AC 2007-2366: CAPSTONE DESIGN PROJECTS WITH INDUSTRY: USING RUBRICS TO ASSESS STUDENT DESIGN REPORTS

Patricia Brackin, Rose-Hulman Institute of Technology

M. PATRICIA BRACKIN is an Associate Professor of M.E. at Rose-Hulman Institute of Technology where she teaches a variety of design courses, and graphics. Her BS and MS are from the University of Tennessee in Nuclear Engineering and her Ph.D. is from Georgia Institute of Technology in ME. She has also been an Associate Professor at Christian Brothers University. Her industrial experience includes Oak Ridge National Laboratories and Chicago Bridge and Iron. She is a registered PE.

J. Darrell Gibson, Rose-Hulman Institute of Technology

J. DARRELL GIBSON is a Professor of M.E. at Rose-Hulman Inst of Tech where he teaches design, noise control, and structural mechanics. His BS and MS are from Purdue in Aero Engineering and his Ph.D. is from the University of New Mexico in ME. He has also been an Associate Professor at the University of Wyoming and a Visiting Professor at Colorado State Univ. His industrial experience includes General Dynamics Corp, J.I Case Co, Sandia Labs, NASA/Langley Research Center, and NASA/Jet Propulsion Lab. He is a registered PE.

Capstone Design Projects with Industry: Using Rubrics to Assess Student Design Reports

Abstract

The benefits of company sponsored capstone design projects, both to academia and to industry, have been well established. Specific benefits to students include the broadening of their engineering skills, the required interaction with practicing engineers, the strengthening of teaming skills by working in design groups, the development of communication skills with required oral and written reports, and the experiences of project management. At the authors' institution these projects are "owned and managed" by the student teams with company contacts providing appropriate data and information and with faculty serving as advisors only. The authors have developed and improved these student/industry interactions over the last few years with over 130 students working with about 30 different companies each year. ABET 2000 requires that graduates demonstrate the ability to design a system, component or process to meet a given need and this capstone design course is a natural place to assess whether or not these outcomes are met.

Outcomes may be assessed by direct or indirect methods. Direct measures of student outcomes are based on student work, and for capstone design courses a natural work product to examine is the design written report. Typically, performance criteria are established and then rubrics are written to insure the consistency of the assessment. The purpose of this paper is to show how such rubrics were developed for senior mechanical engineering design reports and then how they were used by three different groups: the course instructors, other departmental faculty, and outside engineering practitioners. Each of these three groups was given the same set of design reports and then was asked to evaluate the reports by using specifically these scoring rubrics. This paper details the performance criteria, shows the rubrics used, and then reports on the consistency in scoring between these groups.

Introduction

Capstone Design is defined as those senior projects which attempt to provide a culminating experience to students' undergraduate engineering education. Traditionally these projects provide team experiences in utilizing engineering concepts learned in previous courses to provide solutions to real world design problems. This general philosophy is followed in the authors' department but with the following requirements:

1. all project problems are provided by external industrial clients,
2. formal design methodologies must be followed,
3. economics must be considered in design decisions,
4. the improvement of teaming skills must be emphasized, and
5. all design solutions must be communicated through both oral and written reports.

The provision of project problems by external industrial clients is not a simple undertaking but is possible with industrial/academic relationships and partnerships that are fostered over time[1]. These partnerships are essential for technological development, regardless of the discipline. At the authors' institution student design projects with industry are performed as part of course requirements for Engineering Design and for Machine Design, both 4-credit-hour courses. In each course the design project represents 50% of the course grade and the enrollment is approximately 130. The students work in groups of 3 to 5 and are normally expected to spend approximately 4-6 hours per person per week on their projects. Because these courses are taught in two consecutive quarters, the projects last approximately 20 weeks. The teams are required to meet with their instructors weekly and submit written status reports. Formalized design methodologies are emphasized and required. (See for example Pahl and Beitz[2] or Dekker and Gibson[3].)

At the end of the 10 week Winter Quarter the student teams give interim oral presentations and submit interim written reports to their instructors and to their client companies. Final presentations and written reports are due again at the end of the 10 week Spring Quarter. The presentations and written reports are the only deliverables that participating companies are guaranteed. Because every company receives a final design report, these reports have become logical work products to use to evaluate the students' performances.

Rationale

As with all capstone design courses the authors are striving to improve the quality of their design reports and of their courses as a whole. Current methods used in their capstone courses to gather feedback include student evaluations, external client surveys, and the completion of a department mandated feedback sheet used to satisfy accreditation requirements. Although student evaluations and external client surveys contain useful information, it is anecdotal and difficult to use in a systematic way to improve course quality. Therefore the authors have begun to use ABET developed performance criteria, along with appropriate rubrics to gain feedback from both external practitioners and other departmental faculty. The ABET developed performance criteria were selected because they are patterned after the desired design methodology. For example, one of the first actions that students must perform is to understand their client's problem. The entire team is required to visit the client company without any faculty present. This requires that the team generate questions prior to the visit, that they actively listen to their client's concerns, and that they restate the problem in its most general form.

This strategy for design improvement has the following three major objectives:

1. Develop rubrics that clearly describe our course outcomes so that these rubrics can be used by parties external to the course to provide us with meaningful assessments.
2. Use the rubrics to implement peer assessment of the design reports, which has proved helpful in other areas.
3. Use the rubrics to give to students, as a means of communicating to them more clearly how design reports can be assessed, and therefore what they should emphasize.

ABET Criteria

ABET Criterion 3, Program Outcomes and Assessment requires that all engineering programs demonstrate that their students possess "(c) an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability." Consistent with this requirement, the authors' ME Department developed Performance Criteria which are measurable attributes used to define these outcomes. These are stated as:

When given the opportunity, students will:

1. Understand the problem.
2. Develop a product or process design specification that addresses customer/client needs and constraints.
3. Carry out a conceptual design by generating multiple solutions that address the issues above,
4. Evaluate the feasibility of the solutions, and
5. Choose the appropriate solution.
6. Carry out a detail-level design using appropriate design tools and methodologies.
7. Test and refine the implementation until the product or process design specifications are met or exceeded.
8. Document the finished product or process as appropriate for the discipline according to standard practice.

The ME Department desired a method for obtaining external feedback to validate that the quality of the program is being maintained. The intent was not to obtain additional opinions on team grades, but rather to obtain indications of whether or not the performance criteria had been demonstrated. Therefore, the authors decided to use rubrics to determine if desired outcomes are being met. These rubrics were developed to distinguish three levels of performance: "Excellent", "Satisfactory", and "Needs Improvement". In addition, Performance Criterion 8 (above) was modified since outside practitioners may not be aware of the departmental standards for documentation. The rubric therefore was changed to emphasize "conclusions and recommendations". The rubrics that finally were developed to demonstrate these Performance Criteria are listed below and are italicized for emphasis.

1. Understand the problem

Excellent: *The students presented the problem in its most general form and the context for the problem was well defined.*

Satisfactory: *The students presented a problem and gave a general context.*

Needs Improvement: *The students presented a solution rather than a problem and/or the context was not clearly defined.*

2. Develop a Product Design Specification (PDS) that addresses client needs and constraints

Excellent: *The PDS was specific and included both goals and quantifiable, measurable constraints. Applicable codes and standards were appropriately referenced.*

Satisfactory: *The PDS was specific and measurable but a clear distinction was not made between goals and constraints. Applicable codes and standards were mentioned but were not fully referenced.*
Needs Improvement: *The PDS was not specific and measurable. Applicable codes and standards were not referenced.*

3. Generate multiple solutions

Excellent: *The students presented at least three feasible solutions.*
Satisfactory: *The students presented multiple solutions but not all solutions were feasible.*
Needs Improvement: *The students presented only one feasible solution*.

4. Evaluate feasibility of solutions

Excellent: *The students used acceptable techniques such as Pugh Concept Selection or decision matrices to evaluate their design concepts. A rationale was given for every rating.*
Satisfactory: *The students used acceptable techniques to evaluate competing solutions, but no rationale was given for ratings.*
Needs Improvement: *The students do not have a systematic method for evaluating design decisions.*

5. Choose an appropriate solution

Excellent: *The solution was clearly documented based on the authors' feasibility and economic evaluation.*
Satisfactory: *The solution was not fully documented.*
Needs Improvement: *The students had an insufficient rational basis for their recommendation.*

6. Carry out a detail level design using appropriate tools and methodologies

Excellent: *Sufficient detail was presented so that the company can easily implement the solutions. Drawings were sufficient so that a skilled technician could understand the implementation of the solution. Design calculations showed the sufficiency of the proposed solution.*
Satisfactory: *The idea was clear, but additional steps and/or calculations were needed before the solution could be implemented.*
Needs Improvement: *The desired implementation could not be discerned from the information provided.*

7. Test and refine

Excellent: *The product or process was systematically tested and improvements were made and verified.*
Satisfactory: *The product or process was systematically tested and improvements are suggested for future implementation.*
Needs Improvement: *The product or process was not tested.*

8. Present Recommendations
Excellent: *The conclusions and recommendations were complete and logical.*
Satisfactory: *The conclusions and recommendations left some unanswered questions.*
Needs Improvement: *The conclusions and recommendations were confusing and incomplete.*

Method
Inter-rater reliability is often obtained by having raters discuss why they gave certain ratings to individual work products. This study examined the use of rubrics without prior discussion or training to assist in the evaluation of design reports for the purpose of continuous improvement. The authors hoped to obtain meaningful feedback from external sources without putting undue hardship on those external sources. We feel that we will be able to recruit 5-6 external sources each year to help with assessment. Because we routinely have more than 30 projects, each external source will be asked to evaluate 5 reports. What is a reasonable time commitment to ask for from each of our participants? Our initial estimate is that a one hour time commitment is reasonable.

Samples of design reports, in this case five, were used to evaluate the consistency of the rubrics. Outside practitioners (five), other departmental faculty (four), and the two course faculty used these rubrics to evaluate the same reports. The participants in rating were not given any instructions other than to spend no more than one hour on rating the five reports and to indicate any rubrics that they considered confusing. They were also told that the authors expected some of the evaluations to be "needs improvement." The selection of the reports for evaluation included one report judged to be poor by the instructors and one judged to be excellent. No instructor comments or grades were shown on the reports that were evaluated. In order to analyze the results a numeric value was given to each response. A "3" designates a rating of excellent, a "2" designates a rating of satisfactory, and a "1" designates a rating of needs improvement. One rater chose to circle two values for a given criterion and wrote that he felt that the rating was between the two levels. When this occurred, a number equal to the average of the two values was assigned. For example, if the rater circled both a 1 and a 2 indicating that they felt the category was marginally acceptable, the number entered for the criterion was 1.5.

Results
The raw data obtained was evaluated by report and by group. All names of companies and raters have been removed. Table 1 shows the results generated by the course instructors for each report. For example, instructor CA gave a rating of 3 to report R1 for Criterion 1. In addition, it took instructor CA 7 minutes to evaluate R1. Initially, the instructors felt that they would have identical ratings for all questions on all reports. This did not prove to be the case. For example, on report R1, CA rated criterion 1 a 3 and CB rated criterion 1 as a 2. A difference in the rating between instructors indicates that there is not a consistent interpretation of the rubric or that evidence presented for demonstration of the rubric is not consistently interpreted. The instructors gave identical ratings in 21 out of 40 ratings.

Further examination of Table 1 reveals that there are some criteria that are consistently evaluated for all 5 reports. Table 2 shows the number of times that a disagreement in rating exists between the instructors. Row 1 of Table 2 shows the number of occurrences of a numerical difference. The greatest disagreement between the instructors occurred on criterion 4, "Evaluate the feasibility of solutions." The closest agreement occurred for criteria 6, "Carry out a detail level design using appropriate tools and methodologies," and 7,"Test and refine." For purposes of program improvement it is essential to identify reports that need improvement. Tabulation of the number of occurrences where the instructors disagreed concerning whether or not the students needed improvement is shown in Row 2 of Table 2. The instructors agreed on criteria 3, 5, 6, and 7. There was one difference in the ratings for criteria 1, 2, and 8. There were four differences in the criteria for criteria 4.

Table 1: Results from Instructor Ratings

Report	R1								Time
Criteria	1	2	3	4	5	6	7	8	(min)
Instructor									
CA	3	1	3	1	2	2	1	3	7
CB	2	1	2	2	3	3	1	3	8
Mean	2.50	1.00	2.50	1.50	2.50	2.50	1.00	3.00	7.50
σ	0.71	0.00	0.71	0.71	0.71	0.71	0.00	0.00	0.71
Report	R2								
CA	2	1	2	2	2	2	1	3	6
CB	2	2	3	1	2	2	1	2	8
Mean	2.00	1.50	2.50	1.50	2.00	2.00	1.00	2.50	7.00
σ	0.00	0.71	0.71	0.71	0.00	0.00	0.00	0.71	1.41
Report	R3								
CA	3	3	3	2	2	3	2	3	6
CB	3	3	3	3	3	3	3	3	10
Mean	3.00	3.00	3.00	2.50	2.50	3.00	2.50	3.00	8.00
σ	0.00	0.00	0.00	0.71	0.71	0.00	0.71	0.00	2.83
Report	R4								
CA	3	2	2	1	3	2	1	1	5
CB	2	3	3	3	3	2	1	2	8
Mean	2.50	2.50	2.50	2.00	3.00	2.00	1.00	1.50	6.50
σ	0.71	0.71	0.71	1.41	0.00	0.00	0.00	0.71	2.12
Report	R5								
CA	3	1	1	2	2	1	1	2	5
CB	1	1	1	1	2	1	1	2	6
Mean	2.00	1.00	1.00	1.50	2.00	1.00	1.00	2.00	5.50
σ	1.41	0.00	0.00	0.71	0.00	0.00	0.00	0.00	0.71

In addition to comparing the two instructor ratings, comparisons were made between the three groups for each rubric. Because the data is non-parametric, the Mann-Whitney test was used. A number less that 0.05 indicates a 95% confidence level that there is a significant difference in how the two groups rated the performance criterion. These comparisons are shown in Table 3. Table 3 indicates that there was a significant difference in rating between practitioners and other faculty for criteria 1, 2, 4, and 5. There is no significant difference in the rating for any criteria when practitioners are compared to course instructors and when other faculty are compared to course instructors.

Table 2: Number of Times that Instructors Differ for a Given Criteria

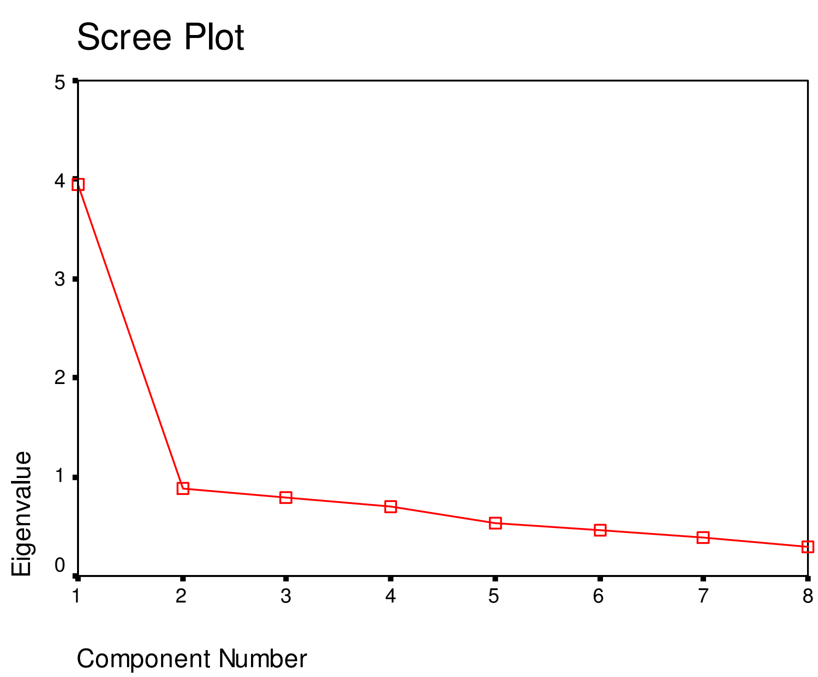
Criteria	1	2	3	4	5	6	7	8
Number of Differences	3	2	3	5	2	1	1	2
Difference in Rating Needs Improvement	1	1	0	4	0	0	0	1

Table 3: Comparisons between Groups to Test for Differences in Rating Each Rubric

Question	1	2	3	4	5	6	7	8
Practitioners compared to other faculty	.003	.021	.728	.012	.000	.321	.364	.071
Practitioners compared to course instructors	.381	.223	.752	.137	.322	.445	.467	.985
Other faculty compared to course instructors	.270	.733	.577	.679	.083	.900	.815	.270

A Scree Plot was generated to examine the correlation between the test questions. The result is shown in Figure 1. The plot indicates that the way in which raters answered question one was strongly correlated to the manner in which the raters answered all other questions.

Figure 1: Scree Plot Indicating Correlation between Performance Criteria



Finally, the average time that raters took to review the reports was 10.9 minutes with a standard deviation of 5.1 minutes.

Several raters found aspects of the rubrics confusing. This feedback is summarized below:

1. When you say, "State the problem in the "most general form," I'm not sure how to judge that.
2. I don't really understand what you mean by "feasible" when presenting solutions.
3. When you say, "Recommendations were complete and logical," what do mean?
4. Under *generate multiple solutions* an excellent indicates that the students presented three feasible solutions. A satisfactory indicates that the students presented multiple solutions, but not all solutions were feasible. I don't know what to do if the students presented two feasible solutions.
5. For rating the Product Design Specification, what should I rate if part of the PDS was specific and measurable, but not all?
6. I found that students included elements of satisfactory and needs improvement for the Product Design Specification

Recommendations and Conclusions

Having established the benefits of capstone design projects with industry, the authors have been attempting to continuously improve methods for the evaluation of the projects, particularly the written reports. Rubric, therefore, have been written to facilitate these evaluations and to provide feedback, particularly from outside practitioners.

First, the time required for using rubrics to evaluate design reports is not prohibitive. External practitioners have indicated their willingness to spend up to an hour helping the authors assess senior design reports. The average time of approximately 11 minutes to evaluate a report means that external practitioners can evaluate several reports within an hour.

Second, the current rubrics are not satisfactory for obtaining consistent results without inter-rater reliability. Refining the rubrics or instituting a method for insuring inter-rater reliability will be necessary.

This paper has presented how the quality of student design reports required in rather large sections of senior capstone design courses can be improved by using rubrics. This strategy, together with the rubrics written and used, is consistent with the requirements of ABET 2000. Although some rubrics can be improved, the authors are quite encouraged with the support from external practitioners and other departmental faculty and plan to continue this strategy. Discussion of feedback on the rubrics has already led to suggestions for course improvement.

Bibliograpy

[1] Brackin, P.D. and Gibson, J.D., "Techniques for the Implementation, Administration, and Evaluation of Industrially Sponsored Capstone Design Projects in the ME Curriculum", Proceedings of 2006 ASME International Mechanical Engineering Congress and Exposition, Chicago, Illinois, November, 2006.

[2] Pahl, G. and Beitz, W., Engineering Design: A Systematic Approach, Edited by Ken Wallace, Springer-Verlag, The Design Council, 1988.

[3] Dekker, D., and Gibson, J. D., "Learning Design in a College Setting", International Conference on Engineering Design, Tampere, Finland, August, 1997.