Individual Capstone Assessment Using Z-Scores

M. Austin Creasy*
Purdue University
mcreasy@purdue.edu

L. Eric Stacy
Purdue University

Abstract
Capstone courses and the associated projects are the culminating learning experience for many engineering programs. Students are placed in teams with an assigned project that simulates working in an industrial setting. Grading individual students within these teams can be challenging for an instructor because many of the course deliverables are the product of a team effort. A grading scheme has been developed, over several years of capstone courses, that assists in grading the individual student relative to the team's success on the project and the individual team member's contributions as assessed by the team. This process uses a variety of survey questions related to the project where each team member rates themselves and their peers over a 10-point scale. Stipulations are placed on the rating to ensure that students provide a range in the assessment for the group members. The survey scores are normalized using the z-score in an attempt to remove individual grading variances between each team member. Each student's individual grade is calculated from the team average grade, as assessed by the instructor, and the individual's z-score calculated from the survey responses. This paper will show the developed survey questions, how the z-score is calculated, how the z-score is combined with the team assessed average score to provide an individual grade for a student, and how this calculated grade compares to the instructor's qualitative observation of individual success within the course. This method provides a means for differentiating the high performers on a team from the low performers (when a majority of the assignments and assessments are group related) and also eliminates the need for an instructor to provide grades on pure qualitative observation.

Introduction

Capstone courses are the culminating learning experience for many engineering and engineering technology programs. These courses place students in teams with an assigned project that simulates working in an industrial setting; fulfills the ABET requirement of working in a technical team;[1-3] and is an intentional contrast to the lectures, homework and tests which the typical student has experienced over the previous three years. Assessment of the individual student on a technical team is challenging when submitted content is a team effort.[4] Some students in these teams become "hitchhikers"[5] or "free riders"[6] and cause other team members to resent working with these individuals. Peer evaluation or rating is one way that instructors gain insight into the team dynamics that may identify "hitchhikers" or "free riders." Adjusting individual grades based on the evaluations or ratings is agreed upon by many experts to be beneficial and may reduce the resentment of team members for both the instructor and the poor performing team-member.[7]

Numerous articles have been published that discusses peer evaluation and the use of peer evaluations in assessments. A majority of this discussion is found outside of the engineering education literature and only a few are referenced here. One system, developed by Robert Brown at the Royal Melbourne Institute of Technology, measures individual performance in cooperative learning from peer evaluations.[8] This system was designed to have nine discrete measured levels where students rated each other between "No Show" and "Excellent." Numerical values are assigned to each level that are used to define a weighting factor for each individual student. This weighting factor is calculated by dividing the individual's average rating by the team's average rating. If the individual average is equal to the team average, the result will be 1.0. If the individual performs below the team average, the result will be less than 1.0 and if the individual performs above the team average, the result will be greater than 1.0. This method has been used by other researchers in engineering to assign individual grades on team assignments by multiplying the weighting factor by the average team grade.[2, 7] Some researchers indicate capping the weighting factor to limit the increase for the individual higher performers' grades. Typically this system has been used with a single question that students use to assess each of their peers and students are not typically provided with specifics of how the levels are numerically defined and used in the grades.[7]

Other peer evaluation methods use Likert scales with multiple questions. Ohland et al.[2] discussed a method developed for mechanical engineering courses at North Carolina A&T State University. Instructions were provided to each student to rate each team member on 10 characteristics of team work on a 5-point rating scale. The characteristics were all related to attendance, contribution, communication, responsibility, performance, and timeliness. No information was provided about how the scores were used to obtain an individual student grade.

Another well-known peer evaluation tool in engineering education is the Comprehensive Assessment of Team Member Effectiveness (CATME) tool.[9] This tool was developed using National Science Foundation funding and was deployed in 2005.[10] This tool has numerous items over five categories that are used in a Likert-style evaluation. The tool is able to provide instructors with a grade adjustment factor that relates an individual student to the average of the team. The adjustment factor is calculated by dividing the average individual student score by the average of each team member and is used by numerous universities for peer evaluations.

The capstone courses in the School of Engineering Technology (SoET) of Purdue University is typically a two-semester sequence with teams ranging in size based on the needs of the project. Over the course of several semesters at the Columbus, Statewide campus, a grading scheme was developed to assist in grading the individual student in relation to the overall team's success. This method uses a variety of survey questions where each team member rates themselves and their peers over a 10-point scale. These scores are calculated into a z-score that is used as the adjustment factor to provide an individual with a final grade. This paper will discuss the SoET capstone

course and discuss how the z-score is calculated. The paper will also discuss the difference in the z-score and benefits of using the z-score.

Methodology

The SoET capstone sequence is a 2-semester senior-level project course for engineering technology students. Students may be from any of the SoET programs (Electrical Engineering Technology, Industrial Engineering Technology, Manufacturing Engineering Technology, or Mechanical Engineering Technology). At the Columbus, Statewide campus, the students are assigned to teams by the instructors of the course, who are called “managers” to further the industry simulation. These assignments are typically based on the projects available, the background of the students, and the academic accomplishments to date. These assignments mirror industry where managers will place employees into teams in an attempt maximize the output of the team. The projects range from local industry sponsored projects to faculty defined projects related to research interests. Teams typically range from 4 to 6 students.

Learning outcomes for the first semester course are: 1. Apply the engineering design process to engineering technology problems. 2. Evaluate customer needs when defining a solution. 3. Determine a preferred solution. 4. Define and develop a plan for implementing a technical solution. The learning outcomes for the second semester course are: 1. Participate effectively in teams. 2. Use appropriate research and discipline processes to design a solution. 3. Develop a final project report, properly acquiring, using, and citing sources. 4. Identify limitations/improvements as well as strengths of the design solution. 5. Present final solution recommendations. In both courses, grades are determined from group presentations, group reports, and course exams. Rubrics have been developed for the reports and each group member receives the equivalent grade since the report is a group effort. Rubrics have been developed for the presentations as well. A majority of the presentation assessment is identical except for communication skills that can be assessed on an individual basis during the presentation. The exams are peer evaluations that are used to grade the individual contributions of each student.

The peer evaluation questions have been developed over several years to accommodate the students and types of projects that are generally assigned to these students. Eighteen questions have been developed with a set of instructions about how to complete the questions. Two evaluations are given in the course as the midterm and the final exams. The midterm exam uses a subset of the questions that are applicable to the development of the project at that time of the semester. Appendix A shows the most recent instructions and the list of all questions currently developed. The questions have been limited and revised to ensure that the instructions and questions are placed on a single sheet that is distributed to the students during the exam. The questions review different aspects of a generic capstone project and include question related to teamwork and project deliverables.

The completion of an exam is an individual effort and the individual results are not shared with the team. A 10-point scale is used for the evaluation where 1 is the lowest

value and 10 is the highest value. The larger scale was used to provide more variation in the responses over the typical Likert 5-point scale. To mitigate students assigning all team members with the same grade, the instructions force the students to have a difference of at least 4 points over all team members related to a single question. This instruction forces students to provide some variation in the scores and therefore consider how each team member functioned with respect to that question. The other major instruction for the exam stipulates that a single student may not receive the same score for all questions. These instructions work well with the z-score because some variation is mandated by the instructions and the z-score can use the information to differentiate how individual students are performing.

The z-score, or equivalently, z-statistic, is a standardized score that measures the number of standard deviations the observed value is away from the overall sample average. Therefore, a z-score of 0 is average, a z-score of -1.0 is a standard deviation below the average, and a z-score of 1.0 is a standard deviation above the average. The z-score (z) is found by

$$z = \frac{X - \mu}{\sigma} \tag{1}$$

where for this process, X is the score, μ is the sample mean, and σ is sample standard deviation.[11] Equation 1 is used with each student's evaluation to compute a z-score for themselves and all of their team members based on the individual student's 1-through-10 evaluation. Each team member's evaluation of themselves and their teammates for each question is then replaced by a z-score. This method allows the z-scores to be standardized for that individual student based on the sample standard deviation of the individual's response. Therefore, if a student uses a larger range of scores, that student's calculated z-scores will not be influenced by another student's evaluation that only uses the minimum range of scores. Once the z-scores are calculated, each team member's evaluation of an individual (themselves or each of their teammates) is the average of the z-scores for that individual for all the exam questions.

Table 1 shows a simulated example of the z-score calculation of an individual for a 3-member group with 3-questions. The raw scores for the self-evaluation are in the third column of the table, the raw scores for the other two team members are in the fourth and fifth columns respectfully. The average scores for all of the questions for each team member are in column 6 with the average and sample standard deviation of those scores in the same column. Equation 1 is used with each raw score, the overall average, and sample standard deviation to produce a z-score for each question in columns 8 through 10. These z-scores are averaged to produce the average z-score for all of the questions for each student in the last column. In this simulated example, the evaluating student (A) rated themselves 1.0 std over the average, rated student B 0.873 std below the average, and rated student C 0.218 std below the average for all of the questions. For each student evaluator, their raw scores are used to create a similar table of z-scores like the scores shown in Table 1 for student A.

Table 1. Example z-score calculation for a simulated evaluation of 3 questions with 3 team members.

Reviewer	Question No.	Team Member	Team Member	Team Member	Raw Data	Avg.	Z-Score	Z-Score	Z-Score	Z-Score Eval	Avg.
A		Self	B	C	Avg. A	8	Self	B	C	Avg. A	1.09
	1	6	9	5	Avg. B	7	-2.84	3.06	-4.80	Avg. B	-0.873
	2	10	6	7	Avg. C	7.33	5.02	-2.84	-0.873	Avg. C	-0.218
	3	8	6	10			1.09	-2.84	5.02		
					Average for All	7.44					
					Std. Dev for All	0.509					

Table 2. Average individual z-score calculation for a simulated evaluation of 3 questions with 3 team members.

Evaluator	**Evaluation**		
	Team Member A	Team Member B	Team Member C
Team Member A	1.09	-0.873	-0.218
Team Member B	0.524	0.112	-0.636
Team Member C	0.642	-0.987	0.345
Average	0.752	-0.583	-0.170

Summary Note: The final score for this exam for team member A is 0.752, for B, -0.583, and for C, -0.170. Team member A's score is relatively high based on experience, indicating that she/he is a relatively strong contributor. B's score suggests that she/he does the minimum amount of work needed to get by.

Once each students' z-scores are calculated, the results are tabulated as shown in the simulated results of Table 2. Here, the simulated results of team member A, from Table 1, are shown as well as results for team members B and C (not shown). Once the z-scores from every member are calculated, the average of those individual calculations shown as the average scores in Table 3 are the z-scores that are used in calculating the exam grades. Each team is assessed and provided an average team exam grade. This team exam grade is an assessment from the instructors on how well the team is completing the project deliverables. The z-score is used to increase or decrease an individual student's grade from the average team grade. The authors are currently use a ±5% to minimize the weighting of the z-score as the authors assess how the evaluations are related to the observation assessments of individual contributions to the team's success.

Results

The instructors of the course compared the individual z-scores to the observed student success on the project teams. The student success was observed from the group presentations. Teams typically assign individuals the assignment of discussing their specific contributions to the project. Perceived "hitchhikers" were typically given the introductory and outline information and did not present information related to the deliverables. The students that the instructors hypothesized from this observation as "hitchhikers" typically received a z-score around -1.0, or one standard deviation below the group average. These scores provided the instructors with a measurable method to justify student grades for substandard performance in a team setting.

An interesting result from reviewing the z-scores scores indicates that students typically rate themselves higher than their ratings from their peers. Almost 60 percent of the student rated themselves higher than their ratings from their peers where the largest difference were over 1.5 standard deviations. Some of the "hitchhikers" did acknowledge in the evaluations their lack of effort, but the majority rated themselves around average or in some cases significantly above average. The high-performing students typically rated themselves similarly to their peers or assessed themselves below their peers' assessments.

Discussions/Conclusions

The instructors have been using this technique since the spring semester of 2019 and have been working on the questions and how to implement the results into the students' grades. Different variations of how to implement the z-score as a weighting factor have changed during this time. The initial implementation started all team exam scores at 75 percent and used the z-scores to increase or decrease the individual scores. This process was switched to an assessed team exam score because high performing teams with significant contributions from all team members were penalized for not having a "hitchhiker." The current team exam with a ±5% grade appears to be too narrow for "hitchhikers" and provides them significantly higher grades than observations indicate. Future changes are being considered and current scores will be used to model potential scores based on brain stormed changes.

The majority of the referenced literature does not force variation in the peer evaluation. As stated, here we force students to leave at least a 4-integer variation on all questions. Prior to this instruction, the majority of students would only give scores in the 8, 9, and 10 range. This range appears to be implemented on the typical 100-point grading scale where all students want to keeps scores in the A and B range. This peer assessment was also observed for questions where students did not actively participate in that part of the project. Once the 4-integer variation was mandated, the instructors noticed that students were using a much larger range and, in some instances, using the entire 1-10 range.

Another consideration, that most of the weighting factors reviewed in the referenced peer evaluation literature does not consider is the range variations used by the students. The z-score is a standardized score that uses the individual's assessment to normalize their scores before averaging the individual group scores. The authors need to review the reliability and validity of this developed method. The authors also need to use some of the published peer evaluation methods to compare the peer results of this method to those methods.

In conclusion, the presented z-score peer evaluation method measures "hitchhikers" in the team environment, but also shows very high performing students on a team. A number of "hitchhikers" tend to rate themselves higher than their peers' evaluations of the "hitchhikers", which is another indication of their role in the team. This method provides a means for instructors to specify an individual's grade from accomplishments as viewed by their peers, when an instructor may be unable to assess the individual contribution. The authors are continuing to review the results of this method and look to improve the method.

Appendix A: Exam Questionnaire

The following is a sample of an exam that has been given to SoET capstone students. In the case of this exam the last student finished around 25 minutes after the test was distributed to the class.

Instructions: Evaluate and grade your team according to the questions on the next page using an integer scale of 1 (= lowest/least) to 10 (=highest/most). Write the names and initials of your team mates below and put the initials of your teammates next to their names and at the top of the columns in the grading matrix on the next page.

This is to be an individual effort. You might want to fill scores out in pencil to allow for adjustments.

Consider your grading carefully. Blanket scores, like all 10's for your team members for a question, will result in your score being a one (1). Ungraded boxes or boxes with unreadable names or grades on your evaluation will result in your score being one (1). The *difference* between the highest and lowest grades *on any row* should be *at least four (4)*, for example, 10, 9, 8, 7, and 6. Another way of saying this is that your grading should span at least five numbers for any question.

No column of the grading matrix on the next page shall be *all the same number*. One way to evaluate this that might be useful to you is to assign grades to the top/most and bottom/least first, and then fill out the folks who are in between.

In the initials column below, please use two (2) initials (first and last name; yes, we have had the case where a student entered only one).

If you wish to see how your "management" judged your team's effort, which provides the center-point percentages for your grade, complete the first page, initial the second, and return your test to the instructor who will fill out the center-point score on the next page. This routine is for FERPA reasons.

Print Your Team Member's Name	Team Member Initials

No.	Question	Self	Initials	Initials	Initials	Initials
1.	Would you want to work on a project with this person again? Use your grading scale to indicate the degree to which you would want to work with this person and **do not grade yourself on this question.**					
2.	This person did the most work on reports. For this, and all other "This person …" questions, use your high value to indicate 'most' and your low value to indicate 'least'.					
3.	This person did the most work on presentations.					
4.	This person was most creative or inventive on our project. They came up with one or more ideas that were adopted by the team.					
5.	This person provided leadership for our team.					
6.	This person did most of the presentation graphics on our team.					
7.	This person did most of the design sketches/drawings on our team.					
8.	This person did most of the hardware creation and assembly work for our team.					
9.	This person did most of the software creation for our team.					
10.	This person did the most library/online research for our team.					
11.	This person did the most in-lab test work for our team.					
12.	This person did most of the data reduction work on our team.					
13.	This person was helpful to the other members of the team.					
14.	Did the person contribute their fair share, in terms of hours, to the work that has been done so far, including research, prototype work, part models, experiments, presentations, and written assignments?					
15.	Do you trust this person to do what they say they will do by (nearly) when they say they will do it?					
16.	Did the person volunteer for any part of the project? Use your grading scale to indicate the frequency with which the person volunteered for work.					
17.	Was the work done by this person of high quality? That is, did the work done by this person meet your minimum quality norms? Or did the work performed by this person need significant revision to meet your expectations?					
18.	Did the person keep you informed of their status and progress of their work in sufficient time to reprioritize work or whatever other steps were needed to meet unexpected problems or delays? **Do not grade yourself on this question.**					

Your Team's Center Point Score:

References

(1) Ohland, M.; Loughry, M.; Carter, R. L.; Yuhasz, A. G. Designing a peer evaluation instrument that is simple, reliable, and valid. In *2004 Annual Conference*, 2004; pp 9.392. 391-399.392. 311.
(2) Ohland, M. W.; Layton, R. A.; Loughry, M. L.; Yuhasz, A. G. Effects of behavioral anchors on peer evaluation reliability. *Journal of Engineering Education* **2005**, *94* (3), 319-326.
(3) *Criteria for Accrediting Engineering Technology Programs, 2020 – 2021*. 2021. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-technology-programs-2020-2021/ (accessed Janurary 31, 2022).
(4) Somerton, C. W. Using Student Evaluations for Individual Grading in Team Projects. In *2000 Annual Conference*, 2000; pp 5.702. 701-705.702. 709.
(5) Kaufman, D. B.; Felder, R. M.; Fuller, H. Peer ratings in cooperative learning teams. In *1999 Annual Conference*, 1999; pp 4.416. 411-414.416. 412.
(6) Dingel, M. J.; Wei, W.; Huq, A. Cooperative Learning and Peer Evaluation: The Effect of Free Riders on Team Performance and the Relationship between Course Performance and Peer Evaluation. *Journal of the Scholarship of Teaching and Learning* **2013**, *13* (1), 45-56.
(7) Kaufman, D. B.; Felder, R. M.; Fuller, H. Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education* **2000**, *89* (2), 133-140.
(8) Brown, R. Autorating: Getting individual marks from team marks and enhancing teamwork. In *Proceedings Frontiers in Education 1995 25th Annual Conference. Engineering Education for the 21st Century*, 1995; IEEE: Vol. 2, pp 3c2. 15-13c12. 18 vol. 12.
(9) Ohland, M.; Pomeranz, H. R.; Feinstein, H. W. The comprehensive assessment of team member effectiveness: A new peer evaluation instrument. In *2006 Annual Conference & Exposition*, 2006; pp 11.1262. 1261-1211.1262. 1266.
(10) Loughry, M. L.; Ohland, M. W.; Woehr, D. J. Assessing teamwork skills for assurance of learning using CATME team tools. *Journal of Marketing Education* **2014**, *36* (1), 5-19.
(11) Tokunaga, H. T. *Fundamental statistics for the social and behavioral sciences*; Sage Publications, 2018.