooperative Learning in general and Biomedical Engineering Design curriculum is that students that may provide little or no contribution when working in teams. Students who fail to contribute to project activities continue to receive high team marks. This would often cause resentment with hard working students towards their peers and affect the team dynamics in the long term. To address this problem, a self and peer review scheme is often introduced [1-8]. Self and peer assessment both give some control and responsibility back to the student, emphasising an increased sense of autonomy in the learner [1,9]. Peer assessment method has been undertaken to assess team member's performance via online (secure and anonymous environment) as opposed to team member face-to-face decision of a proposed fair contribution [6]. The peer review methods implemented around the experimental work compared the two peer assessments [4]. Previously, Self and Peer Assessment Resource Kit (SPARK) has been utilised as a confidential online system for self and peer assessment [8,9]. This system allows students to rate their own and each other's contributions to different aspects of a team task, using multiple assessment criteria. The student ratings are then combined to produce a self and peer assessment factor used to moderate the team mark for each individual team member. 'Autorating System' was another effective peer rating system, developed by ex-colleague Dr. Robert Brown [1]. This method and tool was applied at author's school in the early 2000, for the similar Engineering Design 3 and 4 curriculums.

*Research supported by RMIT STeLR grant (2011).
RMIT University, School of Electrical and Computer Engineering, GPO Box 2476V, Melbourne, VIC 3001, Australia.
Corresponding author name: Dr. Dean Cvetkovic;
Email: dean.cvetkovic@rmit.edu.au

II. Methodology

A. *Study Scope and Study Design*

The team projects which are the part of this Engineering Design curriculum encourages students to work in teams to design and develop their engineering product, as well as present and document that work while performing timely self and peer assessment. This study includes the implementation and embedding of new innovation (i.e. Learning and Teaching activities) within the Engineering Design 3 course curriculum while simultaneously evaluating it. The sample size initially calculated for this study during the ethics application process was 53 students with most ideal case of 100+. While in 2011, 210 students were enrolled in the course, the study data was received from all 38 teams and 210 students. RMIT Ethics approval was granted for the collection all data and all human subjects signed the consent letters.

An original 'Fair Contribution Scoring' (FCS) method was applied in this particular study. This method was initially implemented in RMIT University (Australia) in 2009, and adopted from the University of Wollongong (Australia). Since then it has been enhanced, identified as a Teaching and Learning innovation with the potential to be transformed into a commercial online tool. The FCS method was re-designed from a simple formula (used in 2009 and 2010) to a formula embedded in an Excel spreadsheet which students can use to enter their self and team ratings based on their contribution to their project. This template would automatically calculate their individual mark for their report. Our spreadsheet FCS method consists of a number of stages. The first stage consists of a self and team mark (out of 100%) that can be assigned to each student member, decided from their weekly or fortnightly work contribution to their team project.

A typical team size in 2011 was 6-7 members as compared to 7-8 in 2009 and 5-6 in 2010. For example, for a 7-member team, there would be a pool of 140 points to distribute amongst them. All members start with 20 points, but if someone has not contributed much, they may end up with 14 points, which means there are 6 points (20-14=6) to add onto other members who have worked harder (eg. FCS may look like for members 1-7: 22, 21, 20, 20, 14.4, 21, 21.6, etc).

The second stage of our FCS method, consists of FCS and 'Adjusted FCS' (AFCS) points being automatically calculated from those team ratings (first stage) which are out of 100%. The AFCS limits from over-contribution. For example, in a normal condition (i.e. if all members contributed to their project), if someone ends up with 26 or 27 FCS, it would be reduced to a maximum of 25.5. Whereas, 0 FCS entry can be assigned as a minimum FCS if a member has not contributed at all. In this case it would allow other team member(s) to increase the limit to 28 FCS.

These conditions and formulas have been embedded in the Excel spreadsheets as shown in Table I. The actual table in this Table I extends to week 12, where the AFCS would be averaged over each semester. In the first half of semester 1, this self and peer assessment was conducted weekly. From the second half of semester 1 (week 7) and throughout semester 2, the FCS was conducted every 2 weeks (fortnightly). This AFCS would be used as a multiplier for the mid-year and final Report deliverable. For example, the average 23 AFCS would be translated as 23/20 (starting points) x 75% (team report mark) = 1.15 x 75% = 86% (individual mark for report deliverable).

The interest in this FCS method stems from a need to evaluate the self and peer assessment by computing the Standard Deviation (S.D.) of AFCS for each week between all members. For example, the smaller the S.D., the better cooperation and vice versa - the bigger S.D., cooperation is weaker. This is just one of the evaluations of interests in this study. Therefore we designed the study for this activity, by creating dual Control and Experimental groups, as shown in Table II and III.

TABLE I. SELF AND PEER ASSESSMENT ACTIVITY STUDY DESIGN FOR SEMESTER 1 AND 2.

Semester 1 & 2, 2011		**Self and Peer Assessment**	**Team SET No.**
Semester 1	Week 1-12	Experimental (1 & 2) *	SET 1 & 2
Semester 2	Week 1-6	Control (1) & Control (2) Experimental (1) & Control (2)	SET 1 & 2
	Week 7-12	Control (1) & Experimental (2) Experimental (1) & Experimental (2) *	SET 2 & 1

TABLE II. SELF AND PEER ASSESSMENT ACTIVITY STUDY DESIGN FOR SEMESTER 2.

Semester 2	**Control (1)** (Logbook method)	**Experimental (1)*** (FCS method)	
Week 1-6	SET 1	SET 2	**Control (2)** (peer-blind assessment using Google doc)
Week 7-12	SET 2	SET 1	**Experimental (2)*** (face-to-face peer assessment)

The Control (1) group needed to conduct their individual self rating (out of 100%), based on their weekly/fortnightly project contribution, documented in their timely Logbook Wiki. In week 1-6 of semester 2, Control (1) SET 1 group needed to rate their own self logbook entries (refer to Table III). This rating was conducted using individual and peer-blind submission process using timely 'Google doc' tool. These self ratings were processed and averaged along with their other team member ratings, determining the team rating out of 100%, then be manually entered in the FCS spreadsheet by the investigator in order to calculate the fortnightly AFCS. In semester 1, timely peer-to-peer ratings were made 'transparent' to own and paired team members. In semester 2 the Google doc technology enabled to conduct a peer-blind rating process, as part of Control (2) group (refer to Table III). The Experimental (1) SET 2 group in week 1-6 of semester 2 also required members to submit their self rating out of 100% based on FCS work contribution in that week (not logbook rating). Similarly, these ratings were processed with Google doc, averaged and manually calculated the fortnightly AFCS. From semester 2 week 7-12, the Experimental (2) SET 1 group assessment was conducted in face-to-face environment where members met to complete their own fortnightly team ratings and compute FCS. SET 2 group was slightly different, as they needed to enter their self ratings based on logbook marking criteria and items. The team rating was calculated from investigator's rating of individual's logbook entry in that two-week period. As referred in Table II, throughout semester 1 week 1-12, the face-to-face FCS method was utilised for both SET groups to conduct their self and peer assessment (same as SET 1 group in semester 2 week 7-12). The study design for this self and peer assessment required pairing of Control and Experimental groups.

III. RESULTS

A. *Weekly/Fortnightly Distribution of AFCS*

Firstly, an evaluation of the dynamics of semester 1 weekly team FCS and adjusted FCS (using max, min, S.D. and N values) was conducted. The S.D. values of FCS and adjusted FCS were the main results indicating that the smaller the S.D. FCS, the more cohesive team contribution to their project. Refer to Fig. 1 a) for AFCS results in semester 1. The results indicated that the S.D. was slighty lower in week 1-6 than week 7-12. In week 7-12, the AFCSs' S.D. was constant from start to end of semester 1. Week 1-6 AFCS was 'cyclic' or fluctuating for S.D., min and max data. In interprating these results, the cohesive team contribution was different in the first half than the second half of semester 1. Due to pressure for teams to complete their deliverables to best possible quality and deadlines, the team cohesiveness decreased in the second half of the semester 1. The lowest S.D. was evident week 4-6, at an early project stage. For semester 2 (refer to Fig. 1 b)), the AFCS was lowest in week 1-2 and similar in week 7-10. Whereas, the highest S.D. was observed in week 3-6 and 11-12. Again, the high S.D. in the final weeks of semester 1 only may have showen that pressure to complete their project deliverables. One would argue that this is also part of normal cooparative learning and peer assessment. In comparing semester 1 and 2 S.D. results, it is clear that in

semester 1 the S.D. was around 2.5, whereas in semester 2 it was around 5. It was obvious that team cohesiveness in its peer assessment was lower in semester 2 but there could other reasons for it – the peer-blind self and peer assessment would have contributed for members to rate each other within a larger rating range than normally done when meeting face-to-face.

B. *Comparison of Weekly/Fortnightly AFCS Standard Deviations*

Another statistical analysis was conducted to compare the statistical difference between semester 1 and 2 weekly/fortnightly AFCS and 'AFCS total average' (S.D.) parameters. Correlations between the two semesters show significance for AFCS in last 4 weeks, as well as its AFCS total average of all weeks. These results revealed the following: week 9-10 (*corr.* = +0.482, p = 0.004), week 11-12 (*corr.* = +0.638, p = 0.001) and total average (*corr.* = +0.329, p = 0.044). This higher correlations suggests that the higher the AFCS S.D. in semester 1, the higher the AFCS S.D. in semester 2, in the last 4 weeks. Clearly, this indicates that the high S.D. in the final weeks of semester 1 and 2 show student pressure to complete their project and deliverables and that they strive to assess themselves based on their actual and fair contribution. One would argue that this is also part of normal cooparative learning and peer assessment. Multiple paired-sample t-tests were performed to compute the significant difference between AFCS S.D.s at the two semesters and it revealed that there was a significant difference (p<0.05) between all fortnights and total averages. In semester 2, in order to compute AFCS for both Control and Experimental groups, consisting of SET 1 and 2 groups, students needed to enter their 'self' and 'peer' ratings, fortnightly. The 'team' rating would generally be derived from average peer ratings during the peer-blind assessment process in weeks 1-6. Whereas, in week 7-12, these team ratings would be derived from their own rating (SET 2) and staff rating (SET 1) as part of face-to-face assessment process.

C. *Comparison of Weekly/Fortnightly Self and Team Ratings*

Another statistical analysis was conducted to compare the statistical difference between fortnightly 'self rating' and 'team rating' for all teams (teams 1-38). Fig. 2 a) shows the SET 1 results which indicate the significant difference between the self and team ratings in weeks 1-6, and very similar ratings in week 7-12. Multiple correlation and paired sample t-tests were conducted and results show that there is a high and significant correlation (*corr.* > 0.9, p = 0.001) in weeks 7-12 which show no difference between self and team ratings. Whereas, the t-tests show the significant difference (p = 0.001) only in week 1-6. This SET 1 belongs to Control (1 & 2) week 1-6 and Experimental (1 & 2) groups for week 7-12. The reason for such great discrepancy between self and team ratings in week 1-6 was due to *peer-blind* assessment but also due to student perception that their logbook entries are of higher mark than actually asssessed by the instructor (used as a team rating). The second part of semester 2 basically proves that student self and team ratings did not show much difference was because it was face-to-face assessemnt process. In the case with SET 2 results, shown in Fig. 2 b), the self ratings were significantly higher than team ratings in all weeks 1-12. Multiple t-tests were also conducted and show significant difference (p<0.005) in all weeks and a significant correlation in week 1-2 (*corr.* = 0.879, p = 0.001) and 5-6 (*corr.* = 0.602, p = 0.018). Again, the reason for this difference was that in week 1-6, there was a similar peer-blind assessment (Experimental (1) and Control (2)) conducted, using average peer ratings to derive AFCS. Whereas, in week 7-12, the face-to-face self ratings (Control (1) and Experimental (2)) of individual logbook entries with team ratings (conducted by investigator).

IV. Discussion and Conclusion

This study was designed to evaluate self and peer assessment in team based projects which can easily be applied to Biomedical Engineering curriculum. A quantitative analysis on this evaluation has determined that the self and peer assessment improve the cooperative learning. In comparing the 'peer-blind' and 'face-to-face' FCS methods, both had advantages and disadvantages. The 'peer-blind' self and peer assessment method applied would cause high discrepancy between self and team ratings. One would argue that this 'peer-blind' assessment is probably the 'fairer' assessment, where students in their team would not feel intimidated when rating each other. The second part of semester 2, the 'face-to-face' method on the other hand did not have the discrepancy issue and had actually proved to be a more accurate and effective, indicating team cohesiveness and good cooperative learning. The survey evaluation revealed that self and peer assessment methods helped students 'interact' and 'discuss' their issues with their own team peers and in receiving 'peer feedback' about their project contribution. Discussing their self and peer assessment ratings with their team peers addressed the perception issue and pointed out their realistic peer assessment and contribution to the project. This FCS assessment was applied to deliverable that is 40% of overall grade for the course and there are plans to apply it to 80% overall grade.

Acknowledgment

The author acknowledges the support from Dr. Jaideep Chandran, Prof. Iouri Belski, RMIT's Sports Statistics team and Prof. Ian Burnett.

References

[1] R.W. Brown, "Autorating: Getting individual marks from team marks and enhancing teamwork" *1995 ASEE/IEEE Frontiers in Education Conference*, Atlanta, Georgia, USA Nov, 1995.

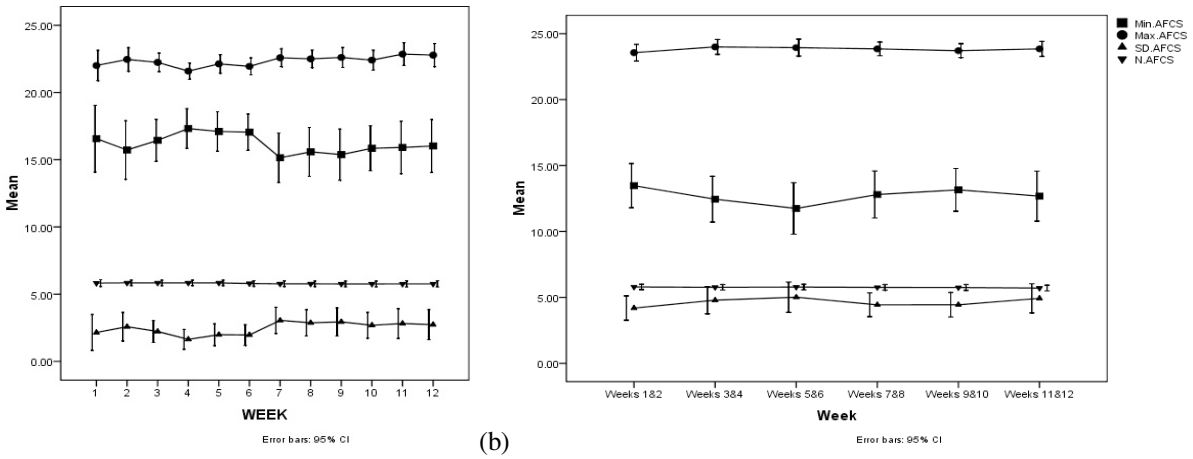
[2] K. A. Smith, "Cooperative Learning: Effective Teamwork for Engineering Classrooms", *1995 ASEE/IEEE Frontiers in Education Conference*, Atlanta, Georgia, USA Nov, 1995, pp. 2b5.13-18.

[3] N. Falchikov and J. Goldfinch, "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks", *Review of Educational Research*, 70(3), pp. 287-322, 2002.

[4] M. Lejk and M. Wyvill, "Peer Assessment of Contributions to a Group Project: a comparison of holistic and category-based approaches", *Assessment & Evaluation in Higher Education*, vol. 26, no. 1, pp. 61-72, 2001.
[5] K. Willey & M. Freeman, "Improving teamwork and engagement: the case for self and peer assessment", *Australasian Journal of Engineering Education*. Online publication 2006-02, 2006.
[6] N. Felchikov, "Peer Feedback Marking: Developing Peer Assessment", *Innovations in Education & Training International*, vol. 32, Issue 2, pp. 175-187, 1995.
[7] B. Oakley, R. Brent, R. M. Felder, I. Elhajj, "Turning Student Groups into Effective Teams", *Journal of Student Centered Learning*, vol. 2, no. 1, pp. 9-34, 2004.
[8] M. Freeman and J. McKenzie, "SPARK, A Confidential Web-Based Template for Self and Peer Assessment of Student Teamwork: Benefits of Evaluating across Different Subjects," *British Journal of Educational Technology*, vol. 33, pp. 551-69, 2002.
[9] SPARK, (2007) Self and Peer Assessment Resource Kit, http://spark.uts.edu.au/

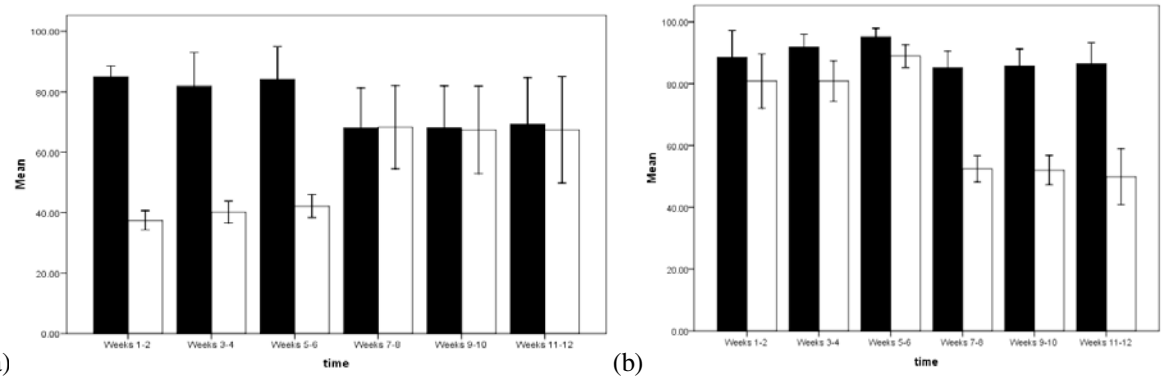
TABLE III. FCS METHOD AND FORMULA EMBEDDED IN A MATRIX TEMPLATE (EXCEL SPREADSHEET).

Team Fair Contribution Scoring (FCS)				**Note:** These spreadsheet columns (→) extend to week 12.					
Team No.	2								
	WEEK 1/1-2			**WEEK 2/3-4**			**WEEK 3/5-6**		
	Team Rating	FCS (20)	Adjusted FCS	Team Rating	FCS (20)	Adjusted FCS	Team Rating	FCS (20)	Adjusted FCS
Member 1	100	20.0	**20.0**	100	22.1	**22.1**	70	19.5	**19.5**
Member 2	100	20.0	**20.0**	100	22.1	**22.1**	70	19.5	**19.5**
Member 3	100	20.0	**20.0**	90	19.9	**19.9**	78	21.7	**21.7**
Member 4	100	20.0	**20.0**	90	19.9	**19.9**	75	20.9	**20.9**
Member 5	100	20.0	**20.0**	95	21.0	**21.0**	70	19.5	**19.5**
Member 6	100	20.0	**20.0**	93	20.6	**20.6**	70	19.5	**19.5**
Member 7	100	20.0	**20.0**	65	14.4	**14.4**	70	19.5	**19.5**



(a) (b)

Figure 1. The weekly distribution of AFCS min, max, S.D. and N (count) data for semester 1 (a) and 2 (b).



(a) (b)

Figure 2. Self and team ratings for SET 1 (a), SET 2 (b). Black bars represent self rating and white bars the peer rating out of 100%.