Classroom Assessment and Design Education

Robert W. Brennan
Schulich School of Engineering, University of Calgary
rbrennan@ucalgary.ca

Abstract

Classroom assessment is a challenge in engineering design given the open-ended nature of student projects, the reliance on multiple assessors, and difficulties around the judgement of "quality". In this paper, we attempt to place the more general work on classroom assessment in the context of engineering design education. Examples are provided from the author’s experience teaching a mechanical and manufacturing engineering capstone design course and recommendations are made based on this experience.

1 Introduction

The open-ended nature of engineering design, and in particular, the capstone design course sets it apart from other courses in the engineering curriculum. By definition, these courses should "provide students with a significant design experience based on the knowledge and skills acquired in earlier course work, and give students an exposure to the concepts of team work and project management" [1]. Although capstone design courses often include classroom instruction on design methodology, design theory, and project management, the focus tends to be on hands-on project work by student teams. In other words, teaching and learning in capstone design is characteristically inquiry-based [2].

Given the nature of capstone design courses however, the design of student assessments can be a challenge. In particular, capstone design course instructors face the challenge of coordinating multiple student projects: each of which involves a separate student group and faculty “project advisor” or “coach”. The instructor must develop assessment tools that facilitate consistent assessment across multiple projects, multiple teams, and multiple assessors. Additionally, these tools must be appropriate for the typical range of student deliverables for capstone design courses: e.g., reports, presentations, reviews, written assignments, and/or prototypes.

Given these challenges one might ask, why assess project work at all? For example, in most capstone design courses the coach and in some cases the project sponsor or “customer” serve as the examiners. Each of these individuals typically has intimate knowledge of the project and the team’s performance and could make a good argument that the team’s result is a foregone conclusion.

As Powell notes [3], there are a number of substantial challenges to this line of thinking. In particular, it is tricky being facilitator and judge at the same time. To address this, it is common to use a report or series of reports and design reviews for assessment [4]. The key advantage to this approach (over the “foregone conclusion”) is that the student team has the opportunity to defend their decisions and also learn from their mistakes.

The challenge to teaching faculty is to create assessment tools that, (1) have clear requirements, (2) are clearly linked to the course objectives, (3) are flexible, and (4) are fair. This paper does not attempt to provide specific solutions to this problem, but instead draws from the work on classroom assessment and attempts to put this in the context of capstone design. The paper begins by focusing on classroom assessment tools in Section 2. An important aspect of this is the validity and reliability of these tools. This topic is introduced in Section 2, and discussed in more detail in Section 3 around this issue of assigning grades to students and student teams. The paper concludes with a brief summary of key issues associated with capstone design assessment that have followed from the author’s experience.

2 Assessment in Capstone Design

The term “assessment” varies somewhat from author to author in the general education and engineering education literature, and as a result, it is important to be clear on what is meant by “assessment” in this paper. For example, Olds et al. [5] view assessment broadly as the "act of collecting data or evidence that
can be used to answer classroom, curricular, or research questions”.

In this paper we focus specifically on measuring individual students’ competencies. More specifically, we use Brookhart’s [6] definition that assessment involves “gathering and interpreting information about student achievement”. As a result, this section focuses on the assignments and tasks that course instructors develop to provide this information on student achievement. This should be distinguished from “evaluation”, which refers to the judgements that instructors make about how well students attain the course learning goals.

Ideally, classroom assessment should be an ongoing process, used to collect evidence of student achievement while students are in the process of learning (i.e., “formative” assessment) as well as at the end of a learning activity (i.e., “summative” assessment) [7].

In this section we begin by looking at the role classroom assessment plays in course design, and in particular, its link to learning outcomes and activities. Next, we review the typical forms of assessment that may be used in capstone design, then conclude with a discussion of the issues around validity and reliability of assessment.

2.1 The Link to Learning Outcomes

When designing an assessment tool for any course, it is important to match the tool with the learning objectives/outcomes for the assessment [5, 6, 7]. In this way, the instructor can determine the extent to which students have achieved the learning objective, and ideally, the learning objective can be reinforced for the student (e.g., via feedback within the assessment). As illustrated in Figure 1, the instructor’s learning outcomes for a course must be clearly stated at the outset and these learning outcomes should be used to inform the development of appropriate learning activities and assessment tools.

Once developed, the assessment tools “should be the focus of ongoing communication with students about their achievement of the course’s objectives” [6]. This is particularly relevant to design courses, which often span two full semesters and typically involve less in-class activity than traditional engineering courses.

For capstone design in Canada, the CEAB (Canadian Engineering Accreditation Board) has been fairly explicit about these learning goals: i.e., capstone design courses should “provide students with a significant design experience based on the knowledge and skills acquired in earlier courses, and give students an exposure to the concepts of teamwork and project management” [1].

2.2 Forms of Assessment

Before tackling the question of appropriate assessment for capstone design, it is useful to look at the common forms of classroom assessment that are used across various disciplines. Brookhart [6] identifies four categories: paper-and-pencil tests, performance assessment of processes or products, oral communication, and portfolios. Arguably, the most common forms of assessment in engineering science courses are the first two from this list. Although many engineering science courses may have student or student team projects, the majority of assessments developed for these courses are in the form of written assignments (performance assessment of processes or products) and pencil-and-paper tests (e.g., quizzes, midterms, and finals).

Ideally, all student assessment should be both “authentic” and “educative” [8] whether one assessing students in engineering, science, the humanities, or any other discipline. In the context of engineering education, authentic assessment tools are those that are grounded in the actual practice of engineering, while educative assessment tools are those that are used to help teach and improve student performance.

Written exams and assignments are certainly appropriate when the learning goals focus on problem solving and analysis typical of many engineering science courses. However, in the case of engineering design, instructors struggle to provide assessments that mirror the work that practicing engineers actually do in the workplace (i.e., authentic assessment). Although educative assessment is still relevant to engineering design, it is arguably less of a focus given that these courses integrate material from other courses in the engineering curriculum where considerable educative assessment is already done.
Given the nature of engineering design, student assessment in these courses tends to cover all four of Brookharts’ [6] categories, with distinctly less (and often no) reliance on pencil-and-paper tests. More specifically, capstone design instructors typically focus on design reviews (i.e., oral communication), logbooks (i.e., portfolios), and design reports (i.e., performance assessment of processes or products).

### 2.3 Validity and Reliability

Once an appropriate form of assessment is determined, one must ensure that the resulting assessments are valid and reliable given course objectives for capstone design. This is where we start to move into evaluation, or judgement, of student performance. As noted previously, the challenges relate to the open-ended and diverse nature of student projects, the need for team-oriented assessment, and the reliance on multiple instructors.

Validity is related to the “degree to which a score is meaningful and appropriate for its intended purpose” [6]. Although students may view design reviews, design reports, and logbooks as painful, they are certainly inline with professional practice and give instructors with the opportunity to provide informed and timely feedback on student performance. In other words, they appear to be valid forms of assessment: whether or not the resulting scores are valid will be discussed in the next section.

Reliability however, tends to be more problematic in capstone design. Here, we are referring to “the degree to which a score is consistent across time or judges or forms of assessment” [6]. Each of these areas is potentially problematic. First, “consistency across time” is difficult to achieve given that students assessment must often be scheduled over days (or weeks in large classes). More specifically, design reviews must be scheduled in numerous slots to accommodate instructor, coach, teaching assistant, and/or sponsor time constraints. It is unlikely that the time of a design review would influence a student team’s score, however this should be taken into consideration when designing the assessment.

The second area, “consistency across judges”, is typically the most problematic. Also referred to as “inter-rater reliability”, this occurs in capstone design through the assignment of faculty advisors or coaches to student teams. Ideally, a given student team would obtain the same score from each and every one of the coaches: in practice this is often not the case. Engineering faculty assigned to capstone teams often comment on the difficulty of assessing something as subjective as the quality of a design. Most will say that they recognize quality when they see it, however each faculty member’s perception of “quality” may vary considerably.

One approach to addressing the issue of inter-rater reliability – especially when assessing quality – is through the use of rubrics. Rubrics are “descriptive rating scales that are particularly useful for scoring when judgement about the quality of an answer is required” [6]. An example of an analytic scoring rubric used to assess student case studies in the author’s capstone design course is shown in Figure 2.

![An analytic scoring rubric](image-url)
The author’s experience with this form of assessment in a large capstone design course (i.e., 140 students, 30 student teams) [9] has been very positive in terms of improving inter-rater reliability as well as in terms of improving the educative aspects of the assessment. In particular, rubrics can be shared with students ahead of time: this can be used to facilitate discussion on the assessment (and learning objectives) and also provide students with clear benchmarks.

The third aspect of reliability, “consistency across forms of assessment”, at first appears to be of little concern in capstone design. For example, one would worry about consistency across forms of assessment in the case of a make-up exam or a substitution of an assignment for a particular student. Since every team is typically assessed using the same assessment tools (e.g., all case studies are assessed with the rubric shown in Figure 2), this should not be a problem in capstone design courses.

However, it is very common in these courses that every team has a different project: typically, the “form” of these projects will vary considerably. For example, student projects may vary in scope (e.g., full system v. sub-system design), in type (e.g., product v. process design), in context (e.g., industry-sponsored project v. faculty-sponsored project), or in some other way. Although the assessment tools may be consistent across forms in the strict sense, reliability of assessment across student teams may come into question given the nature of the design projects.

The bottom line is that “a score cannot be any more valid than it is reliable” [6]. In other words, when choosing an assessment tool, issues around its consistency must be first addressed. The question of the overall validity of the assessment can then follow.

3. Grading Capstone Design

In this section, we continue with the discussion on assessment validity by looking more closely at the assignment of grades in capstone design, and specifically, at the meaningfulness and appropriateness of these grades.

In engineering courses, it is very common to determine a percentage score for a student’s term work then convert this score into a final letter grade. For example, a combined weighted score of 85/100 or greater may equate to an “A”, while a 50/100 or less would be assigned an “F”.

Given that the combined weighted score is an aggregate of individual assessments, it is not uncommon to have scores like “87.32” or “92.61”. What happens though, when a student receives 49.95/100 given the criteria noted above? Is this a “pass” or “fail”? How does this student’s performance compare to a student who receives 50.01/100? Strict adherence to the policy would lead to the first student receiving an “F” and the second student passing: from a practical point of view though, their performance is effectively the same.

Understandably, students are prone to quibble about scores that are close to breakpoints between grades; consequently, instructors must have their stories straight to maintain consistency and fairness. However, in most courses “percentages and points imply much more precise measurement than may exist” [6].

It seems reasonable that this is particularly relevant to capstone design given the reliability concerns noted previously. However, this does not answer the question of how engineering design instructors can avoid this problem of precision (or lack thereof).

The issue is not necessarily related to the validity of the individual assessments, but to the act of combining individual assessments into an overall score. In other words, after weighting individual assessments, then adding them together to obtain an aggregate score (e.g., out of 100), this resulting score may not be as precise as the one or two decimal places imply. When a variety of assessment types are used (e.g., 5 scale rubrics as shown in Figure 2, percentage scores, complete/incomplete scores, pass/fail scores, etc.).

Instructors use various methods to combine term work into a final letter grade. From the literature on classroom assessment, Brockhart [6] identifies four common approaches: (1) median, (2) weighted letter grade, (3) total possible points, and (4) holistic. In order to illustrate these approaches a simple example of a set of capstone design course grades is shown in Table 1.

As shown in this table, student teams are assessed using design reviews, design reports (an “interim report” and a “final report”), and a case study. As well, analytic rubrics similar to the one shown in Figure 2 are used for the design reviews and the case study, while the design reports are graded on a 20-point scale.

In order to assign letter grades for these different types of assessments, the descriptors shown at the top of Figure 2 are used for the rubrics, and a point-to-grade range is used for the point scores. More specifically, the rubric letter grades are assigned as follows:

- “Excellent” (4/4) = “A”
- “Good” (3/4) = “B”
- “Satisfactory” (2/4) = “C”
- “Unsatisfactory (Marginal)” (1/4) = “D”
- “Unsatisfactory” (0/4) = “F”
These descriptors correspond approximately to what would be described as an “A”, “C”, “D”, or “F” in a typical university calendar.

For the design reports, the scores are converted to a percentage and a linear scale is used to determine the letter grade. A typical set of upper and lower limits is: “F” ≤ 50% and “A” ≥ 90%. On this basis, three different student teams are compared in order to illustrate the differences inherent in the methods noted above.

The first two methods are used when a letter grade is assigned to each assessment. As the name implies, the median approach takes the median of the individual assessments. For example, the first team has received effectively four “A” grades (three for the “final report”, which is worth 30%, and one for the “case study”; which is worth 10%), two “B+” grades (for the “interim report”, which is worth 20%), three “B” grades (for the last three “design reviews”, worth 10% each), and one “C” grade (for the first “design review”, which is worth 10%). The resulting median grade is a “B+”.

The weighted letter grade approach differs by taking the average of the letter grades. In other words, the grade point values are used to determine the final letter grade. For example, a typical grade point scheme is “A” = 4.0, “A-” = 3.7, “B+” = 3.3, “B” = 3.0, “B-” = 2.7, “C+” = 2.3, “C” = 2.0, “C-” = 1.7, “D+” = 1.3, “D” = 1.0, and “F” = 0. In this case, Team 1’s final letter grade is determined by: (0.1)(2.0) + (0.1)(3.0) + (0.1)(3.0) + (0.1)(3.0) + (0.2)(3.3) + (0.3)(4.0) + (0.1)(4.0) = 3.36, or a “B+”.

The total possible points approach is fairly common in engineering courses. With this approach, a letter grade is not assigned to each assessment. Instead, the points for each assessment are weighted then added together to obtain a total score. For example, Team 1’s final score would be determined by (0.1)(3/5) + (0.1)(4/5) + (0.1)(4/5) + (0.1)(4/5) + (0.2)(17/20) + (0.3)(19/20) + (0.1)(5/5) = 85.5/100, or a “B+”.

Finally, the holistic method (not shown in Table 1) typically uses a single rubric for assigning overall grades for the course. For example, an “A” may be described as being achieved by “passing 90% or more of the assignments” while a “B” is achieved by “passing 80% or more of the assignments” [6]. This approach is useful for courses that use “pass/fail” assessments or for senior seminars where a very focused rubric for overall assessment can be devised.

As can be seen in Table 1, the “median”, “weighted letter grade”, and “total points possible” methods are consistent for the Team #1 (i.e., a team with “good” to “excellent” performance on the majority of its assessments). However, the methods based on letter grades (“median” and “weighted”) result in higher final letter grades for Teams #2 and #3. In these cases, the teams are showing “marginal” to “satisfactory” performance on most of their assessments with one heavily weighted “fail” and one lightly weighted “good” performance.

The “total possible points” approach averages all of the scores to obtain a final percentage that is then converted to a letter grade. For courses where all of the assessments are points-based (like the design reports here), this may be the most appropriate technique to use. However, it is important that the assessments are well planned in advance and that the instructor is confident with the validity and reliability of these assessments. For example, a well tested set of assessments may fall in this category.

When a mixture of assessment tools is used though (in this case, rubrics and points-based assessments), the total possible points method may not be the best choice. In these cases, the choice will center on the importance that the instructor places on the reports relative to the design reviews and case study presentation and on the confidence in the scoring of the various assessments.

Another factor to consider when determining the best approach to use is the effect of a “low F” vs. a “high F”. For example, Team #2 received a 9/20 while Team #3 received a 2/20 on the “final report”. Since both teams received an “F” for this report, there is no difference between the two teams when either of the letter grade approaches is used. However, the total possible points method shows quite a difference: a “D” for Team #4 and an “F” for Team #3.

Finally, the three approaches shown in Table 1 behave differently when there is one extreme grade. For example, if Team #1 happened to fail the second design review (i.e., received 0/4), the following final grades would result:

- median grade = “B+” (no change)
- weighted letter grade = “B” (from a “B+”)
- total possible points grade = “B-” (from a “B+”)

As can be seen, the median grade approach is not unduly influenced by extreme grades. The instructor may decide that this is very appropriate for assessment, especially in the context of engineering design where students are expected to fail at times. Of course, other techniques may be used such as allowing one assessment to be ignored: however, the instructor must be very clear about how this type of policy will be applied.
The choice of grading technique is dependent upon a number of factors such as the types of assessments used, the reliability of the assessments, and the instructor’s views around the treatment of extreme grades. Given the nature of capstone design assessment, and in particular, its difference from typical engineering science classroom assessment, it is important to understand the impact of the choice of grading method.

4. Summary

In this paper we have provided a short overview of assessment in capstone design. We used the general literature on classroom assessment as a basis to attempt to raise some important issues in the area of classroom assessment in capstone design courses. As well, this work was placed in the context of the author’s experience with a capstone design course in mechanical and manufacturing engineering.

The key assessment issues for capstone design noted here are:

- classroom assessment is a challenge in these courses given the open-ended nature of student projects, the use of multiple assessors, and the inherent difficulty in assessing overall quality of a student project;
- capstone design courses lend themselves to a broad range of assessment tools (most notably: performance assessment of processes or products, oral communication, and portfolios);
- reliability is a key concern in capstone design course assessment and hinges most notably on consistency across judges and forms of assessment;
- given the broad range of assessment tools used in capstone design, it is important to take into account the impact of how individual assessments are aggregated into a final grade.

5. References


