PEER ASSESSMENT OF STUDENT TEAMWORK: CASE STUDIES INVOLVING MULTICULTURAL PROJECT-BASED LEARNING

Ronald J Hugo and Robert W Brennan

Schulich School of Engineering, University of Calgary

Peihua Gu

College of Engineering, Shantou University

ABSTRACT

This study compares peer assessment data on team performance from four different Project Based Learning (PjBL) experiences. Teams consisting of Chinese and Canadian undergraduate students, mostly from engineering programs, were studied. In all cases examined students were found to assess themselves higher than their peers. In general, the Chinese students were found to assess the Canadian students positively, while the Canadian students were found to assess the Chinese students negatively. This cultural difference was found to increase as project difficulty decreased. Outlier analysis indicated that Canadian males were most likely to give strong negative feedback, while Chinese females were the least likely to give this type of feedback. Chinese males were found most likely to give strong positive feedback, while Canadian females were the least likely to give this type of feedback.

KEYWORDS

teamwork, peer assessment, project-based learning, multicultural

INTRODUCTION

The use of active learning methods, be it Problem-Based Learning (PBL) or Project-Based Learning (PjBL), are recognized as strong pedagogical techniques by which students are able to learn both the academic content as well as the professional skills required to function in their field of study [1,2]. PBL has been in use for over 40 years in fields that include medicine, engineering, science and economics [3]. The fundamental premise behind both PBL and PjBL is that real-life problems become the context in which students learn [1,4].

In the development of professional skills through either PBL or PjBL, it is common for students to work in teams, thereby providing them with experiences that mimic the environment that they will encounter when they enter the workforce [5]. The ability to assess the teamwork skills of individuals who are working as part of a group require special assessment methods in order to properly evaluate the intended learning outcomes of the course, provide sufficient motivation to the students through the realization that these skills do influence their final grade, and provide compensation for the extra effort required in order to engage in a peer-learning activity [5, 6, 7, 8].

The current investigation examines the use of a team performance instrument that involves both self- and peer-assessment. The instrument enables students to assess both their performance and that of their teammates on a four-point scale. Evaluation dimensions include Participation, Leadership, Listening, Feedback, Cooperation and Time Management. The results are examined by considering the influence of University affiliation, degree program affiliation, gender, student friendship, student roommates, and country affiliation.

The team performance instrument is applied to four different PjBL experiences. The projects are of varying degrees of difficulty, and consequently it is also possible to examine if project difficulty has influence on the results. An attempt is also made to compare how Canadian and Chinese students approach the process of peer and self assessment.

Research Questions:

Given the importance of teamwork in PjBL and the growing interest in multicultural student learning experiences in a globalized economy, the assessment of teamwork skills in four Canadian-Chinese PjBL experiences is examined in this study. The investigation examines teamwork peer assessment data segregated for university affiliation, gender, friendship, and student roommates to determine if bias errors exist. The investigation explores if a relationship between teamwork peer assessment and the level of difficulty of a PjBL experience exists, and how cultural differences may influence teamwork peer assessment data. Statistical comparisons and outlier analysis are used to examine these research questions.

BACKGROUND

Four project-based learning experiences were performed as part of a May-June 2012 Group Study program at Shantou University (STU) in China that lasted for five weeks. The program involved 20 University of Calgary (UofC) engineering students (15 male, 5 female) and 20 Shantou University students (14 male, 6 female). Of the 20 STU students, 4 studied in the Faculty of Science.

During the five-week program, STU students remained in their regular student dormitories while UofC students lived in an International Student Dormitory. Canadian students were randomly assigned to rooms with two students of the same gender per room, and four students of the same gender per apartment (two bedrooms per apartment). Student friendship was assessed through instructor observations of student behaviour during the five week period, which included time at the Shantou University campus as well as time on three field trips within the Guangdong province (Chaozhou - 1 day, Guangzhou - 4 days, Nan'ao Island - 2 days).

METHODS

This section discusses the nature of the student sample, the instruments and measures used, and the procedures by which the instruments and measures were delivered to the student sample.

***Sample*:**

Students from both STU and the UofC applied to take the five-week long Renewable Energy Practicum course (along with a second Innovation and Entrepreneurship in Renewable Energy course not described in this paper). STU students took their normal course load in addition to the two five-week courses whereas UofC students were only enrolled in the two five-week long

courses. UofC students were selected based on a number of factors that included year in program, (third-year [Junior] students receiving preference), letter grade in an earlier Thermodynamics course, and student responses in the program application that included questions like “Reasons for wanting to participate in the program” and “What do you expect to gain by participating in this program?” STU students were selected based on their field of study, their year in the program (third-year students receiving preference), and their English language abilities - an important prerequisite given that all course instruction was provided in English.

Students worked on projects in teams of 5 students per team. Group assignments included at least one female student member per team and a minimum of at least two students from either UofC or STU on each team. Student teams were determined randomly at the start of each project. Although information about student personality was known using a MBTI survey, it was not used during the team formation process.

Instruments and Measures:

1) Instructor Observations

The instructional team was in the laboratory during scheduled laboratory periods, amounting to approximately 26 hours of student observations. Instructor observations were recorded daily in a journal and used to infer aspects of the PjBL experiences that were difficult to capture with the other instruments. Examples of instructor observations included the distribution of work within a team, the level of intensity of student activity, the need to force students out of the lab at the conclusion of a 4-hour laboratory session, and the occurrence of euphoric outbursts when a group of students would solve a challenging problem. Other observations included taking note of students who appeared to be either high or low achievers and the amount of peer instruction taking place during a particular project.

2) Team Project Review Survey

In order to quantify the PjBL learning process, a short (11 question) survey was developed [9]. This survey was completed by students at the end of each one-week long PjBL experience. It was used to report on the level of difficulty of each step of the exercise, the level of student involvement (number of tasks per team member), the level of learning associated with non-technical attributes, and the level of learning associated with technical attributes. Of importance to this investigation was a question that examined whether or not the number of students on a particular team was properly sized given the level of difficulty of the project.

3) Self- and Peer-Assessment

Students were asked to complete a self- and peer-assessment survey at the conclusion of each PjBL experience [10]. The peer assessment asked students to self assess as well as evaluate their peers on participation, leadership, listening, feedback, cooperation, and time management. These results were then combined for each student and used as a percentage of their final grade (5% per PjBL experience). Students were provided with the averaged results of how their peers evaluated their performance.

Procedures:

Each of the four PjBL experiences lasted for approximately five days. Each PjBL experience began with a lecture that introduced students to the project, distributed team assignments, and

highlighted information pertaining to areas of safety for that particular project. Laboratory sessions typically lasted for four hours during which students could work on building their system, testing their system, or working on the course deliverables (presentation and assembly / test document). On the final day of the week, students presented their projects in 15-minute presentations (10 minutes presentation, 5 minutes questions and answers). At the conclusion of each project, students were given the peer assessment survey as well the team project review survey. In order to comply with requirements for research involving human subjects, the results of the team project review survey were not released to the course instructional team until after the final grade had been determined and submitted.

RESULTS AND DISCUSSION

Instructor Observations and Project Execution:

The peer-assessment value for each project represented 5% of the total course grade. For each of the six categories (Participation, Leadership, Listening, Feedback, Cooperation, Time Management), students evaluated their peers and themselves on a four-point scale, as shown in Table 1. The maximum value that a student could receive from each of their peers was 24 points (6 categories and a four-point scale). The total that a student could receive from their entire team was 120 points (4 peer assessments and 1 self assessment). If a student received 100 out of 120 total potential points, they would receive 4/5 of the 5%, or 4% for the peer-assessment on that project.

Table 1. An Example Peer Assessment Data Set

Group Member	Participation	Leadership	Listening	Feedback	Cooperation	Time Management
Student A	2	3	3	2	3	2
Student B	4	3	2	3	2	3
Student C	3	3	2	3	2	2
Student D	1	3	4	2	3	3
Student E (self eval.)	3	4	3	2	3	4

Note: Be honest in your assessment. If a score of 4 is entered for every team member, all scores will be set to zero. This helps to prevent collusion within the group.

The peer assessment data was analyzed in an effort to quantify the difference between how a student assessed themselves with respect to their peers. This was quantified numerically by computing and summing the numeric difference between the self-assessed and peer-assessed values for each student pair. This process is illustrated in Table 2 using the data set from Table 1. If a student (the assessor) viewed their peer's (the assessee) performance negatively, a low negative number would result (-10, for example), while if the assessor viewed their peer's performance positively, a high positive number would result (+10, for example). On this scale, a summed difference of -10 translates into a 0.42% reduction in a student's grade through the peer assessment value (-10/120 * 5% = -0.42%).

Table 2. Example Data Set Analyzed with respect to Student E

Group Member	Participation	Leadership	Listening	Feedback	Cooperation	Time Management	Summed Difference
Student A	-1	-1	0	0	0	-2	-4
Student B	1	-1	-1	1	-1	-1	-2
Student C	0	-1	-1	1	-1	-2	-4
Student D	-2	-1	1	0	0	-1	-3

The data in Table 2 was categorized further by identifying the associations between students. For example, if Students A and E from Table 1 were *Roommates*, the summed difference value appearing in the first row of Table 2 would be flagged and put into the *Roommates* category. Given that the association in this example is *Roommates*, the summed difference value of -4 from Table 2 would contribute to the *Roommates* data set for that specific project. By segregating data for each student association for each project, it is possible to examine if student associations result in biases on a project-by-project basis.

Statistical Data Analysis:

All of the summed-difference values (Table 2) for all student pairs and projects were categorized by student association (University - N = # of samples = 245, Degree - N = 278, Gender - N = 332, Friends - N = 8, Roommates - N = 16, Chinese-Chinese - N = 130, Chinese-Canadian - N = 190, Canadian-Chinese - N = 190, Canadian-Canadian - N = 132) and averaged. The bar graph shown in Figure 1 presents the results of this categorizing and averaging process.

Examining the University, Degree, Gender, Friends and Roommates student associations in more detail, only the Roommates association was found to be positive, although of low statistical significance given that it is based on only 16 samples. The University association resulted in the most negative of the five main associations. Examining evaluations within cultures, students were found to also evaluate peers from their own country negatively (Chinese-Chinese and Canadian-Canadian). Evaluations across cultures were, however, split with Chinese students evaluating Canadian students positively and Canadian students evaluating Chinese students negatively.

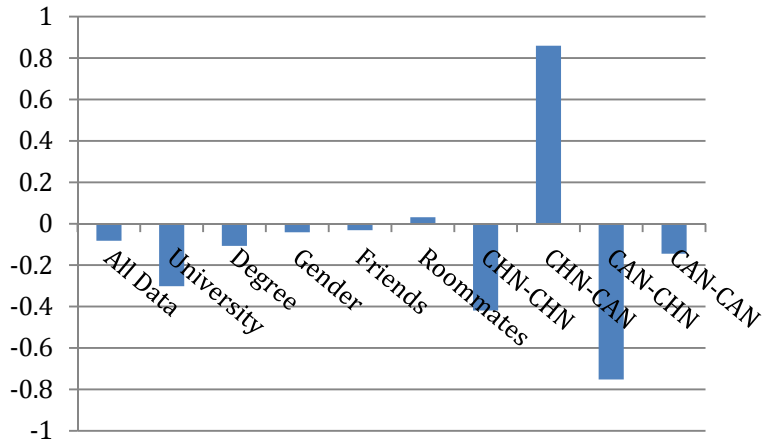

Figure 1. Average Peer Assessment Values for All Projects

The All Data category indicates that on average the students assess the teamwork skills of their peers lower than their own teamwork skills. The exception to this is the Chinese students evaluating the Canadian students. It is interesting to note that Chinese students are more critical of other Chinese students (-0.419) than Canadian students are of other Canadian students (-0.144). Other student associations like Gender, Friendship, or Roommates were found to be relatively negligible.

In order to understand the results in more detail, the data was further examined on a project-by-project basis. The All Data category from Fig. 1 is expanded on a per-project basis in Fig. 2. The average, although negative for all projects, remains close to zero. The largest standard

deviation is for Project 2 with a value of 2.15. It is speculated that Project 2 has the largest standard deviation given that, as will be shown later in the paper, it is the most challenging project and consequently has the ability to generate the largest amount of team conflict.

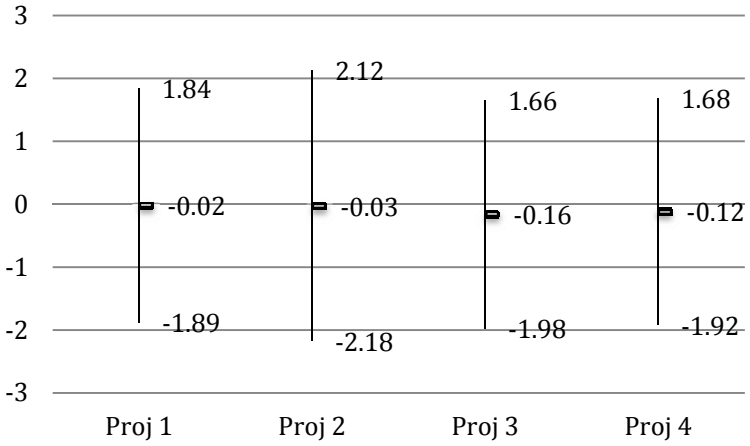

Figure 2. Peer Assessment Data by Project - All Data

The project-by-project breakdown for University Affiliation and Degree Program Affiliation is shown in Fig. 3. The University Affiliation data is consistent with the results for All Data in that students on average evaluate their peers lower than themselves. The Degree Program data, shown on the right in Fig. 3, reveals results that average close to zero for all four projects. Consequently neither University Affiliation nor Degree Program indicate any sort of systematic bias in the peer assessment data.

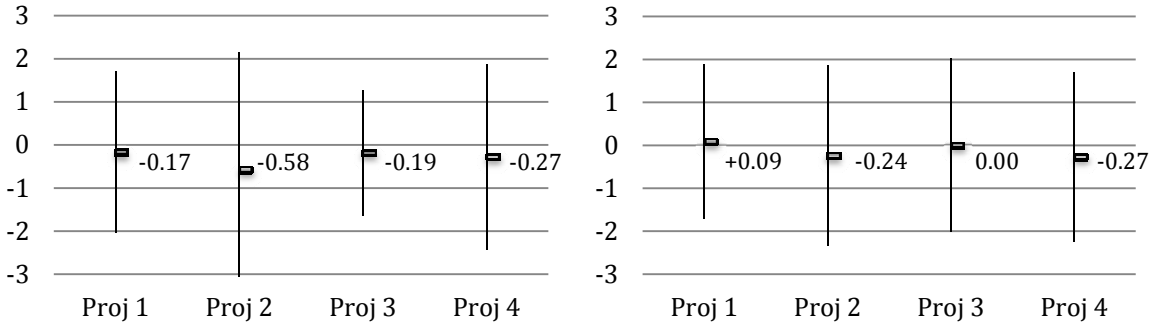

Figure 3. Peer Assessments - University Affiliation (left) and Degree Program (right)

The peer assessment results of students from the same country (and culture) are shown in Fig. 4. The Chinese students are found to consistently evaluate their Chinese peers lower than themselves, whereas the Canadian students are seen to evaluate their Canadian peers lower than themselves on all projects with the exception of Project 4.

The cross-cultural comparison of students from different countries is shown in Fig. 5. The Chinese students are noted to consistently assess the Canadian students in a positive manner, whereas the Canadian students are found to assess the Chinese students in a generally negative manner. The only exception to this trend is Project 2 where the Canadian students assess the Chinese students in a near neutral manner (near zero mean).

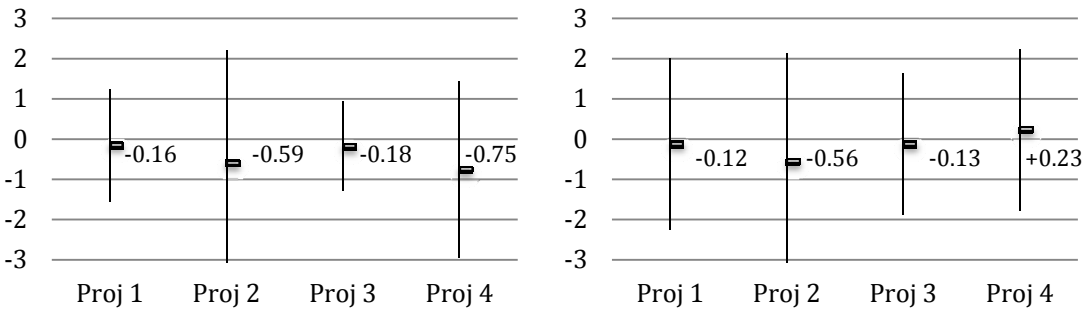

Figure 4. Peer Assessments - Chinese-Chinese (left) and Canadian-Canadian (right)

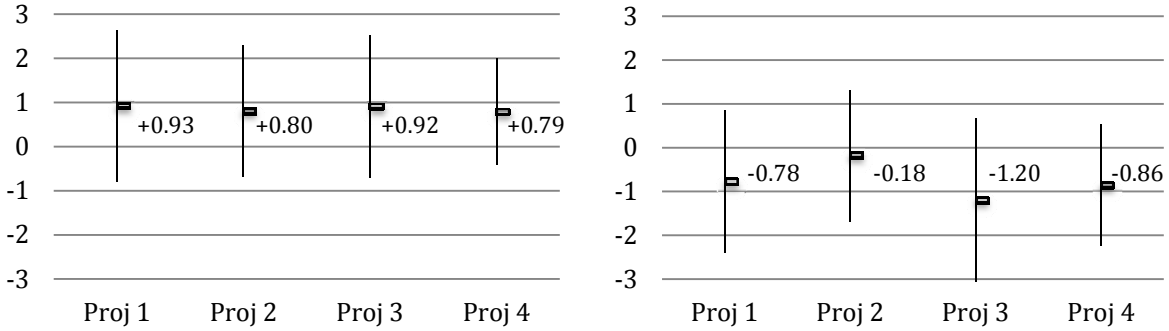

Figure 5. Peer Assessments - Chinese-Canadian (left) and Canadian-Chinese (right)

Examining in finer detail the distribution of Peer Assessment data when Chinese students assess Canadian students (Chinese-Canadian), Fig. 6 demonstrates what approximates a normal distribution with a positive offset of approximately 1.0 for Projects 2-4. Project 1 shows a more one-sided distribution. This indicates that, in general, the Chinese students assessed their Canadian peers stronger than themselves in teamwork skills.

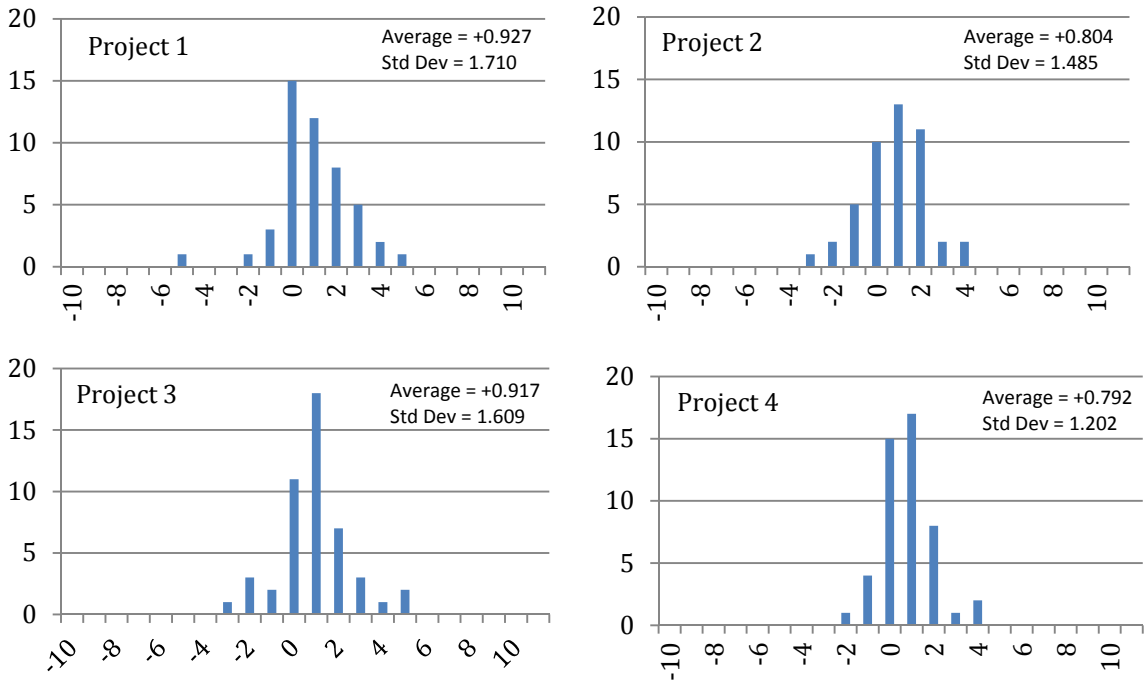

Figure 6. Chinese-Canadian Peer Assessment

As mentioned, the Canadian students were shown to assess their Chinese peers lower than themselves. The exception to this was for Project 2. In order to explore this result further, histograms of Canadian-Chinese peer assessment data for all four projects were produced as shown in Fig. 7. By inspecting the histograms, it becomes apparent that, in addition to the near zero average, the Project 2 distribution was closest to a normal distribution. Although interesting, this does not provide information about why the average for Project 2 is different. In order to attempt to answer this question, project difficulty is explored next.

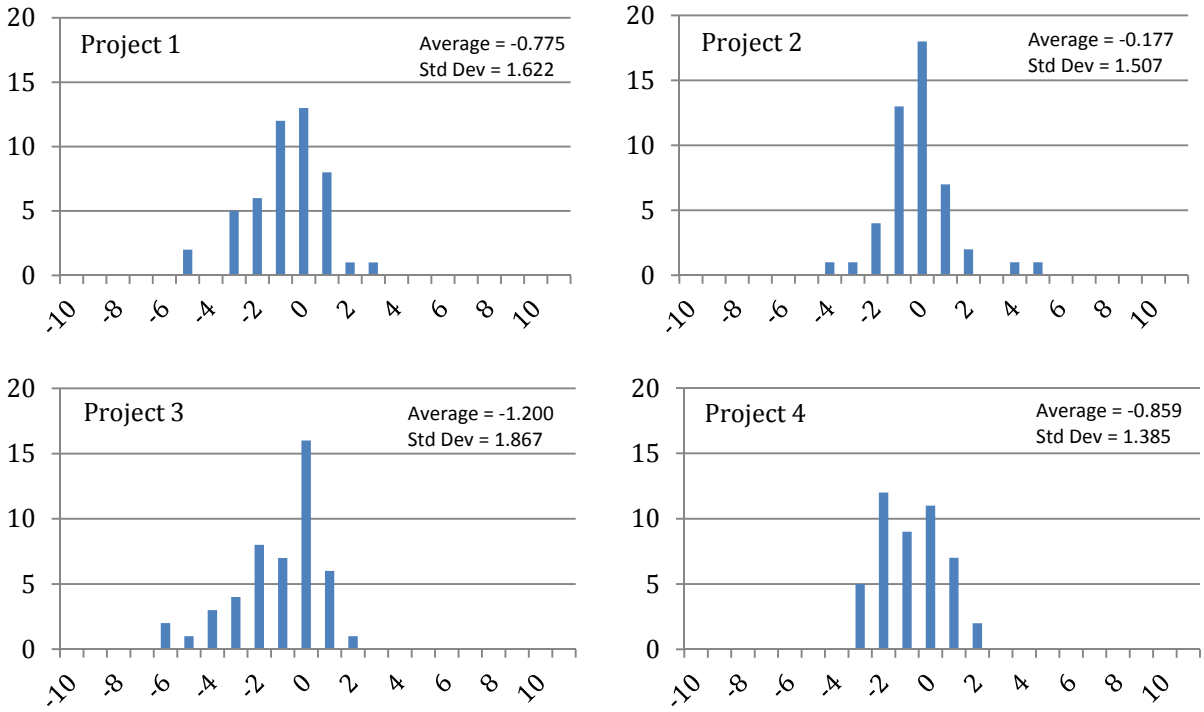

Figure 7. Canadian-Chinese Peer Assessment

Project 2 resulted in Canadian students assessing their Chinese teammates more positively (-0.177) than for the other projects (-1.20 for Project 3, for example). During the course, a survey was conducted after each project concluded where the students were asked a series of questions, one of which was "Our team had the right number of students for the scope and allotted time." to which the students responded on a 5-point scale that ranged from 0% (strongly disagree) to 100% (strongly agree). The last column of Table 3 shows the results of this survey question for all four projects. It is noted that for Project 2 the students felt that their team was most properly sized given the complexity of the project.

Table 3
Project Difficulty: Number of Steps, Hours to Complete, Correct Team Size

Project	Number of Steps to Complete	Hours to Complete Project	Lab Overtime Requests	Correct Team Size? % YES (UofC / STU)
#1 - Solar Cell	13	4	None	85% / 93%
#2 - Wind Turbine	18	10-12	Often	94% / 98%
#3 - Solar Thermal	11	2	None	76% / 88%
#4 - Solar Fan	11	8	Some	61% / 86%

This finding is supported by comments made in Ref. [11] stating that social loafing and free-riding do not occur when (1) group goals are compelling; (2) the task involved is challenging, appealing, or involving; and (3) individual performances are identified, monitored, and rewarded. The other data provided in Table 3 indicates that Project 2 took the longest to complete and posed the greatest challenge to the student teams given the number of lab overtime requests. Consequently, it can be concluded that the more challenging nature of Project 2 resulted in the Shantou students becoming more involved in the project, thereby receiving more positive Peer Assessments by the Calgary students. This indicates that there is a relationship between project difficulty and Peer Assessment results, with more difficult projects forcing the more passive students to become engaged, thereby increasing their Peer Assessment values.

Outlier Data Analysis:

A final form of analysis was performed by examining statistical outliers in the peer assessment data. In doing this, an outlier was approximated as a data value that was more than two standard deviations away from the mean. In order to simplify matters, this was taken as data points with an absolute value equal to or greater than 4.0. Negative data points were considered as Strong Negative Feedback and these results are shown in Table 4, while positive data points were considered as Strong Positive Feedback and these results are shown in Table 5. The last column in both tables indicates the percentage of each type of student in the class.

Table 4 - Strong Negative Feedback
Outlier Analysis - Assessor and Assessee [N = 26]

Student	Assessor	Assessee	% Class Representation
STU Male	38.5%	38.5%	35.0%
UofC Male	50.0%	23.1%	37.5%
STU Female	0.0%	26.9%	15.0%
UofC Female	11.5%	11.5%	12.5%

The data shown in Table 4 indicates that UofC Males gave strong negative feedback more often than their class representation, and they received strong negative feedback less than their class representation. STU Female students gave no strong negative feedback, but yet they received it more often than one would expect given their class representation. Of the 26 incidents reported in Table 4, 7 were directed towards the non-engineering STU students. Consequently 26.9% of the strong negative feedback incidents were directed towards only 10% of the total student population.

Table 5 - Strong Positive Feedback
Outlier Analysis - Assessor and Assessee [N = 15]

Student	Assessor	Assessee	% Class Representation
STU Male	46.7%	13.3%	35.0%
UofC Male	33.3%	73.3%	37.5%
STU Female	20.0%	0.0%	15.0%
UofC Female	0.0%	13.3%	12.5%

Data for strong positive feedback is shown in Table 5, from which it can be seen that STU Males were the ones to most often give strong positive feedback. UofC Males received strong positive feedback more often than one would expect given their class representation. Surprisingly, STU

Females never received strong positive feedback while Canadian Females never gave strong positive feedback.

The results of Tables 4 and 5 provide insight into which student groups gave and received strong feedback, however, it does not indicate how this feedback flowed between the student groups. In order to more fully understand how the student groups interacted, the magnitude and direction (i.e. the flow) of strong feedback is next examined.

The flow of strong negative feedback is illustrated in Fig. 8. The flow diagram shows that Canadian Males gave strong negative feedback to the other three student groups in a manner that was consistent with the student representation of each group. The Chinese Female students received strong negative feedback, with the majority of it coming from the Chinese Male students. There was no flow between the two Female groups.

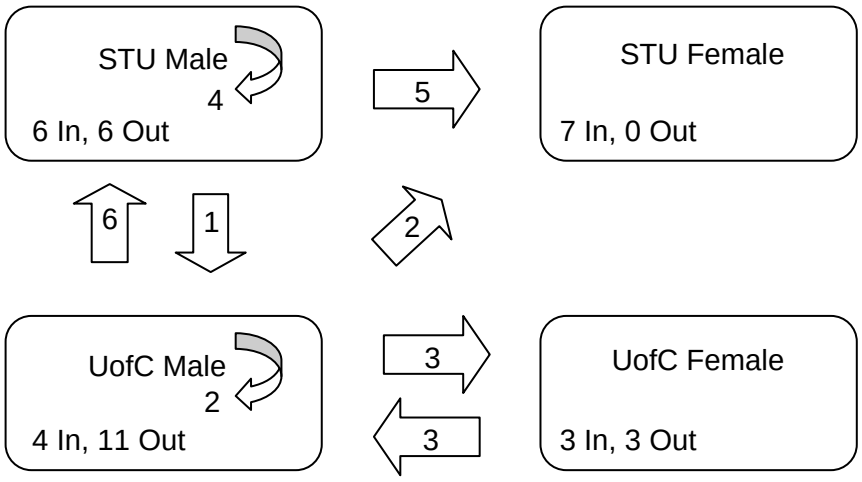

Figure 8: Strong-Negative-Feedback Flow Diagram

The flow of strong positive feedback is illustrated in Fig. 9. This shows that Canadian Males received all of their strong positive feedback from the Chinese students, and again, the two Female groups did not exchange feedback.

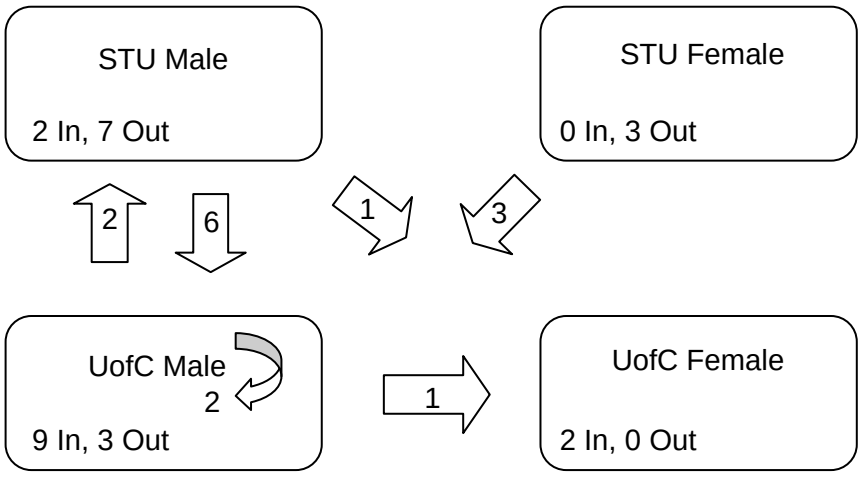

Figure 9: Strong-Positive-Feedback Flow Diagram

These results show that the female students did not have strong positive / negative feedback between one another. This results due to the fact that only 11 of the 40 students were female, and for each project, only 3 teams had the ability for female-female interaction with two female students. The other 5 teams had only 1 female student each. The majority of the strong negative feedback received by the Chinese Female students is seen to come from the Chinese Male students.

CONCLUSIONS

This paper has investigated teamwork peer assessment for multi-cultural teams of students involved with Project-Based Learning experiences. The results indicate that, in general, students tend to evaluate their own teamwork skills higher than that of their peers. Based on the entire data set, there was no strong evidence of bias error with university affiliation, gender, friendship, or student roommates. The strongest source of bias was found to exist between Chinese students themselves and also between Canadian and Chinese students.

The Canadian students assessed the Chinese students negatively for all of the projects with the exception of the second project, which was shown to be the most difficult project. This result indicated that for the second project the Canadian students evaluated their Chinese classmates on par with themselves (near zero mean). This implied higher team functionality for this project. This result agrees with results that have been found by other researchers and demonstrates a relationship between teamwork peer assessment results and project difficulty.

Canadian Males were found to give the largest amount of strong negative feedback, and that they distributed this feedback equally (based on percentage representation) to the other student groups. Chinese Females received the largest amount of strong negative feedback with the majority of the negative feedback coming from the Chinese Males.

The Chinese Males gave the largest amount of strong positive feedback, while the Canadian Males received the largest amount of strong positive feedback. The Chinese Females received no strong positive feedback at all. On average, Male students were more likely to provide either strong positive or strong negative feedback.

ACKNOWLEDGEMENT

The authors wish to thank Shantou University, the University of Calgary, and the Li Ka-Shing Foundation of Canada for providing financial support for this project.

REFERENCES

[1] Biggs, J., Tang, C., Teaching for Quality Learning at University, 3rd Edition, McGraw Hill, 2007.

[2] Shuman, L.J., Besterfield-Sacre, M., and McGourty, J. “The ABET ‘Professional Skills’ – Can They Be Taught? Can They Be Assessed?”, Journal of Engineering Education, Vol. 94, No. 1, 2005, pp 41-55.

[3] Savery, J.R., “Overview of Problem-based Learning: Definitions and Distinctions”, Interdisciplinary Journal of Problem-based Learning, Vol. 1, No. 1, 2006, pp 9-20.

[4] Crawley, E., Malmqvist, J., Ostlund, S., Brodeur, D., Rethinking Engineering Education – The CDIO Approach, Springer, 2007.

[5] Boud, D., Cohen, R., Sampson, J., "Peer Learning and Assessment," Assessment and Evaluation in Higher Education, Vol. 24, No. 4, 1999, pp. 413-426.

[6] Conway, R., Kember, D., Sivan, A., and Wu, M., "Peer Assessment of an Individual's Contribution to a Group Project," Assessment & Evaluation in Higher Education, 18:1, 1993, 45-56.

[7] Goldfinch, J., "Further Developments in Peer Assessment of Group Projects," Assessment & Evaluation in Higher Education, 19:1, 1994, 29-35.

[8] Lejk, M., Wyvill, M., Farrow, S., "A Survey of Methods of Deriving Individual Grades from Group Assessments," Assessment & Evaluation in Higher Education, 21:3, 1996, 267-280.

[9] Brennan, R.W., Hugo, R.J., Gu, P., "Reinforcing Skills and Building Student Confidence through a Multicultural Project-Based Learning Experience," Proceedings of the 8th International CDIO Conference, Queensland University of Technology, Brisbane, July 1-4 2012.

[10] Peer Assessment Collaboration Rubric, Intel Teach Program – Designing Effective Projects, http://download.intel.com/education/Common/en/Resources/DEP/density/density_peer_rubric.pdf, website accessed January 2012.

[11] Beatty, J.R., Haas, R.W., Sciglimpaglia, D., "Using Peer Evaluations to Assess Individual Performances in Group Class Projects," Journal of Marketing Education, 18:17, 1996, 17-27.

BIOGRAPHICAL INFORMATION

Ronald J. Hugo is Professor and Head of Mechanical and Manufacturing Engineering at the University of Calgary. He is also the holder of the Engineering Education Innovation Chair in the Schulich School of Engineering. His research interests are in the areas of experimental fluid dynamics, energy systems, and engineering education.

Robert W. Brennan is a Professor of Mechanical and Manufacturing Engineering and the Associate Dean (Academic & Planning) at the Schulich School of Engineering. He has served on the steering committee of the Canadian Engineering Design Education Network (CDEN) and as chair of the Schulich School of Engineering's Engineering Education Summit.

Peihua Gu is Provost and Vice President (Academic) of Shantou University. He was one of the four founders of the Canadian Design Engineering Network (CDEN) which sought to increase the amount of design content in Canadian engineering undergraduate programs. His current scholarly interests focus on adaptable design, life cycle engineering, and engineering education.

Corresponding author

Dr. Ron J Hugo
Schulich School of Engineering
University of Calgary
2500 University Dr NW
Calgary, AB, Canada T2N 1N4
1-403-220-2283
hugo@ucalgary.ca