



A Comparison of Peer Evaluation Methods in Capstone Design

Dr. Joshua A Enszer, University of Maryland Baltimore County

Dr. Joshua Enszer is a full-time lecturer in Chemical Engineering at the University of Maryland Baltimore County. He has taught core and elective courses across the curriculum, from introduction to engineering science and material and energy balances to process control and modeling of chemical and environmental systems. His research interests include technology and learning in various incarnations: electronic portfolios as a means for assessment and professional development, implementation of computational tools across the chemical engineering curriculum, and game-based learning.

Prof. Mariajose Castellanos, University of Maryland, Baltimore County

A Comparison of Peer Evaluation Methods in Capstone Design

Abstract

The final chemical engineering capstone design course at our university consists of a semester-long project where students work in predetermined groups. The emphasis of the experience is to allow students to develop and explore their engineering creativity while at the same time cultivate the importance of communication and teamwork skills. Students receive an industrially relevant open-ended problem statement and face a rigorous schedule with specific weekly deliverables. We concentrate on building teamwork skills, ownership of the project, accountability, communication and self-assessment. We use the projects to incorporate the skills our graduates need to join the workforce or graduate schools while inspiring students to evaluate, defend, and treasure their creations. Measuring group dynamics, division of labor, engagement, group versus individual experiences is a challenging task. Student grades are highly dependent on peer evaluation, as the team does not receive the same grade for group assignments, but instead, group grades are multiplied by a *peer assessment* factor (for example if student's X peer assessment is 50 percent, all group grades for that student will be multiplied by 0.5. In the spring of 2012, two separate methods were employed as part of peer assessment: a traditional "point division" method¹ and CATME, the Comprehensive Assessment of Team Member Effectiveness². The point division method instructs students to divide 100 points between the group members; if there are five team members in a group and a student believes all the students in the group participated equivalently in the project then he/she will assign 20 points to every student in the group. CATME allows the students to evaluate team members in more behavioral areas: contribution to the team, interaction with peers, keeping team on track, quality of work, proficiency in the discipline. Because the point division method is effectively a zero-sum game whose philosophy may contradict the cooperative nature of the course, we are interested in replacing it, and therefore have explored correlations between it and CATME. Additionally, we used the data from both methods to evaluate how students rated themselves compared to the ratings of their peers. In this report we share the results of our analysis. We have determined that the two methods have a statistically significant positive correlation and that, with modest significance, students on average self-evaluate higher than their peers evaluate them.

Introduction

The two-semester chemical engineering capstone design course sequence at our university concludes with a semester-long project where students work in predetermined groups. The emphasis of the experience is to allow students to develop and explore their engineering creativity while at the same time cultivate the importance of communication and teamwork skills. Students receive an industrially relevant open-ended problem statement and face a rigorous schedule with specific weekly deliverables. We concentrate on building teamwork skills, ownership of the project, accountability, communication and self-assessment. We use the projects to incorporate the skills our graduates need to join the workforce or graduate schools while inspiring students to evaluate, defend, and treasure their creations.

Measuring group dynamics, division of labor, engagement, group versus individual experiences is a challenging task. Student grades are highly dependent on peer evaluation, as the team does not receive the same grade for group assignments, but instead, group grades are multiplied by a *peer assessment* factor (for example if student's X peer assessment is 50 percent, all group

grades for that student will be multiplied by 0.5. Traditionally, this factor has been determined through a “point division” method¹ in which all students are tasked to divide 100 points among their team proportional to their perception of how work was accomplished.

In the spring of 2012, CATME, the Comprehensive Assessment of Team Member Effectiveness², was used to supplement this traditional method. The tool is publicly available at www.catme.org and requires an instructor to register for access. CATME allows the students to evaluate team members in more behavioral areas rather than focus strictly on numerical points; additionally the students do not see the numerical ratings they receive or award each other using this method. Because the point division method is effectively a zero-sum game whose philosophy may contradict the cooperative nature of the course, we have explored correlations between it and CATME. Specifically, we are interested in replacing the traditional method with the more behavior-based assessment provided through CATME. Additionally, we used the data from both methods to evaluate how students rated themselves compared to the ratings of their peers. In this report we share the results of our analysis.

Methods

Our institution is a medium-sized, Northeastern, public institution whose student body is comprised of approximately 75% full time and 25% part time students. The overall student population is 53% male and 47% female and is diverse with about 40% of students representing minority populations. In chemical engineering, the student population is 63% male and 37% female with 46% minorities. In the Spring of 2012, 39 students completed the capstone design sequence. This population was 51% male and 49% female, 28% Asian, 13% black, 10% other and 49% white as reported by the students.

Process design is a two-semester sequence in the senior year; its prerequisites include thermodynamics, reaction kinetics, fluid dynamics, and heat and mass transport. Separations is a co-requisite for the first semester of design, and while process control is not a strict pre-requisite, more than 75% of students have completed it prior to the first semester of design. In the 2011-2012 academic year, Design I focused primarily on piping and instrumentation diagrams and process economics. Design II served as the capstone course to bring together all other aspects of the curriculum. It is a four-credit course that meets 4.5 academic hours per week. Class time is largely devoted to project updates, and all students are required to give four presentations over the course of the semester.

The Comprehensive Assessment of Team Member Effectiveness, CATME, has been developed by faculty from several universities, through support of the National Science Foundation, as an effort to develop a peer evaluation instrument that is reliable and valid. Its tools include a survey called Team Maker³ to collect demographic data, such as gender and race, as well as logistics like a weekly schedule of times students are free to work together. There is also a series of survey questions developed for Peer Evaluation⁴, which can be used to explore student perception on group work, team cohesiveness, and other team-related attitudes.

Project teams for Design II are initially formed using the Team Maker survey and input from the academic advisors of the chemical engineering department. Team Maker is used to create a “first draft” of project teams, and members are rearranged based on advisors’ knowledge of previous chemical engineering teamwork experiences. Team Maker is then used to compare the schedules

of availability for all project teams to ensure that they are able to meet for a minimum of ten hours per week outside of class.

The CATME Peer Evaluation survey was assigned three times across the semester, roughly once every five weeks, to monitor team dynamics. Students were encouraged to use the comments section to voice any concerns that they had at each point in the term. The results of the standard behavioral categories are shared graphically with each team member, letting each know how their peers rated themselves and one another in five categories: Contributing to the Team's Work, Interacting with Teammates, Keeping the Team on Track, Expecting Quality, and Having Related Knowledge, Skills, and Abilities.

The standard point division survey was assigned once at the end of the semester, as has been done traditionally in our department every year. In this survey, students submit the 100 points division among the members of their team, including themselves, to represent their perspective on the equity of contributions to work across the semester through email to the instructor, written feedback is not required, in 2012 at least 50% of the students included written feedback in the submission.

The CATME Peer Evaluation survey automatically computes a series of numerical scores called "Adjustment Factors" based on the behavioral ratings of each student. Adjustment Factors are computed both with and without a student's self-ratings, and by default are "capped" at a maximum of 1.05 (also rounding any numbers between 0.95 and 0.99 up to a 1.0). It is possible to also obtain "raw" Adjustment Factors where no capping or rounding occurs. More information about how these factors are computed can be found in the original literature.³

Adjustment Factors from the point division method are likewise computed in two ways, both with and without a student's self-ratings. They are computed by taking the average of all numerical scores and dividing by the highest average in the group. In this way, the highest Adjustment Factor is always a 1.0, and there is always a 1.0 in each group.

For the purposes of comparison to the point division method, we thought it would be most appropriate to use the "capped" Adjustment Factors in CATME, since the other method is effectively capped. Two comparisons are made between the two methods, one factoring in a student's own ratings into the computation, and one ignoring self-assessment. Further, two more comparisons can be made with a single method each, by examining the relationship between scores with and without a student's own assessment scores factored in.

Results and Discussion

For the following comparisons, the entire population of the Spring 2012 course (n=39) is used. A one-sample t-test is conducted on the population correlation coefficient to determine if there is a statistically significant correlation between the adjustment factors computed by both CATME and the point division method.

Point Division vs. CATME (including self-evaluation)

Figure 1 shows the results of student ratings including their own self-assessment scores in the computation. There are fourteen students who received ratings of 1.05 from the CATME computation and eight students with ratings of 1.0 (since there are eight groups). The point

(1,1.05) in Figure 1 is actually seven points. The correlation coefficient r for the data is 0.86, which virtually guarantees statistically significant correlation ($\alpha < 0.001$). The R^2 -value for the data is about 0.73, which means that while there is definitely a positive correlation between the data, one method should not necessarily be used to predict the result of the other.

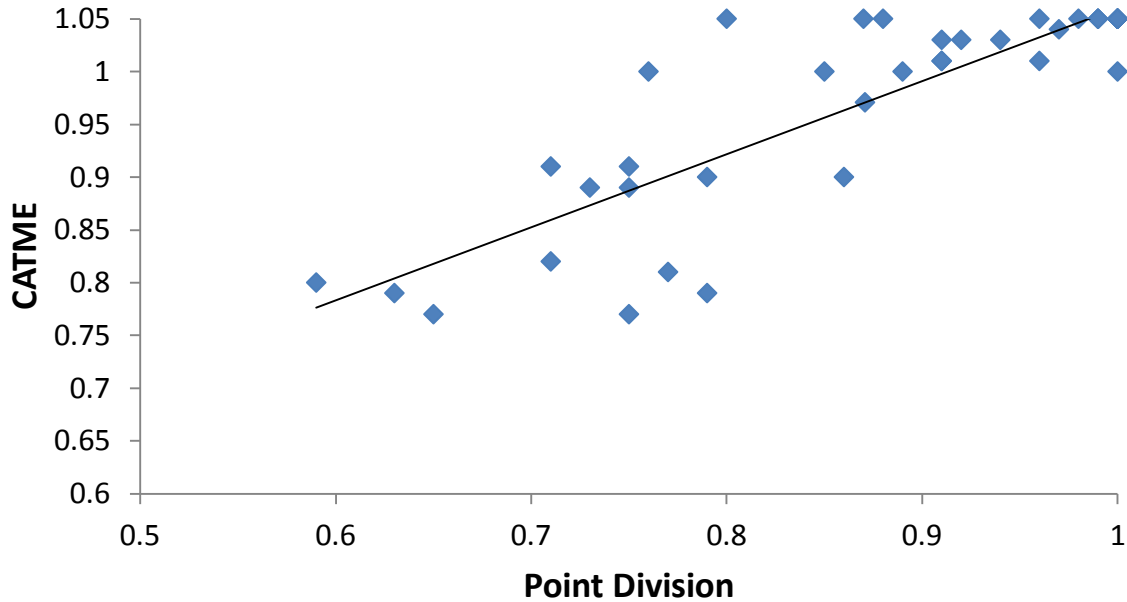


Figure 1: Point Division and CATME Adjustment Factors with Self-Evaluation

Point Division vs. CATME (without self-evaluation)

Figure 2 shows the results of student ratings excluding their own self-assessment scores in the computation. There are fifteen students who received ratings of 1.05 from the CATME computation. The point (1,1.05) in Figure 2 is actually eight points. The correlation coefficient r for this data is 0.83, which also virtually guarantees statistically significant correlation ($\alpha < 0.001$), but again a lower R^2 value suggests a linear regression of the two methods is unhelpful in predicting specific results from one based on the other.

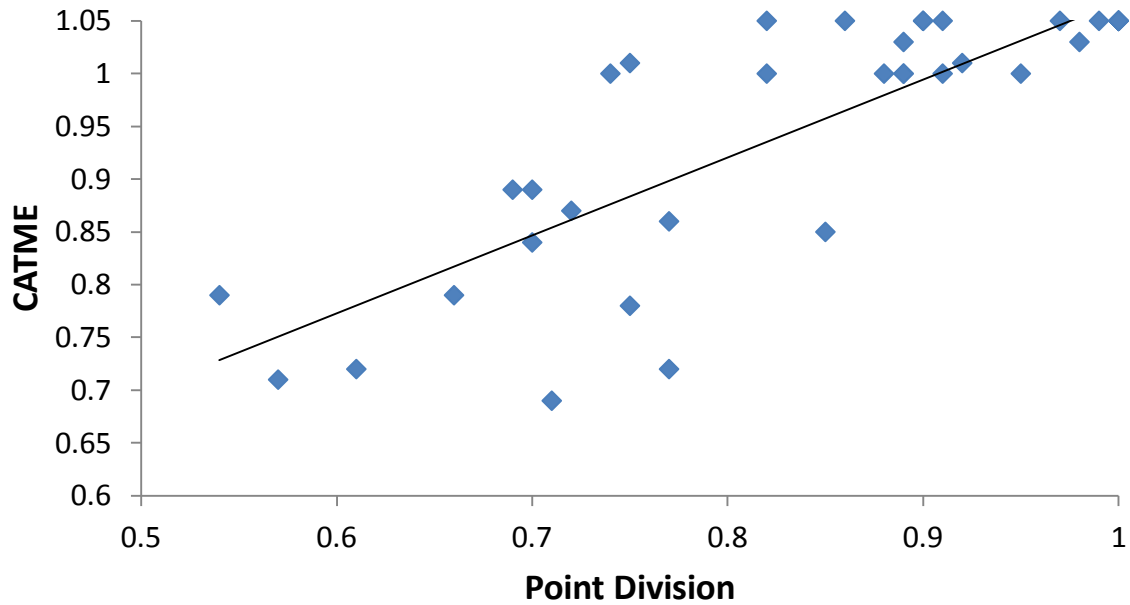


Figure 2: Point Division and CATME Adjustment Factors without Self-Evaluation)

Ratings Including vs. Excluding Self-Evaluation (using Point Division)

We hypothesized that on average, student scores including self-evaluation would be higher than those that do not include a student's own self-evaluation in the Adjustment Factor computation. A plot of the two evaluation scores as computed using the traditional point division method is shown in Figure 3. The 45° diagonal line represents the line where a student's Adjustment Factor is unaffected by the choice to include or exclude self-evaluation in the computation. The line of best fit through the data has a slope of 1.08, meaning that on average it would appear that our hypothesis is correct. However, in computing the confidence interval on the slope, we find that we can only claim a slope statistically greater than 1 for a modest level of significance ($\alpha < 0.2$). This suggests that we cannot conclusively say that the point division method is different when choosing to include or exclude self-evaluation. By inspection, it appears this happens for two reasons: the highest performing group members are often also modest, and the lowest performing group members have a skewed interpretation of their own contributions to their group.

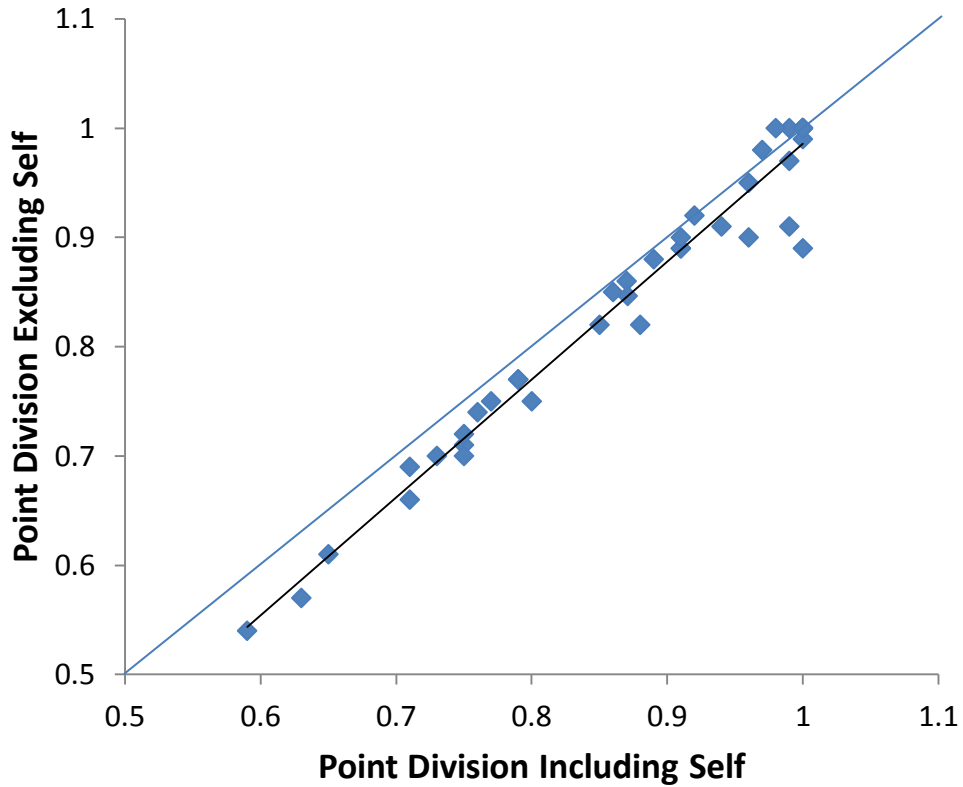


Figure 3: Adjustment Factors Including and Excluding Self-Evaluation in Point Division Method

Ratings Including vs. Excluding Self-Evaluation (using CATME)

The results using CATME for Adjustment Factor computations are similar as seen in Figure 4, though the slope of the line is a little steeper, at 1.13. Again, this slope is statistically greater than 1 for only a modest level of significance ($\alpha < 0.16$). In our set of 39 data points, only one lies above the 45° diagonal, meaning only one person had a lower Adjustment Factor when including his or her own self-evaluation.

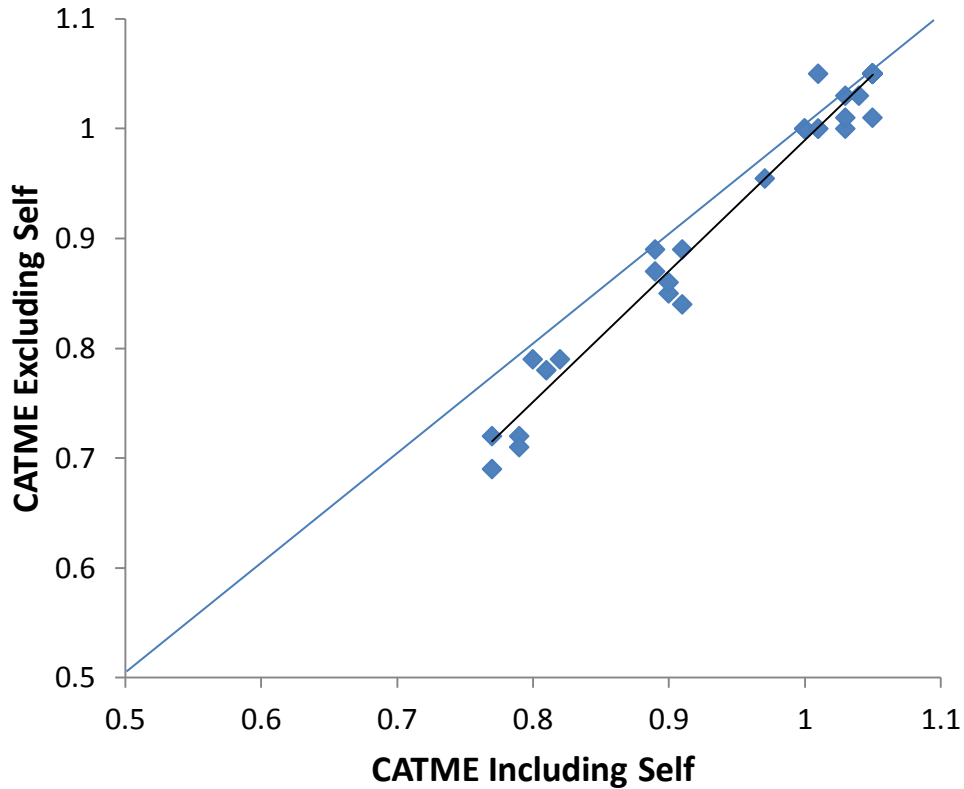


Figure 4: Adjustment Factors Including and Excluding Self-Evaluation in CATME

Conclusion

The main purpose of this investigation was to determine if there is a relationship between the results of the traditional “point division” method of peer evaluation and the Comprehensive Assessment of Team Member Effectiveness and whether it is suitable to outright replace the former with the latter for our final senior capstone design course. Using data from 39 students in the spring of 2012, we have discovered that there is a strongly statistically significant correlation ($\alpha < 0.001$) between the two methods, but that one does not easily predict the result of the other. CATME provides the benefit of automatic computations of peer evaluation scores and feedback for multiple behaviors, versus the single numerical value from the point division method, though both peer evaluation systems are supplemented with additional written comments.

We also looked at the differences between peer evaluation scores when a student’s self-evaluation is either included or excluded. While the majority of students seem to have higher scores when their own self-evaluation is included in the computation of this score, the statistical significance of this observation is more modest, on the level of $\alpha = 0.2$.

We did not implement any measure to determine student preferences for the two types of evaluation systems. Both the point division and CATME systems are being implemented again in the Spring 2013 semester and plans for a student attitude survey are ongoing.

References

- [1] Michaelsen, L. K., A. B. Knight, and L. D. Fink (eds.), *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*. Sterling, Va.: Stylus, 2004.
- [2] Ohland, Matthew W., H. R. Pomeranz, and H. W. Feinstein, "The Comprehensive Assessment of Team Member Effectiveness: A New Peer Evaluation Instrument," *Proceedings of the American Society of Engineering Education Annual Conference*, Chicago, IL, June 2006.
- [3] 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), 1-28.
- [4] 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.