AC 2012-5186: EVALUATION OF DESIGN WORK AND THE ACHIEVEMENT OF LEARNING OUTCOMES IN SENIOR CAPSTONE COURSES

Prof. Carl D. Sorensen, Brigham Young University
Dr. Robert H. Todd, Brigham Young University

Robert H. Todd is a professor of mechanical engineering at Brigham Young University and the Founding Director of BYU's capstone program Integrated Product and Process Design. Todd received his Ph.D. from Stanford University in mechanical engineering design, where he also received a postdoctoral fellowship, taught engineering courses, and served in department and college administration at BYU, Idaho (then Ricks College), before spending 10 years in industry in senior engineering and management positions with both the General Motors Corporation and the Michelin Tire Corporation in both the U.S. and Europe. His research and teaching interests include hybrid and hybrid/electric vehicles, manufacturing process machine design, product design, and the improvement of engineering education in a global setting. Since coming to BYU in 1989, he has been the senior author of two manufacturing processes books. One, Manufacturing Processes Reference Guide, is a best seller used thought out the world. He has also authored more than 100 technical articles with his students. He has served as a Department Chair and Undergraduate Coordinator at BYU and a member of the Engineering Accreditation Commission (EAC) of ABET (the Accreditation Board for Engineering and Technology). He is a recipient of BYU's Karl G. Maser Excellence in Teaching Award, the Mechanical Engineering Department's Outstanding Teacher and Rudy Awards, the Department's Outstanding Achievement Award, and BYU's Blue Key College of Engineering and Technology Outstanding Faculty award. He is a member of ASEE, ASME, SAE, SME, and a Fellow of the American Society for Engineering Education.

Taylor Halverson, Brigham Young University

Taylor Halverson earned a double major Ph.D. at Indiana University in instructional technology and design and Judaism and Christianity in antiquity. He earned master's degrees from Indiana University and Yale University. His bachelor's degree was earned at BYU. Halverson spent several years working for Cisco in Silicon Valley, where he designed creative learning experiences for thousands of customer service agents spread across the globe. Halverson currently works as a Teaching and Learning Consultant at BYU, assisting faculty members to enhance the student learning experience. He also is a part-time faculty member at BYU, teaching a variety of courses including "The History of Creativity in the Arts, Sciences, and Technology," and a part-time faculty member at Capella University, teaching online Ph.D. learners in instructional technology and design. Halverson regularly presents at academic conferences and has a published a book on instructional design theory and practice.

Evaluation of Design Work and the Achievement of Learning Outcomes in Senior Capstone Courses

Abstract

The two-semester Mechanical Engineering Capstone course at Brigham Young University (BYU) was created in 1990 to help students learn a structured design process and assist them in developing design skills for the practice of engineering. Course outcomes were established based on stakeholder input and students receive classroom instruction as well as do project design work on projects provided by industry sponsors. To date more than 575 design and build projects have been completed for more than 300 industry project sponsors with more than 3500 mechanical, electrical, manufacturing and other students working in multidisciplinary teams.

One of the more significant challenges in offering a course of this type is developing a method to accurately evaluate student work. Students are typically very interested in their grades and faculty are interested in accurately and fairly assessing student work as it relates to the achievement of course outcomes by students. Fair and accurate assessment can be particularly challenging when much of the student work is performed working in teams.

This paper presents a brief literature review on the topic of assessment and evaluation of engineering design work by students, a set of three historical evaluation schemes used in BYU's Capstone, including the one now in use, and relevant observations and conclusions we have drawn through the years.

Introduction

During the fall semester of 1990, BYU initiated a two-semester design course entitled, Integrated Product and Process Design.[1] The main objective of the course was to strengthen perceived weaknesses in new graduating engineers. The course includes classroom instruction, individual assignments, an examination, and project work by student teams performed on industrially sponsored design and build projects. Essentially the course involves students learning how to integrate design skills in four major areas: technical analysis, design creativity, project management and teamwork.

Each year approximately 180 students take the course from a variety of majors including mechanical engineering, manufacturing engineering technology, electrical engineering and a variety of other majors. Students are divided into cross-functional teams and typically about 30 industrially sponsored projects are recruited from industry for student teams to work on.

Each team is assigned a faculty coach to mentor the students in learning how to apply the design process taught in the course. About half of the faculty coaches come from full-time faculty where coaching is considered part of their normal teaching load and the other half are practicing engineers or engineering managers from local industry, who are hired as part-time faculty. One reason for soliciting real projects from industry and in hiring part-time faculty from industry is to bring industry and academia together in a partnership to improve engineering education.

Through the years, the course has received a number of awards and positive feedback from both industry and academia. Approximately 3500 students have completed the course.

The instructional design for the class includes both general objectives and specific educational outcomes. The general objectives are the hardest to measure effectively when evaluating students; therefore this paper focuses on evaluating the accomplishment of these objectives. The general objectives for the current year are shown in Table 1.

The specific course outcomes for the first and second semesters of the Integrated Product and Process Design course are shown in Table 2 and Table 3, respectively.

Table 1: The general educational objectives of the Integrated Product and Process Design program at Brigham Young University

The overall objective of Capstone is to help each student become a successful, practicing professional who can make a positive difference in the world. To achieve this, each student is expected to learn to:

1. Understand and apply a structured design process to complete a design project that adds real value to a project sponsor.
2. Understand and apply principles of project management and scheduling to ensure that the right work is accomplished at the right time.
3. Integrate the knowledge and skills developed over the course of prior education and experience to achieve high-quality engineering designs that meet customer needs.
4. Participate synergistically as a team member, whether leading or following, in order to help the team succeed at the highest level.
5. Grow personally and professionally, taking the responsibility to learn and work independently; seeking outside help, advice, and feedback as needed to complete the design project.
6. Work hard on a challenging project, in spite of the difficulties that will arise, and couple that work with faith to accomplish an outstanding solution.

Assessment of student work in the course is based on three areas of evaluation:

1. Individual learning of course material
2. Assessment of team project work, and
3. Individual contribution to the team

Evaluation in a design class differs from that in “traditional” engineering classes. In traditional classes, much of the grade is based on the student’s ability to solve problems that generally have one right answer, and the grade can be quite objective. By contrast, in design work, there is no unique correct answer, so that evaluation is necessarily subjective. The quality of an individual’s or team’s work is evaluated by external parties. Different external parties may have different opinions about the quality of the work. In such situations, students can feel that the evaluations are arbitrary or capricious, and the evaluations then may be seen as invalid or not helpful.

Table 2: Specific course outcomes for the first semester of Integrated Product and Process Design

1. Describe and apply the steps of the concept development phase on the design process, including:
    a. Establish a project objective statement as a framework for the project.
    b. Identify customer needs, develop target specifications that assure meeting the needs, and determine test criteria and procedures to determine whether the specifications are met.
    c. Generate multiple concepts for meeting the design objectives, and use a structured process to select one or two concepts for further testing.
    d. Test product concepts to further refine the concept selection, create a preliminary design of the concept in an assembly drawing, and establish final specifications.
    e. Select, build, and test physical and analytical prototypes to answer questions about the performance of the concept.
    f. Develop a work breakdown structure and a dependency chart for a design project.
    g. Justify a project financially using fundamental engineering economics principles and practices.
2. Demonstrate an understanding and ability to use effective team processes, communication, and conflict resolution skills to enhance synergy on a project team.
3. Working with a project sponsor in a team setting, apply the design process from outcome 1 to achieve superior project results, as demonstrated in a semester-end project review.
4. Contribute to the team's production of professional-quality written reports and oral presentations to the various project stakeholders.
5. Accept, evaluate, and act upon feedback given in design reviews in order to improve the final outcome of the project.

A Student Guidebook has been created and modified through the years, which is essentially a syllabus for the course.[2] The Guidebook contains five sections including: *Getting Started, Schedule, Assignments and Evaluations, Policies and resources, and Record Book.* The Guidebook contains about 75 pages making up the first four sections of the Guidebook, plus 100 pages in the fifth section which are numbered blank pages, using a quad ruled format, which make up the Record Book section of the Guidebook. These pages are used by the students to record notes for their project as they work on their project over the course of the two semesters. The Assignments and Evaluations section of the Guidebook explains the current grading scheme used in BYU's capstone course and provides information so that students can track their grade and understand the process and its rationale.

This paper describes some of the attempts we have made over the last decade and a half in adjusting our evaluation practices to improve the quality of student learning in our capstone class. It begins with a review of some of the relevant literature, describes our objectives in evaluation, presents some of the options we have tried in the last 15 years, and shares observations about the effects of the various grading practices. We believe that these observations provide an important contribution to the ongoing discussion of evaluation of design work and the improvement of engineering design education.

Table 3: Specific course outcomes for the second semester of Integrated Product and Process Design

1. Describe and apply the steps of the detailed design phase of the design process, including:
    a. Develop a product information package containing the detailed information necessary for producing the final project, including assembly drawings for the final product and subassemblies; part drawings for custom parts; purchase information for vendor supplied parts; schematic, piping and/or wiring diagrams for electrical, pneumatic, or hydraulic systems; block and/or logic diagrams for software systems; and similar materials necessary for unique projects.
    b. Develop and carry out statistically-based experimental plans for determining the values of critical design parameters, using focused physical or focused analytical prototypes.
    c. Perform a basic FMEA analysis to identify critical risks that should be addressed in a project.
    d. Manage the complexity of the detailed design phase through the use of a Work Breakdown Structure and project schedule based on that structure.
    e. Create a final prototype that reflects, to the extent possible, the product information package.
    f. Perform and document acceptance tests that determine how well the final prototype meets the final specifications developed in the concept development phase.
2. Check an engineering drawing, identify weaknesses, and recommend changes to bring the drawing up to professional standards.
3. Describe the fundamental principles of contract law.
4. Describe the protections, limitations, and methods for using different means of protecting ideas: trademarks, copyrights, patents, and trade secrets.
5. Working with a project sponsor in a team setting, apply the design process from Outcome 1 to achieve a superior project outcome, as demonstrated in a final project review.
6. Contribute to the team's production of professional-quality written reports and oral presentations to the various project stakeholders.
7. Accept, evaluate, and act upon feedback given in design reviews in order to improve the final outcome of the project.

Literature Review

The fundamental principles of assessment involve identifying measurable student outcomes and then devising methods to measure how well each student has achieved these outcomes. This sounds simple in principle but in practice it is often much more difficult. Using a structured design process to develop a new product is often easier than designing a method to accurately and fairly measure how well students have learned principles of a design process and learned how to apply them creating a successful, real-life project.

Literature from the field of assessment in Higher Education suggests that frequent assessment with informative and "learning-centered" feedback promotes and improves student learning.[3] Principles, methods and recommendations for good assessment that are accepted and used today include the actual student performance of the skills they have learned in the context of projects, group work, research, and exams.[4,5,6] The main challenges that exist are in aligning learning outcomes with meaningful and measureable assessment data points that inform student learning, help students improve their learning, and enable program improvement. An additional challenge to this is in defining useful rubrics that can be used to assess and judge student performance on project, design, writing, or other performance assessments.[7,8,9]

Bailey and Szabo declare "Rigorously assessing students' design process knowledge is essential for understanding how to best create learning environments to facilitate the development of such knowledge. Such assessment is also quite difficult and hence there is a lack of assessment tools capable of measuring the design process knowledge of every student in a large college".[28] This statement concisely depicts the wide open field of opportunity for engineering educators to devise, validate, and publish assessment instruments targeting design skills and knowledge.[35, 36] Some efforts have been made to address this opportunity through the development of standardized tests of engineering design skills.[10,11,12,13]

Nevertheless, standardized approaches to assessing engineering design skills are the exception. Engineering and technology education literature on the topic of assessing engineering design skills and knowledge parallel what is currently being said for assessment in Higher Education generally. The most common forms of assessment of engineering design skills and knowledge cited in the engineering education literature are proposals, design reports, final reports (oral and/or written), presentations, competitions, student self-assessments, sponsor/client evaluations, status reports, peer evaluations, portfolios, design notebooks, prototypes, expert review by faculty or professional engineers, quizzes and exams,[14-24] and the judicious development and use of rubrics.[25-29] There have been mixed successes with these forms of assessments as documented in the literature cited. Furthermore, the pros and cons of some of these assessment methods are described by Bailey and Szabo.[30]

One of the challenges of evaluating student work in team projects is the difficulty of fairly evaluating the performance of individual team members. The course instructors who are expert in the subject matter have less knowledge of the individual student performance. The students who have the most knowledge of the differences in performance are the less expert in assessment. Given this dilemma, Kruck and Reif identify five possible methods for evaluating individual performance and recommend a mix of a group grade and peer evaluation.[31]

Various methods used for calculating individual grades from team and individual assessment have been summarized in the reviewed literature.[32-34] We have tried the most commonly cited methods, as will be described in the body of the paper.

Our approaches to assessing engineering design skills are of the forms mentioned above. But what have we done to improve these assessments to increase the chances of a better learning and evaluation experience? This article highlights modifications to typical engineering design skills assessments in terms of implementation, grading schemes, and motivational apparatus in order to improve the likelihood of learner success and the assessment viability of capturing learners' design skills.

Objectives

Evaluations in BYU's engineering Capstone course are based on individual and team performance and project results that meet the project sponsor's needs. Our assessment method strives to evaluate each student and project team fairly and to provide the best possible feedback to help students improve in their learning effort *and* project success. Three major factors with which we have struggled over the years include the need to emphasize the importance of individual learning, fairly evaluate the work of the team, and provide fair feedback and grades for the individual contributions of the team and its members. In addition, we have found that the mechanics of grading can affect the attitude and performance of the students.

Emphasize importance of individual learning

Capstone is not just a class to complete a sponsored design project. Capstone is a class to *learn a design process* and to *learn to apply the design process* to a sponsored project. Both aspects of learning in the class should be measured by the assessment and grading process.

In developing the outcomes for the course, certain capabilities of the students are described. As part of our requirement to educational integrity, we should do our best to evaluate to what degree individual students have achieved the desired capabilities, independent of whether the team may have demonstrated these capabilities.

Fairly evaluate the work of the team

One of the challenges of capstone grading is the need to fairly evaluate the work of each team in the class. Different teams have different projects, with different degrees of difficulty. They have different sponsors, who may provide more or less help on the project. With all of the variation between teams, it can be a challenge to evaluate the work of each team fairly.

In design, as in life, excellence requires going beyond what is "expected". Excellence involves work that goes beyond the norm, delighting and inspiring those for whom the work is performed. Often the customer cannot define in advance what would make the work excellent, but they recognize excellent work when they see it. Excellence requires creativity and initiative. Excellence is more than just avoiding mistakes; it requires creatively finding ways to exceed the expectations of the customer.

Grade fairly team members' contributions

The ability to work effectively in teams is integral to the Capstone course outcomes. The best way to measure this ability is to measure the performance of the team on which the students are working. All students share responsibility for the performance of the team, so the performance of the team must affect the individual's grade.

In an ideal world, team members would all contribute equally to the project. In a real world, the contributions are sometimes unequal. Sometimes a team member decides to do the bare minimum to get by. Other times a team member decides to go beyond the typical team performance to ensure a successful outcome. In both cases, the team evaluation should be adjusted when assigned to an individual. A team member should not be able to get an excellent grade by freeloading just because he or she may be part of an excellent team.

Grading Mechanics

Students should be able to tell at any point in the semester how well they are doing in the class. Further, heroic effort at the end of the semester should not make up for lackluster effort at the beginning of the semester. The grades should be cumulative, with early grades combining to give a final grade.

The idea of giving periodic grades throughout the semester is not universally accepted. Some stakeholders argue that the final result is all that matters in a design context, and that the final grade should be based solely on the final design results. We believe that the final results are important, but so is the work leading to the final results. By evaluating the work throughout the project, we encourage consistent, high-quality work and the development of increasing competence in our students.

Students, coaches, and instructors should be able to understand the grading algorithm and the effect of individual scores assigned. Once the individual scores are assigned, there should be no surprises in the final grade received. Combining individual and team grades fairly will require some complexity, but the complexity should be managed as well as possible.

Identifying Appropriate Evaluators

It is important to identify appropriate evaluators for completing the evaluation. The evaluators should be chosen to be those who have the most understanding, and ideally, the least bias concerning the learning obtained and work performed by individuals and teams. In general, this means that there will likely be different evaluators for different aspects of the capstone experience.

Because the instructors see the learning of all of the students in the class, they are best equipped to ensure that individual learning grades are fairly allocated.

Because coaches are intimately aware of their team's activities, they can effectively assign credit for how well the activities of the team match the needs of the project.

Coaches generally have experience with a wide variety of teams, and are well-equipped to evaluate the performance of individual team members in this context. Peer team members generally have good knowledge of the contributions of the other team members, because they have worked the most closely with them. Together, evaluations by coaches and peers should be able to fairly evaluate individual team member's contributions.

Instructors see the broad range of projects, and are thus well-equipped to evaluate the relative performance of the different teams. Coaches of other teams can compare the resulting work of the team being evaluated with the work of their own team, and are thus in a good position to evaluate relative team performance as well.

Options

Since our capstone course at BYU was created more than two decades ago, a number of different grading schemes have been tried. Each has demonstrated strengths and weaknesses, and minor refinements to each of the fundamentally different schemes have been tested from year to year. In reviewing all of the evaluation schemes that we have explored, they can be largely classified into three schemes: The dominant project grade scheme, the team grade plus individual grade scheme, and the individual grade plus multiplier on team grade scheme. Each of these general schemes is described below.

Dominant project grade

In the dominant project grade scheme, the effect of the teamwork on the project dominates the student's grade. Typically about 2/3 of the grade comes from the project, and 1/3 comes from individually evaluated assignments and tests. Generally, the 2/3 of the grade on the project is assigned by the coach, since the coach is most familiar with the student's individual work on the project.

In this scheme, the coach assigns each team member an individual project grade. The course instructors discuss their perceptions of the overall team performance, but the coach has the freedom to assign grades irrespective of the instructors' observations.

In the dominant project grade model, feedback is given on the project deliverables (such as reports, drawing packages, analysis results, etc.), but there is no formal calculation of the semester grade based on the deliverable grades. This decision was made because overall performance on the project, rather than work on individual components of the project, is considered to be the key aspect of this grading scheme.

Table 4: Representative grading scheme for a semester under the dominant project grade model

Item	Evaluator	Weighting
Participation	Instructors and Reviewers	5%
Exams (2 exams equally weighted)	Instructors	10%
Early Design Exercise	Instructors	4%
Self-Education (12 documented hours required)	Self-evaluated (completed or not)	6%
Progress Reports (5 reports equally weighted)	Instructors	10%
Mid-term team and project performance	Project Coach	20%
End-of-semester team and project performance	Project Coach	45%

The dominant project grade model puts the responsibility for assigning most of the student's grade on the project coach. In some respects this is desirable, as the coach is most familiar with the team member's individual work. However, the coach is less familiar with the work being done in all of the project teams. This makes it difficult to assure grading consistency throughout all of the teams that makeup the class. In addition, the coach faces a conflict of interest, as he or she must be both an advocate for the team and an evaluator of the team's performance.

Another weakness of this grading model is the lack of an explicit connection between the individual's performance on the team, the team performance on the project, and the individual's grade.

Team grade plus individual grade

In this grading scheme, there are three components that affect an individual student's grade. As in the dominant project model, there is an individual grade given on learning activities related to the class. However, unlike the dominant project grade model, there is also a specific project grade that is assigned to a team and given to all members of the team. In addition, individual grades are assigned by the coach to reflect the individual's contributions to the project.

In this grading scheme, the results obtained by the team are graded by individuals outside the team. Final project reports are graded by the coaches of three other teams, and final project review is graded by the course instructors.

The performance of the team and the individuals on the team are graded by the coach, in the form of team and individual process steps.

Table 5: Representative grading scheme for a semester under the team grade plus individual grade model

Item	Evaluator	Weighting
Assignments (3 equally	Instructors	15%

weighted)		
Product and Process Development Exam	Instructors	18%
Team Process steps	Coach	17%
Individual process steps in pursuit of team progress	Coach	16%
Semester final report	Three other project coaches	21%
Semester project review	Instructors	12%

The team grade plus individual grade model addresses some of the weaknesses of the dominant project grade model. It contains an explicit grade element for the individual's contribution to the team. It uses other project coaches as well as the class instructors to assign part of the team project grade. It provides explicit evaluation of key project deliverables. In our experience, this model greatly increases the reliability of the resulting grades.

Nevertheless, there are two major weaknesses of this grading model. First, the individual project grade is assigned solely by the coach. This is of concern because the coach may be less aware of the individual student's performance than the student's fellow team members. It is possible for students to fool the coach by looking better than they really are.

The second major weakness of this model is that it allows a student to earn an undeserved grade by taking advantage of the work of other team members. Consider a student who works very hard on the assignments and exam, earning 100% on these items. Suppose the team does very well on the team process steps, final report, and project review, earning 100% on these items. Finally, suppose a student earns only 50% on the individual process steps, which is a failing grade in this area. Such a student would receive a grade of 94%, or an A in the class, even while failing in their contribution to the team's success. In our opinion, a team member who is making only marginal contributions to the team should not receive full credit for the team's performance.

Individual grade plus multiplier on team grade

This grading scheme is closely related to the Goldfinch method.[34] It can be considered an extension of the team grade plus individual grade scheme. As in all of our grading schemes, a fraction of the grade is based upon the student's performance in individual learning. As in the dominant project grade scheme, part of the grade is based on the coach's assessment of the overall performance of the team. As in the team grade plus individual grade scheme, other coaches evaluate the team's final report and assign a grade, and the instructors assign a semester project review grade.

However, in this scheme, rather than assigning an individual grade for team performance, the coach assigns an individual contribution multiplier that describes the performance of the individual student relative to the norm of the team. If the student is making higher than average contributions, the multiplier is greater than 1.0. If the student is making lower than average contributions, the multiplier is less than 1.0. The individual grade for team performance is then calculated as the product of the individual contribution multiplier and the team performance grade.

This scheme addresses the two major weaknesses of the team grade plus individual grade model. The evaluation of the individual's contribution to the team is equally weighted between the project coach and the peer team members. This gives both a say in the individual contribution, and requires both types of evaluators to agree to make a strong difference in the individual's grade. In addition, this model prevents a minimally-involved team member from benefiting from the strong work of the team. This scheme also tends to foster more communication between coaches and their team members which can enhance feedback and learning in how to apply the design process to a meaningful project.

Consider a student who gets 100% on all assignments and is a member of a team that gets 100% on all evaluations, but who receives a 50% evaluation from the coach and team members on the individual performance. Such a student would receive a grade of 66.5%, a D. In reality, 100% scores for a semester in any area are virtually nonexistent. A student scoring 95% on individual assignments, 50% on team contribution, and 95% on team performance would receive a semester grade of 63%, a D-. A student scoring 90% on individual assignments, 50% on team contribution, and 90% on team performance would receive a semester grade of 59%, which is a failing grade.

The main difficulty with this grading scheme is its complexity. Students are unfamiliar with the concept of an individual contribution multiplier, and are unsure exactly how to assign a number. They are much more likely to think in terms of a grade for each individual team member. However, allowing students to directly assign a grade for their peers decouples the grade from the team performance. Therefore, we believe this complexity is necessary, although undesirable. We continue to look for less complex ways to achieve the desired outcome.

Table 6: Representative grading scheme for a semester under the individual grade plus multiplier on team grade model

Item	Evaluator		Weighting
Assignments (3 equally weighted)	Instructors		15%
Product and Process Development Exam	Instructors		18%
Team Process steps	Coach	24%	
Semester final report	Three other project coaches	20%	
Semester project review	Instructors	15%	
Team meeting class expectations	Instructors	8%	
Total team score			67%
Individual contribution multiplier	Coach and teammates	0 to 1.2 (not a percentage; average for the team must be <= 1)	
Total student grade			Assignments + Exam + (Individual contribution multiplier) * (Total team score)

In order to help students deal with the complexity of this grading scheme, we email a Microsoft Excel spreadsheet to each student. This spreadsheet includes the coach grade for the team process steps, the names of each team member, and protected cells containing the formulas used to calculate the grade. Students can then assign an individual contribution multiplier for each team member. The estimated team contribution score for each team member is displayed. The student can adjust the contribution multiplier scores until a set of scores that is deemed fair by the student is achieved. The students and coaches also have access to a description of Capstone Evaluation Criteria in the Capstone Guidebook, which they can refer to in assisting them in using this evaluation method.

Observations

As we have used various evaluation and grading schemes during the course of the capstone program, we've learned a few things about grading. Most of our learning is largely anecdotal, because we have been unable to find good statistical methods of evaluating our grading systems. However, we have made use of extensive interviews with faculty members and students in an effort to improve our grading methods. The following sections describe some of the consistent observations we have made. We begin with a discussion of some of the continuing challenges, followed by a summary of some of the successes we have observed.

Challenges

We have noticed three major challenges with the current grading scheme. First, there can be confusion about what a grade means. Second, peer evaluations often tend to be strongly nonlinear. Finally, there is an opportunity and sometimes a tendency for students to punitive in their peer evaluations.

Confusion about what a grade means

Perhaps not surprisingly, we have had consistent challenges with getting our students to understand that an A in capstone requires work that goes beyond just being correct. The assumption that a grade starts at 100 and is marked down for every mistake seems to be strongly embedded in the student psyche. Although we have taken great pains to indicate that mistake-free performance is consistent with a B+, students continue to ask what they did wrong to get a B+. They also ask what they need to do to get above a B+, hoping that they can get specific guidance as to what is wrong that can be fixed. In spite of our best efforts to teach that grades above a B+ require excellence, it is difficult to make this paradigm change.

A second area of confusion is the meaning of the individual contribution multiplier. Because the average of the contribution multipliers for a team must be 1.0, students feel like they must give somebody a C in order for other team members to earn an A. We continually work to reinforce with both general statements and specific examples that if the team earns an A, everybody on the team can earn an A if the multipliers are 1.0. By the end of the class, students understand; but at the beginning this paradigm is often not clear to the students.

Nonlinearity of peer evaluations

An ongoing challenge is the difficulty of getting students to be honest and perceptive in the evaluation of their team members. Even if they feel that a particular team member is not pulling their fair share of the load, student evaluations are likely to be close to 1.0. If a team member is doing virtually nothing, then evaluators will give an extremely low score. So there is a tendency for peer evaluations to fail to give an accurate measure of individual performance.

This challenge is especially pronounced early in the first semester. During the forming and norming stages of team development, students are concerned that frank and honest feedback may cause problems with the team, so evaluations are guarded. By the end of the semester, when the consequences of poor individual performance are more obvious, evaluators are likely to give a non-performing student a lower score. But since the final evaluation is only one-third of the total contribution evaluation, the non-performing student is likely to get a higher grade than the evaluator would like to see.

Potential for punitive behavior

When students have a long-term disagreement with their teammates, they can tend to withdraw from the team and do their work primarily on an individual basis. In such cases, their teammates may respond punitively and give an extremely low contribution multiplier, such as 0.2. There is a concern that such low multipliers are punitive, rather than fact-based. They also have the potential of creating an atmosphere for grading retaliation.

In an attempt to mitigate these challenges, we have taken at least three distinct steps. First, we have minimized the benefit to the team for giving a low multiplier score. Because the average of the individual multipliers must be no higher than 1.0, low multiplier scores for an individual provide extra points that can be used to raise another individual's scores. For a five-member team, if one member had a multiplier of 0, everybody else could have a multiplier of 1.25. In order to minimize this benefit, we treat multiplier scores lower than 0.8 as if they were 0.8 for purposes of calculating the average. Thus, there is no benefit to team members for giving a multiplier lower than 0.8.

A second step we have taken to avoid punitive behavior is to have the coaches work closely with their teams. We try to identify problems and help the team resolve them before evaluation time. We also expect that team members will state their concerns vocally before evaluation time. For most cases, this prevents punitive behavior.

A third step we take to avoid punitive behavior is to counsel individually with the student evaluator whose score appears punitive. Students are promised that their evaluations will be anonymous, and the anonymity is preserved as far as the coach and teammates are concerned. However, in order to give students credit for completing the evaluation, we keep a record of each student's evaluation. If there appears to be a problem with an evaluation, one of the instructors of the course will visit with the student. During this visit, the reasons for the low evaluation will be discussed. The instructor shares the reason for the concern, and the student shares the reason for the low evaluation. If the evaluation is punitive, the student generally feels uncomfortable and will adjust the score. On the other hand, students who feel they have a reasonable basis for giving a low score will likely keep the low score. As the student has the responsibility for the score, we feel it is inappropriate for an instructor to adjust it, so the student score will stand.

Problems with punitive grading tend to show up on rare occasion, occurring in less than 0.5% of peer-assigned multiplier scores.

Successes

The current capstone grading scheme is successful overall. In particular, students are able to understand the grading method and calculate their own grade, given the raw data. In addition, we have observed relatively consistent evaluations when different evaluators evaluate the same characteristic.

Students know the source of their grade

Unlike most classes, it seems like capstone students consistently believe that they should receive an A for their work. If grades are not recorded throughout the semester, receiving a grade lower than A comes as quite a shock. However, when the evaluation items, the evaluators, the evaluation criteria, and the grading scheme are known in advance, students are much more accepting of the grade they have earned.

In the past, it was common for students to come and argue for a higher grade. When everything is explained, students who come to question their grade are asked to find the raw data (which is shared with the students on Blackboard), then run through the calculation themselves to ensure everything is calculated correctly. Once they have done so, they understand the reason for their grade, and there is much less dissatisfaction.

Consistency of evaluations

When evaluations are largely subjective, there is always concern that the evaluations are neither repeatable nor reliable. An unplanned but fortuitous outcome of the current grading scheme is the fact that every component of an individual's project grade is evaluated by more than one person, and the correlations between the grades can be observed. In particular, we have observed good correlations between the coach and instructor evaluation of overall team performance, the various coaches' reviews of the final report, and the coach and peer evaluations of individual students on their team.

Various graders of report

Each team's final report is graded by three coaches of other teams. The coaches are given the capstone evaluation criteria, and asked to evaluate the reports according to their professional judgment.

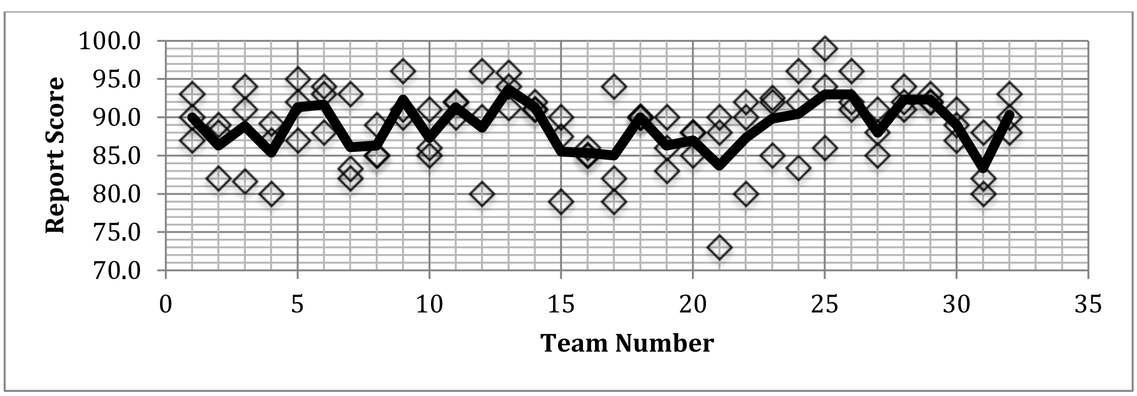

Figure 1: Final report scores vs. team number. Markers indicate scores given by evaluators. The line indicates the average score

The individual coach grades given to each team are shown in Figure 1. For many cases, the spread in the data is small compared with the differences between the teams. However, in some cases the spread is quite large (see, for example, team 21). Having this much variation in the data is somewhat concerning.

Further investigation of the variation was conducted to see if a correlation exists between variation and report grade. Figure 2 shows the absolute deviation of each report score as a function of the average report score. No correlation appears to be present. Figure 3 shows the standard deviation of the set of report scores as a function of the report score. Again, no correlation appears to be present. Given this analysis we infer that the variation in the report scores is due to the inherent variation in the grading process, and we are comfortable with using the grades.

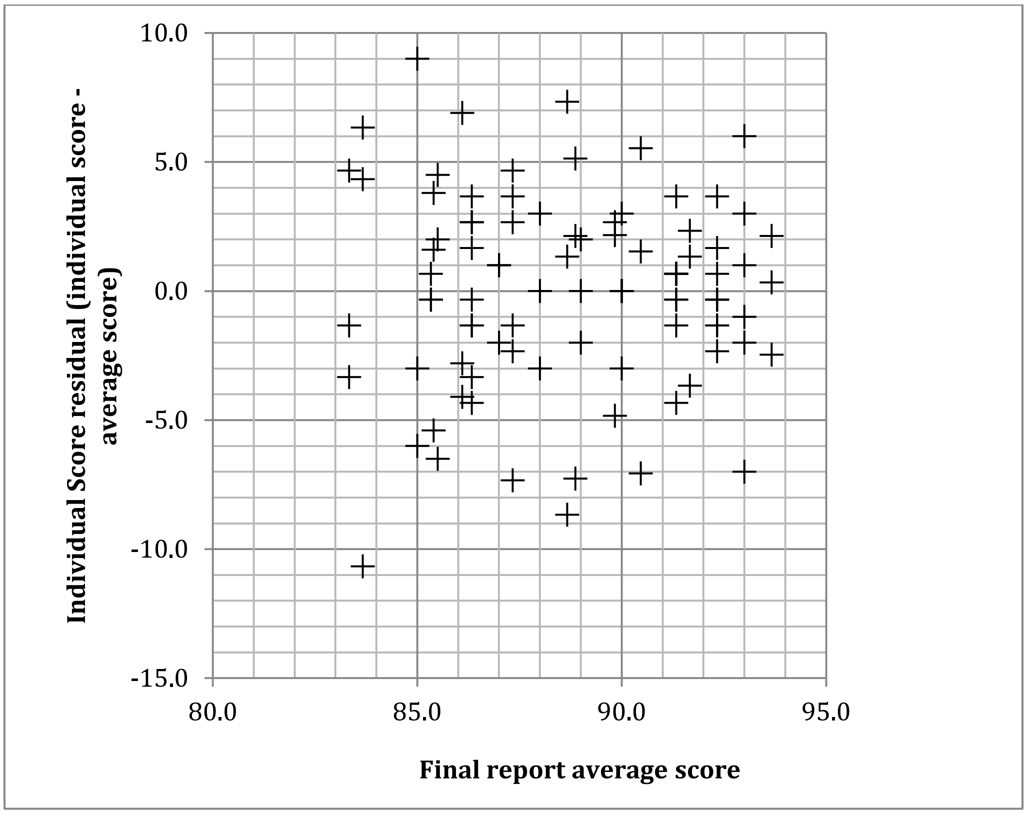

Figure 2: Absolute deviation of individual score from average score vs. average score.

The score range for the final report averages is fairly narrow, from a low of 83 to a high of 94. The lack of average scores higher than 94 is indicative of the graders' unwillingness to automatically give 100 to a report without any errors, as desired in the capstone grading criteria. The absence of scores below 83 is thought to be due to two primary effects. First, a draft version of the report was reviewed before the final submission, giving teams an opportunity to avoid the most egregious mistakes. Second, the internal reviewing of the report within the team leads to a minimum quality in the reports that is acceptable according to the grading criteria.

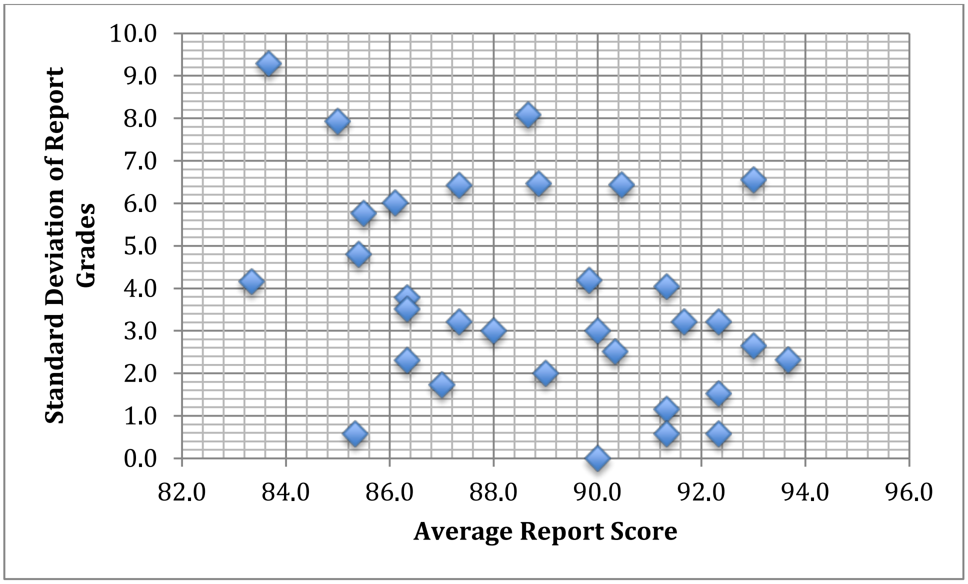

Figure 3: Standard deviation of scores vs. average score

Various evaluators of overall team performance

Overall team performance at the end of the semester is measured by the coach in the final team process score. It is also measured by the instructors in the semester project review. Since both scores are intended to be an evaluation of the overall performance of the team, we would hope that there is correlation between the independent measurements.

Figure 4 shows the instructor project review grade for each team as a function of the coach final team process score. Note that there is good correlation, and that a linear fit to the data has a reasonable R^2 value (indicative of a correlation of about 0.4). It appears that the coaches generally rate the team higher than the instructors (about 3 points higher, on average). Most of the coach scores fall in line with the instructor scores, but there are a few cases where the coach score is significantly higher than the instructor score.

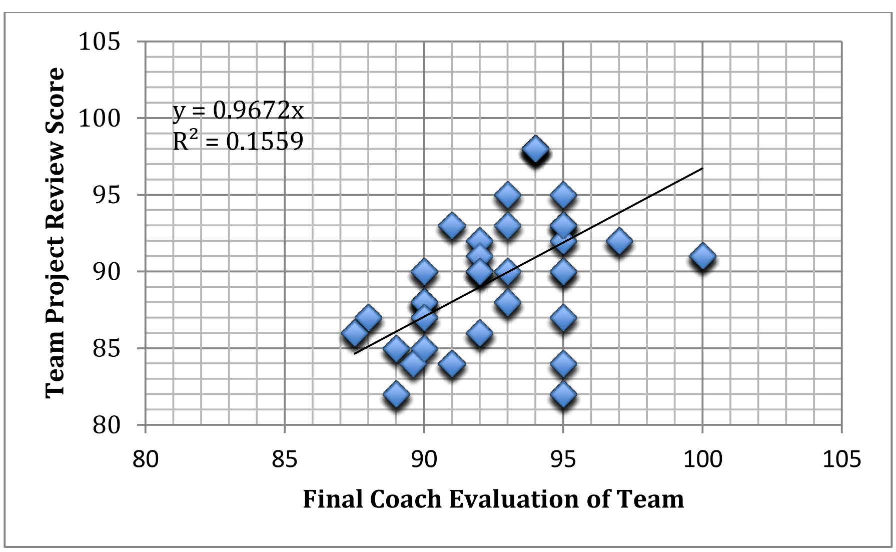

Figure 4: Instructor project review score as a function of final coach evaluation

If we remove the four data points that are farthest from the trend line in Figure 4, the correlation greatly increases as shown in Figure 5. The fit implies a correlation coefficient of nearly 0.7.

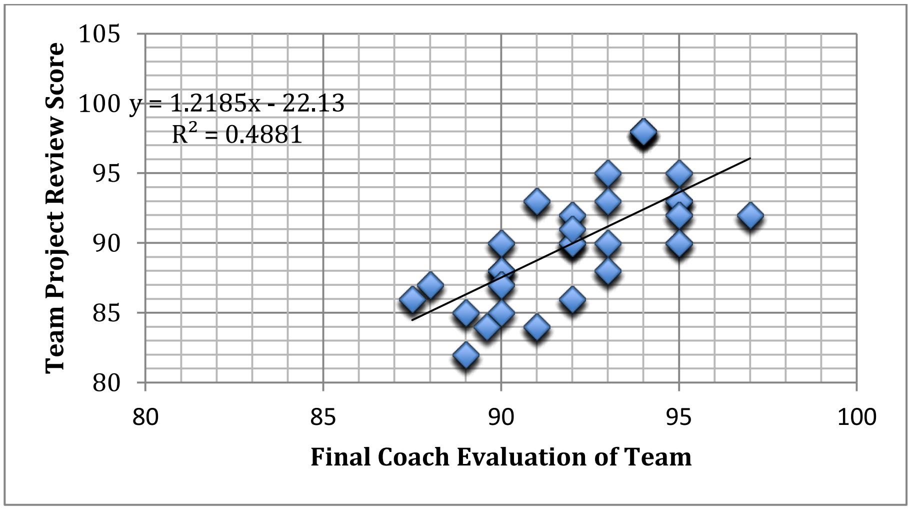

Figure 5: Project review score vs. final coach evaluation with outliers removed

Various graders of individual contribution

The individual's contribution to the team is evaluated by both teammates and the team coach. These constitute independent evaluations, and can be used to assess the validity of the data.

Figure 4 shows the instructor individual contribution multiplier as a function of the peer individual contribution multiplier for all students in the class. Note that most of the scores are very close to 1.0, which indicates that in general, students are contributing equally. There is a general trend that shows students who are highly rated by their peers are likely to also be highly rated by their coach.

The range of coach scores is higher than the range of peer scores. This is expected, since the peer scores that are reported are the average of multiple evaluations. Using an average instead of the individual scores reduces the range of the data.

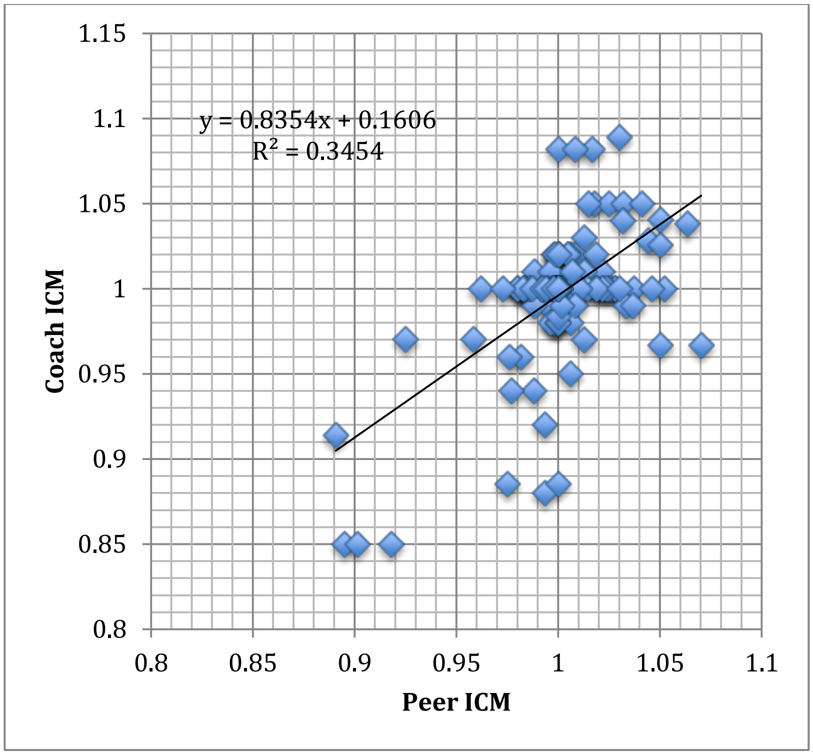

Figure 6: Coach-assigned ICM vs. peer-assigned ICM

Conclusions

Although it is inherently difficult, robust methods for evaluating design team work can be developed.

An evaluation method has been developed at BYU that includes the effects of individual learning, team performance, and individual contributions to team performance. This method minimizes the possibility for a student to skate along and obtain a satisfactory grade in the class without making significant contributions to the design project work of the team.

As part of the evaluation process, independent evaluations are obtained for each of the major contributions to an individual's grade. Examination of these independent evaluations indicates that the grades are likely to be repeatable and reliable.

References

1. Todd, R., C. Sorensen, and S. Magleby, "Designing a senior capstone course to satisfy industrial customers" *Journal of Engineering Education*, 1993, Vol. 82, No. 2, pp. 92-100.
2. *Product and Process Development Guide and Record Book, 2011-2012*, ME Courses 475-476, Department of Mechanical Engineering, Brigham Young University, Provo, Utah 84604
3. Fink. L. D., *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*. San Francisco: Jossey-Bass, 2003.
4. Banta, T. W., E. A. Jones, and K. E. Black, *Designing Effective Assessment: Principles and*

Profiles of Good Practice. San Francisco: Jossey-Bass, 2009
5. Michaelsen, L. K., A. B. Knight, L. D. Fink, (eds.), *Team-based Learning: A Transformative Use of Small Groups*. Sterling, VA: Stylus Publishing, 2004.
6. Wiggins, G., and J. McTighe, *Understanding by design (2nd ed.)*. Upper Saddle River, NJ: Prentice Hall, 2005.
7. Middle States Commission on Higher Education. *Student Learning Assessment: Options and Resources*. 2003.
8. Stevens, D. D., and A. J. Levi, *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback and Promote Student Learning*. Sterling, VA: Stylus Publishing, 2004
9. Suskie, L., and T. W. Banta, *Assessing Student Learning: A Common Sense Guide*. San Francisco: Jossey-Bass, 2009.
10. Shah, J.J., R.E. Millsap, J. Woodward, and S.M. Smith, "Applied Tests of Design Skills—Part 1: Divergent Thinking" *Journal of Mechanical Design*, 2012, Vol. 134, Vol. 2, pp. 1-10.
11. Shah, J.J., S.M. Smith, and J. Woodward, "Development of Standardized Tests for Design Skills" *International Conference on Engineering Design*, Stanford, CA, 2009.
12. Shah, J.J. "Identification, Measurement & Development of Design Skills in Engineering Education" *Proceedings of the 15th International Conference on Engineering Design*, *2005*, pp. 1-15.
13. Shah, J.J., N. Vargas-Hernandez, and S.M. Smith, "Metrics for measuring ideation effectiveness" *Design Studies*, 2003, Vol. 24, No. 2, pp. 111-134.
14. Brackin, M. P., J. D. Gibson, "Methods of assessing student learning in capstone design projects with industry: A five year review" *Proceedings of the 2002 American Society for Engineering Education Annual Conference & Exposition*.
15. Dym, C. L., A. M. Agogino, O. Eris, D.D. Frey, and L. J. Leifer, "Engineering design thinking, teaching, and learning" *Journal of Engineering Education*, 2005, Vol. 94, No. 1, pp. 103-120.
16. Dym, C.L., "Teaching design to freshmen: Style and content" *Journal of Engineering Education*, 1994, Vol. 83, No. 4, pp. 303-310.
17. Gnanapragasam, N., "Industrially sponsored senior capstone experience: Program implementation and assessment" *Journal of Professional Issues in Engineering Education and Practice*, July 2008, pp. 257-262.
18. Kelley, T. R., and R.C. Wicklein, "Examination of assessment practices for engineering design projects in secondary education. (Second in a 3-part series)" *Journal of Industrial Teacher Education*, 2009, Vol. 46, No. 2, pp. 6-25.
19. Maskell, D., "Student-based assessment in a multi-disciplinary problem-based learning environment" *Journal of Engineering Education*, 1999, Vol. 88, No. 2, pp. 237-241.
20. Shuman, L. J., M. Besterfield-Sacre, and J. McGourty, "The ABET "Professional Skills" - Can They Be Taught? Can They Be Assessed?" *Journal of Engineering Education*, 2005, Vol. 94, No. 1, pp. 41-55.
21. Sims-Knight, J., R. Upchurch, N. Pendergrass, T. Meressi, and P. Fortier, "Assessing design by design: Progress report 1" *American Society for Engineering Education Frontiers in Education Conference, 2003.* Session T4B.
22. Sven, G. B., E. C. Kisenwether, S. E. Rzasa, and J. C. Wise, "Developing and Assessing Students' Entrepreneurial Skills and Mind-Set" *Journal of Engineering Education*, 2005, Vol. 94, No. 2, pp. 233-243.

23. Wolcott, M., S. Brown, M. King, D. Ascher-Barnstone, T. Beyreuther, and K. Olsen, “Model for faculty, student, and practitioner development in sustainability engineering through an integrated design experience” *Journal of Professional Issues in Engineering Education and Practice*, April 2011, pp. 94-101.
24. Todd, R.H., “The How and Why of Teaching and Introductory Course in Manufacturing Processes” *1991 Frontiers in Education Conference*, pp. 460-463.
25. Asunda, P. A., and B. R. Hill, “Critical features of engineering design in technology education” *Journal of Industrial Technology Education*, 2007, Vol. 44, No. 1, pp. 25-48.
26. Asunday, P. and R. Hill, “Critical features of engineering design in technology education” *Journal of Industrial Teacher Education*, 2007, Vol. 44, No. 1, pp. 25-48.
27. Bailey, R., Z. Szabo, and D. Sabers, “Integrating education students in the assessment of engineering courses” *Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition*. Session 3130.
28. Deek, F. P., R. H. Starr, H. Kimmel, and N. Rotter, “Cognitive assessment of students' problem solving and program development skills” *Journal of Engineering Education*, 1999, Vol. 88, No. 3, pp. 317-326.
29. Sobek, D. K., and K. J. Vikas, “Two instruments for assessing design outcomes of capstone projects”, *Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition*. Session 2425.
30. Bailey, R., and Z. Szabo, “Assessing engineering design process knowledge” *International Journal of Engineering Education*, 2006, Vol. 22, No. 3, pp. 508-518.
31. Kruck, S.E., and H.L. Reif, “Assessing Individual Student Performance in Collaborative Projects: A Case Study”, *Information Technology, Learning, and Performance Journal*, 19:2 (Fall 2001), 37-47.
32. Lejk, M., M. Wyvill, and S. Farrow, “A survey of methods for deriving individual grades from group assessments”, *Assessment and Evaluation in Higher Education*, 24:1 (1999), 5-14.
33. Gibbs, G. “The assessment of group work: lessons from the literature”, Assessment Standards Knowledge Exchange, Centre for Excellence in Teaching and Learning in Higher Education, Oxford Brookes University, http://www.york.ac.uk/media/staffhome/learningandteaching/documents/keyfactors/The_assessment_of_group_work_lessons_from_the_literature.pdf, accessed 15 Feb 2012.
34. Sergi, M., “Evaluating short-term and long-term peer assessment of Student Teamwork”, *e-Journal of Business Education & Scholarship of Teaching*, Vol.1, Iss.1 (2007), 41-58.
35. Lewis, T., “Coming to terms with engineering design as content” *Journal of Technology Education*, 2005, Vol. 16, No. 2, pp. 37-54.
36. Welch, M., “Assessment in Technology Education: What, why, and how?” *Proceedings of the Second AAAS Technology Education Research Conference*. American Association for Advancement of Science, 2001.

Appendix 1: 2011-2012 Grading Algorithm

With these principles described in this article in mind, a grading algorithm was created based on the three areas of evaluation for the class. As noted previously, the three areas of evaluation are, Individual learning, Team grade, and Individual contribution to the team. The percentage and breakdown of each of these areas is noted below:

Area 1 (33%, graded by instructors) Evaluates each student's individual learning of the materials taught in the course and includes evaluation elements as follows:

Item	Weight	Evaluated by
Product and Process development Examination	16%	Instructors
Assignment 1	15%	Instructors
Assignment 2	5%	Instructors
Assignment 3	5%	Instructors
Credit for completing evaluation surveys	2%	Yes/No

Area 2 (67%) Evaluates the process each team has been using to pursue the objectives of the sponsored project. It also evaluates the results of your team's efforts.

Item	Weight	Evaluated by
Team Process Grade – September	8%	Team coach
Team Process Grade – October	8%	Team coach
Team Process Grade – Nov/Dec	8%	Team coach
Team Shop Performance Grade	4%	Shop TAs
Team Design Milestones	4%	Yes/No
Semester Written Project Report	20%	3 other coaches
Project review	15%	Instructors

Area 3 Evaluates each student's efforts-as an individual, in pursuit of the team's progress.

This is an Individual Multiplier (above or below 1.0) multiplied by the Team Grade. It is determined by averaging three Peer and Coach Reviews during the semester.

The Final Grade = Individual Learning Grade + (Team Grade X Individual Contribution to Team Multiplier)

An individual example and blank Worksheet to enable students to keep track of their scores is included in the Guidebook.

This table, included in the guidebook, provides a mapping function of percent scores to letter grades so the students can calculate themselves how they are performing.

A	93 - 100
A-	90 - 92.99
B+	87 - 89.99
B	83 - 86.99
B-	80 - 82.99
C+	77 - 79.99
C	73 - 76.99
C-	70 - 72.99
D+	67 - 69.99
D	63 - 66.99
D-	60 - 62.99
E	0 - 59.99

Appendix 2: Capstone Evaluation Criteria

A list of capstone evaluation criteria is also included in the guidebook as follows, to assist the students of faculty coaches in their evaluation efforts.

CAPSTONE EVALUATION CRITERIA

This table describes six basic categories when evaluating the items in the grade breakdown. The table provides a quality of work description, grade range, and percentage range for each category. Evaluators will judge the work based on the descriptions provided in the description column of the table, and will assign a percentage to more clearly pinpoint the quality of the work.

Quality of Work *Percentage Range* Grade Equivalent	**Characteristics**
Excellent – meets latent and explicit needs *90 – 100* A- to A+	**Thinking**: The design and engineering thinking reflects exceptional understanding of the project and the process. **Technical Work:** The work covers the essential issues and more, and is correctly performed **Timing**: The timing is excellent; it appears to be ahead of schedule **Communication:** The communication is highly effective, exceeding the expected standards **Teamwork:** The team has worked synergistically; the whole is much better than the sum of its parts. **Initiative**: The team has taken the responsibility for project success, and has eagerly sought to identify things they can do to delight the sponsor, including seeking advice and help from others.
Good – meets explicit needs *80 – 89* B- to B+	**Thinking**: The design and engineering thinking is conscientious, thorough, and correct **Technical Work**: There is good technical work in support of the project. Most of the pertinent areas are covered, and there are very few technical flaws. **Timing**: The timing is correct; it appears the project will be completed on schedule **Communication**: The communication is competent, clear, concise, complete and professional **Teamwork**: The team has worked together effectively to achieve the desired results **Initiative**: The team is engaged in producing a successful outcome, and has done what is expected of them.
Satisfactory – meets some explicit needs *70 – 79* C- to C+	**Thinking:** The design and engineering thinking is superficial in one or more areas, and minimally meets the needs of the project. **Technical Work**: There is technical work in support of the project. Some important areas are missing, or technical flaws that should have been corrected are present. **Timing**: The timing is off to the extent that it concerns the stakeholders **Communication:** The communication may be somewhat superficial, misleading, wordy, incomplete, unclear, or incorrectly focused. **Teamwork:** The team has worked together, with some issues hindering the team progress. **Initiative**: The team has not really accepted responsibility for the project, but generally does what is required in order to get the grade. Regular external prodding is necessary to keep the team moving forward.

Quality of Work *Percentage Range* Grade Equivalent	Characteristics
Poor – partially meets some explicit needs *60 – 69* D- to D+	**Thinking**: The design and engineering thinking is minimal; there appears to be little or no quality thinking related to the project **Technical Work**: There is only superficial technical work done. The work is not relevant to the project. It may have been done just to meet a requirement. Serious technical errors remain **Timing**: The timing is off to the extent that is it likely to significantly affect the desired outcomes **Communication**: The communication is poor or thoughtless. It appears to be primarily an effort to meet externally-imposed requirements **Teamwork**: The team has struggled to work together, with significant challenges in the team hindering work **Initiative**: The team is resistant to doing what needs to be done, and may often look for ways to meet the letter of the requirement, but fails to grasp and/or respond to the spirit of the requirement. The project is worked on at the minimal possible level. The team only gets to a reasonable outcome with significant, repeated external prodding.
Failing – fails to meet explicit needs *0 – 59* E	**Thinking**: There is little or no evidence of careful or creative thought **Technical Work**: There is little or no evidence of technical work **Timing**: The timing is off such that it may cause the project to be unrecoverable **Communication**: The communication is unacceptable **Teamwork**: The team has spent more effort in team issues than in accomplishing project work **Initiative:** The team has failed to do significant work on the project, even with significant external prodding.

Appendix 3: Individual Contribution Materials

The table below, also published in the Guidebook, can also assist as a rubric in helping team members and coaches assign an individual multiplier to represent each team member' contribution.

Individual Contribution Multiplier	**Description of performance**	**Individual contribution score (grade) for a team score of:**		
		85 (B)	90 (A-)	94 (A)
1.2	Exceptional performance – far above normal for the team	102 (A+)	108 (A+)	113 (A+)
1.1	Outstanding performance – significantly above normal for the team	93.5 (A)	99 (A+)	103.4 (A+)
1.0	Normal performance for the team	85 (B)	90 (A)	94 (A)
0.9	Inferior performance – significantly below normal for the team	76.5 (C)	81 (B-)	84.6 (B+)
0.8	Marginal performance – little contribution to the team	68 (D+)	72 (C-)	75.2 (C)
0 – 0.7	Failing performance – insignificant contribution to the team	0 (E) – 59.5 (E)	0 (E) - 63 (D-)	0 (E) - 65.8 (D)

The worksheet below shows a sample of a spreadsheet that is given to students to help them understand the effects of the ICM scores. Students can change the ICM score given to each team member and see the effect on the Individual contribution score.

		Team Score (Assigned by Coach)	Justification of Team Score (Assigned by Coach)
Team	7	93	Team did excellent work
Reviewer	John		

Team Member	Your given ICM Score	Score used for average calculation	Actual ICM after adjustment	Estimated Individual Score (calculated from Team Score and Assigned ICM)	Justification of ICM Score
	Average of Passing ICM				
John	1	1	1	93	doing great
Susan	1.1	1.1	1.1	102.3	went above and beyond
Mark	9	1.2	1.2	83.7	shows up late
Joseph	1	1	1	93	doing great