What's a Good Project? Comparing Student, Faculty and External Evaluations

Bridget M. Smyser
Northeastern University

The goal of this study was to compare student, faculty, and alumni jury evaluations of the same set of 30 projects. Students were asked to identify the most innovative project at a 'prototype preview' day. Faculty assessed communication, project management, and completeness of delivered design. A jury of alumni from industry evaluated the final posters and presentations, particularly focusing on projects that most completely solved the problem or that had the biggest potential impact. Scores and evaluations from all three sources were compared using Pearson's Product Moment correlation and t-tests to determine statistically significant differences. The projects favored by each type of evaluator were compared to look for similarities and differences between the three groups. A positive correlation ($R = 0.47$, $P=0.008$ at $\alpha = 0.05$) was found between high jury scores and high peer scores. There were also significant differences between the top 10 and bottom 10 projects for the 3 evaluators for all criteria, with the exception of prototype score for the alumni jury. Agreement among the evaluators is more pronounced for low ranking projects than high ranking projects. Results indicate that the three viewpoints, while noticeably different in some respects, collectively can be used to identify innovative, well managed, and well communicated projects.

Keywords: External evaluation of capstone, capstone assessment, peer assessment

Corresponding Author: Bridget Smyser, b.smyser@northeastern.edu

Introduction

Capstone design courses commonly use multiple types of reviewers to evaluate capstone projects.[1] A 2004 paper by McKenzie et.al. surveyed a number of institutions and found that on average capstone faculty participate in 90% of course assessment activities, industry sponsors in 68% of assessments, and students in 70% of assessments.[2] Their study focused specifically on ABET requirements that were concerned with communication, recognizing wider contexts, and the ability to solve complex engineering problems.[3] Although this paper was based on earlier ABET requirements, the current student outcomes 1, 3, and 4 cover similar topics.[4] The authors' view was that different stakeholders offered valuable alternate perspectives that aid in completely evaluating capstone projects.

Beyerlien et.al. described three different perspectives offered by various stakeholders in capstone design courses.[5] Professors take on the role of 'educational researcher' and assess projects using knowledge of what the discipline requires, an appropriate set of criteria for assessment, and a method to interpret the results as a way of ranking performance. External reviewers from industry act as 'professional practitioners' They know what is expected in current professional engineering settings, and often focus explicitly on soft skills required for professional success. Student learners must not only receive the criteria for evaluation, but also apply it to their own projects and recognize how projects do or do not meet the criteria. These three stakeholders all have valuable contributions to make in the context of design assessment.

The perspectives and validity of assessments by these cohorts have been studied repeatedly. One study found a high correlation between student self-assessment results and faculty assessments.[6] Other studies indicate that peer evaluations may be more valid than self evaluations in teamwork situations.[7] A study focused on grading written reports compared external judges and faculty scores and found that external judges tended to give higher grades based on perceived project success.[8] However, faculty rewarded academic success, without strong consideration of project success on its own. This relates to anecdotal evidence the author has encountered: students complain that the jury selects good 'salesmen' instead of projects that were much more difficult to build and describe.

The literature seems to indicate that despite both being qualified, faculty and external judges tend to disagree on what constitutes a 'good' project. Students are qualified to evaluate peers and tend to agree more with faculty assessment. The current study compares evaluations by all three groups to determine if there is consensus on what constitutes a successful project.

Course Information

ME4701/4702 Capstone Design is a two-semester sequence. Teams of 4-5 students work on projects that can be proposed by faculty, industry sponsors, or students. At a point 2-3 weeks before the end of term, teams must write an Executive Summary that is sent to the alumni jury in advance of the final presentation day. The state of the prototypes is assessed by the course coordinators at this point and given a prototype score, worth 10 points. This scoring scheme is described more fully in a previous paper[9]. A high score here indicates a project that has been well managed and is likely to be complete and validated by the end of term.

Approximately 1 week before the final presentation day, the students participate in a 'project preview' day which allowed students to demonstrate their prototypes to the other students in an informal session. Each student was asked to choose their top three projects in an anonymous poll, specifically considering which projects demonstrated innovation and 'out of the box' thinking. The project with the most votes received the department's Gorlov Innovation Award.

On the last day of the course, final prototypes were presented in a poster session, and in oral presentations presented in 3 concurrent sessions. Projects were evaluated by a jury of 6-9 alumni per session. The jury for each session chooses two projects to receive awards: one for the project that is the Most Complete and one for the project with the Biggest Impact. Jury members also fill out evaluations of the program and note projects that were particularly successful or unsuccessful.

At the end of term, the advisor for each team assigns a final grade for that team. This is modified based on input from the course coordinators. One form of input is the writing grade, which takes in performance on written and oral reports as well as the Executive Summaries and the final poster presentation. Another coordinator input is the 'Delivered/Initial' score, worth 10 points. This score is a measure of how well the final delivered prototype satisfies the initial goals and problem statement as defined by the team and their advisor and/or sponsor at the beginning of the project.

Research Questions

The present work seeks to answer the following questions:

- Do students, jurors, and faculty identify the same projects as successful or unsuccessful?
- Are there significant differences between high and low ranked projects?
- Is there evidence that the different groups of evaluators favor different types of projects?

Methods

Student ranking of projects was determined by counting the number of votes each project received during the voting for the innovation award. Each student was allowed to vote for 3 projects. This prevents each team from voting only for their own project and ensures that the best projects truly rise to the top of the list.

Faculty assessment was based on the total of the prototype, delivered/initial, and writing scores as a percentage of the total possible points. A perfect faculty assessment score would be 120 points. The prototype, delivered/initial, and writing scores were also compared individually to other scores to determine if certain aspects of the faculty assessment were strongly correlated with assessments from other sources.

Jury assessment was based on reviewing individual juror's evaluation sheets. Projects that were mentioned by name as 'successful' or 'showing consideration of wider impact' were given 1 point per mention. Projects that were mentioned by name as 'unsuccessful' or 'conspicuously lacking in consideration of wider impact' were given -1 point per mention. This accounts for projects that were mentioned positively by one judge and negatively by another judge.

Pearson's product-moment correlation analysis, single factor ANOVA, and t-tests assuming unequal variances were used to compare the results of different evaluator groups. Lists of highly and poorly rated projects were compared to determine any similarities.

Results

The top 10 and bottom 10 projects as ranked by students, jury, and faculty were examined. In the top 10 projects, 4 projects were common to all 3 evaluators. For the bottom 10 projects, 6 projects were common to all 3 evaluators. Table 1 shows the number of projects chosen by each evaluator group sorted by general topic. For all 3 evaluators top projects were heavily weighted toward space and medical/assistive projects. The jury and faculty evaluators each identified service projects in their top 10, while the students did not. The bottom 10 projects for all 3 evaluators had many research and automotive projects.

Table 1: Top 10 and bottom 10 project topic distribution between evaluators

Top 10 project topics			
# of Projects	**Students**	**Jury**	**Faculty**

Space related	5	3	3
Medical/Assistive	4	4	4
Consumer Product	1	0	1
Service	0	2	3
Research	0	1	0
Automotive	0	0	0
Bottom 10 project topics			
# of Projects	**Students**	**Jury**	**Faculty**
Space related	0	0	0
Medical/Assistive	1	0	1
Consumer Product	1	1	1
Service	0	0	1
Research	5	6	4
Automotive	3	3	3

There were statistically significant differences between the top 10 and bottom 10 projects for all criteria with one exception. The only case where no statistical difference was found was between the top 10 and bottom 10 prototype scores for projects rated by the jury. The average prototype scores for the jury were similar for both the high and low scoring groups.

Pearson's product moment correlation analysis was conducted to determine correlation between individual measures. Table 2 shows the results, listing only statistically significant correlations. Most strongly correlated were the faculty scores, writing grades and the delivered/initial score. The prototype score, which favors projects with good project management and early completion of tasks, is positively correlated with the writing score, delivered/initial score, and overall faculty score. The delivered/initial score, indicating a solution that satisfies the initial goals, is positively correlated with student votes, writing score, prototype score, and overall faculty score. The number of jury mentions, indicating projects/teams that were able to communicate their success to educated outsiders, is correlated with student votes, writing grade, and overall faculty score. The number of student votes, which favors innovative solutions, was correlated with jury mentions, delivered/initial score, writing score, and overall faculty score. The overall faculty score considers communication scores throughout the course, project management, project progression, and project completeness. This overall faculty score is positively correlated with all other measures.

Table 2: Statistically significant differences between measures for all projects

Correlation	**Pearson's R**	**P (α=0.05)**
Writing Grade/Faculty score	0.97	0
Delivered/Initial/Faculty score	0.73	0
Writing grade/ Delivered/Initial	0.61	0.003
Faculty score/Prototype score	0.60	0.0003
Student votes/Jury mentions	0.47	0.008
Student votes/Faculty score	0.47	0.008
Student votes/Writing grade	0.46	0.01
Writing grade/Jury mentions	0.45	0.01
Writing grade/Prototype score	0.43	0.005
Delivered/Initial /Prototype score	0.43	0.02
Faculty score/Jury mentions	0.43	0.005
Student votes/Delivered/Initial	0.40	0.03

Discussion

The three evaluators had slightly less agreement about good projects versus poor projects. All three evaluators were impressed by projects that were highly complete and had a wider impact or that solved problems that would help a particular person or group of people. For example, one group developed an adaptive device to allow a quadriplegic father to control a pitching machine in order to play baseball with his son. The team not only created a functional and verified prototype but did it with elegant and innovative controls. This project received both the student-chosen 'Gorlov Innovation Award' and the 'Biggest Impact' award in their presentation session. Overall, students and the jury were impressed with projects that were space related (Mars Rover competition projects or projects with NASA JPL) or that had complicated controls or robotics challenges.

Faculty favored space related projects, but also strongly favored projects with simple and elegant solutions. One such project involved a device to mechanize aspects of cocoa harvesting for small farmers in Ghana. Although the controls and overall sophistication of the device were relatively simplistic, the device clearly solved a real problem. Moreover, one of the students on the team was the daughter of a cocoa farmer in Ghana, and her passion for helping her country spread to her whole team, giving them a clear reason for their project. Faculty and student evaluators tended to favor projects with high prototype scores. Jury favored projects did not show any clear patterns with regards to prototype scores.

Projects ranked poorly by all three evaluators had similarly low writing, prototype, and delivered/initial scores. One set of projects that was universally panned were sponsored by the student SAE Formula One Hybrid team. The three projects sponsored by this team were seen as too constrained, not complete, and not particularly challenging. All three evaluators had the same mix of project sources in the bottom 10: 40% student proposed, 40% faculty proposed, and 20% industry proposed. One low ranked project seemed to be an anomaly. This project was low rated by jury, faculty, and students, but it also won the 'Most complete' award in their presentation session. At their final presentation, however, they were able to convince the group of jury members of the project's completeness. Historically, some students and others have believed that poor projects with good 'salesmen' win over projects that are challenging and difficult to explain. For the current group of projects, that did not seem to be true, with perhaps the exception of this one team.

Conclusions

Certain measures such as writing score and overall faculty score are highly correlated with success. This score tends to be comprehensive, taking into account the entire process. Communication skills are highly valued by ABET and employers, so the fact that these measures contribute strongly to project success is encouraging.

The argument could be made that since the faculty measures are comprehensive and predict success, there may be less use for peer and external reviewer feedback. However, peer feedback is valuable for recognizing work that is innovative even in its uncomplete state. The prototype scores, which measure early project completion, do not correlate with student votes. Students are able to recognize innovation, effort, and complexity of a problem without it being finished and wrapped for presentation. Students in the lab at all hours are often the only witnesses when a breakthrough is made.

The jury was not heavily influenced by a highly complete project. They were influenced by teams that explained and communicated the value of their project to a viewer who has little foreknowledge of it. This perspective cannot be achieved in any other way.

High quality projects tend to be recognized as such by all three evaluator pools. Poor quality projects cannot be hidden through good salesmanship. However, the different nuances of all three evaluators provides a more complete and rounded evaluation of projects at multiple levels and at multiple times during the term. This points to the continuing value of seeking student, faculty, and industry perspectives in evaluating capstone projects.

References

1. Dutson, Alan J., et al. "A review of literature on teaching engineering design through project-oriented capstone courses." Journal of Engineering Education 86.1 (1997): 17-28.
2. McKenzie, Larry J., et al. "Capstone design courses and assessment: A national study." Proceedings of the 2004 ASEE Annual Conference & Exposition. 2004.
3. ABET. 2004. Criteria for Accrediting Engineering Programs. Engineering Accreditation Commission, Accreditation Board for Engineering and Technology, Baltimore, MD. Worldwide web address: http://www.abet.org
4. ABET. 2019. Criteria for Accrediting Engineering Programs. Engineering Accreditation Commission, Accreditation Board for Engineering and Technology, Baltimore, MD. Worldwide web address: http://www.abet.org
5. Beyerlein, Steve, et al. "Assessment framework for capstone design courses." Proceedings of the ASEE Annual Conference & Exposition. 2006.
6. Hylton, Pete, and J. Anderson. "Industry Based Assessments in an Engineering Technology Capstone Design Course." Proceedings of the ASEE IL/IN & North Central Joint Section Conference. 2006.
7. Thompson, Robert S. "Relative validity of peer and self-evaluations in self-directed interdependent work teams." 31st Annual FIE Conference. Impact on Engineering and Science Education. Conference Proceedings (Cat. No. 01CH37193). Vol. 1. IEEE, 2001.
8. Lyerly, A. N., & Dixon, G. (2016, June), Grading the Capstone Written Design Reports: A Comparison of External Judges and Faculty Scores Paper presented at 2016 ASEE Annual Conference & Exposition, New Orleans, Louisiana. 10.18260/p.25416
9. Kowalski, G. J., and B. M. Smyser. "Assessing the Effect of Co-op Sequence on Capstone Design Performance." Proceedings of the ASEE Annual Conference, Vancouver, BC. 2011.