Longitudinal Assessment of External Experts and Teaching Assistants as a Class Resource

Mr. Aldin Malkoc, Arizona State University

Aldin Malkoc, MS is a student in the School of Biological and Health Systems Engineering at Arizona State University. Aldin is enrolled in the 4+1 program to receive his Masters of Science in Biomedical Engineering from Arizona State University in 2017 and will pursue a doctoral degree in Biomedical Engineering from Arizona State University in 2017. The primary focus in his master’s thesis will pertain to the study of a point-of-care insulin sensor in biosensor development. Currently, Aldin is a graduate teaching assistant at the Fulton Schools of Engineering and wishes to develop effective engineering education strategies.

Ms. Mackenzie Honikel, School of Biological and Health Systems Engineering at Arizona State University

Mackenzie Honikel, a current PhD student in the School of Biological and Health Systems Engineering at Arizona State University. Mackenzie graduated from SUNY Binghamton in May 2016 with a Bachelor’s degree in biomedical engineering, concentrating in biomedical devices and biomaterials. Her background is in point-of-care diagnostic sensors, and she aims to continue this work during her time at Arizona State University. Her dissertation focuses on the development towards implantable cardiovascular sensors for continuous patient monitoring and reduced embolism formation at the site of implantation.

Dr. Jeffrey Thomas La Belle, Arizona State University

Jeffrey T La Belle is currently an Assistant Professor in the School of Biological Health and Systems Engineering and the Biodesign Institute at Arizona State University. He holds adjunct status in the School of Energy and Matter Transport (Mechanical Engineering) as well as the College of Medicine at Mayo Clinic. He has a Ph.D. and Masters in Biomedical Engineering from ASU and a MS and BS in Electrical Engineering from Western New England University in Springfield Massachusetts. In the La Belle Group, we are currently developing electrochemical sensors for noninvasive glucose sensing, the novelty of our design is to obtain tear fluid for tear to blood glucose correlation in a noninvasive means to increase patient compliance. The next leap in technology for diabetes care is a multiplexed sensor that will add more depth of information for a self-monitoring blood glucose devices, here five accepted markers for DM care and management, including glucose, HbA1c, among others are simultaneously monitored on a single strip sensor. This technology we are developing could also allow for continuous and single use stress/trauma sensing technologies. Other applications of the sensing technologies include small molecule, DNA, protein, and whole cell detection to address changing climate in point-of-care technologies and medicine. On the activation side of our research, we are fabricating nitinol staggered muscle arrays that mimic skeletal muscle and we have recently demonstrated over 30% compression in our SMA’s similar to muscle bundles. Our approach to design is simple, following FDA guidelines and suggestions from the start, look at what the user needs and/or wants and apply a unique solution. We have a well-diversified group to tackle the challenges in health care today, staff and students come from biomedical engineering, electrical engineering, mechanical engineering, chemical engineering, computer science engineering, as well as biology and chemistry programs at ASU. BME at ASU teaches a 8 semester wide medical device design tract that initiates the students in design, regulations, standards, IP and other aspects from day 1. Dr. La Belle has develop and courses and taught at the freshman, junior, senior and graduate level on these topics.
Longitudinal Assessment of External Experts and Teaching Assistants as a Class Resource

To help with instructor teaching and student learning in STEM courses, various methods such as two-way formative feedback, flipped classrooms, and project-based learning have been used to enhance student learning, participation, attitude, and overall achievement. This longitudinal study, conducted over four consecutive semesters, assesses students' skills in innovation, prototyping, and design through a project-based course. Students were given 16 weeks to develop a functional gamma prototype that would be scored by external experts and teaching assistants. Furthermore, the scores would be compared to actual performance on a design challenge (DC). The gamma prototype is a board game that functions as a teaching tool of the prototyping pathway for medical device design. The use of external experts and teaching assistants for validation asked the following research question: “Can external experts and teaching assistants effectively evaluate student performance in (1) innovation, (2) prototyping, and (3) design?”

The focus of the course is the development of a gamma prototype board game as a surrogate for enhancing skills needed in medical device design. In the course, students are placed into groups, and while completing the board game; students are encouraged to consider aspects of innovation, product design, prototyping, and fabrication. Throughout the 16-week course, students learn various bioengineering design, analysis, and decision-making techniques as well as the fundamentals of business and technical management. Additionally, student’s actual performance in innovation, prototyping, and design is measured based off the instructors scoring of the DC.

The board game prototype assessment by the external experts and teaching assistants is scored and grouped into innovation, prototyping, and design. The study compares student performance on the DC to scores given by external experts and teaching assistants. By comparing students’ actual performance on the DC to the prototype score, the grading accuracy of external experts and teaching assistants can be evaluated. This junior-level biomedical engineering course is tasked with not only teaching important concepts, such as innovation, prototyping, and design, but tighter evaluation of the learned concepts.

The data generated across four semesters showed statistical significance. Teaching assistants across semesters typically ranked students higher in the three categories than external experts. The percent change seen between external experts and teaching assistants was 19% for innovation, 26% for prototyping, and 15% for design. Furthermore, coefficient of variance for external experts was on average below 5% as compared to 10% for teaching assistants. The results from this study offer insight into the benefits of including external experts, as a resource to better evaluate student performance in innovation, prototyping, and design.
Introduction

Many resources are made available to engineering students in hopes of succeeding in their engineering courses. A survey was performed on 13 courses in the biomedical engineering department at Arizona State University in the hopes of determining key resources students use in preparation for exams and comprehension of key concepts. Currently there have been various research assessments on students’ use of resources and longitudinal performance in various engineering classes. Engineering is a profession that asks students to learn theory in a wide range of science courses and demonstrate application of the knowledge they have developed. In addition, engineering students must learn technical skills, such as, problem solving capacity and logical thinking abilities. According to The Taxonomy of Educational Objectives (Cognitive Domain) by Bloom, et al, there are six levels of expertise with an emphasis on proficiency in analysis, synthesis, and evaluation. Within the last decade, there has been a push for all six levels to be incorporated into engineering majors to ensure that students are learning, retaining, and applying material efficiently in their core engineering classes. However, the missing need is having a resource, such as external validators and teaching assistants, effectively assess and ensure students’ proper application of engineering concepts are applied.

Research conducted in the 1990’s suggests that active teaching by the instructor will increase student engagement, which results in a better understanding of material. Other studies performed show that active teaching compared to a traditional style of teaching results in higher grade point averages. A study performed in an engineering course at Utah State University found that students taught via project based learning received equal or higher conceptual understanding as students in a traditional lecture-based course, as measured by concept quizzes after the end of every lecture and a single concept test. Furthermore, the authors want to ensure that the use of external validators is a valid resource for enforcing student knowledge of innovation, prototyping, and design. The three skills are highly important to the development of future engineers. Innovation is learned implicitly as students progress through their engineering core classes. Furthermore, prototyping and design are extremely worthy skills for students to develop as they progress through this course. Lastly, literature shows that innovation, prototyping, and design are highly regarded as skills to learn in an engineering senior design experience.

The objective of the study is to identify if external validators and teaching assistants are useful in identifying students’ proficiency in specific areas such as innovation, prototyping, and design. In order to generate a comparable benchmark, the instructor scored students using an interactive design challenge with the same metrics in mind. We hypothesize that students’ design challenge scores would be similar to both external validators and teaching assistants scores on the gamma prototype. This comparison can help determine if external validators and teaching assistants can accurately assess student’s capabilities and


skills in a project-based learning course. This would further encourage instructors and departments to consider using external experts and teaching assistants for evaluation of student performance in such courses.

**Materials and Methods**

The course uses a project-based learning style to teach and enforce three main ideas; innovation, prototyping, and design. The course focuses on the process behind developing a gamma prototype, specifically of a zombie themed board game, while enforcing principles utilized in medical device design. The longitudinal study determines if external validators and teaching assistants can efficiently assess students’ understanding and integration of critical design aspects.

*Please refer to Appendix 1 Figure 10 for direct comparison made between a board game prototype and a medical device prototype.*

**Gamma Prototype**

Students are posed with the task of developing a final prototype board game using various engineering skills such as Computer Aided Design (CAD) and machining tools. The students are given complete freedom over the development of the game but are required to develop rigorous explanation for all design additions to the board game. Through this process students learn statistical analysis, design of experiment techniques, decision-making techniques, and fundamentals of business and technical management. Additionally, aspects of technical, regulatory, economic, legal, social, and ethical aspects are employed and asked to of students to consider when developing the prototype.

**Design Challenge (DC)**

The design challenge is a single page handout that tested student’s skills in innovation, prototyping, and design. The problem asked students to develop a diagnostic method to detect an infectious agent. The design challenge was given both the 1st week and the 16th week of the semester and students were given 15 minutes to complete. Only the final design challenge score was used for comparison in results. The design challenge was scored on a 5-point scale with a multiplying scale factor of 2 to match the prototype scale. It should be noted that the instructor of the course scored the design challenge.

**Prototype Scoring**

At the end of the course, students showcased their final prototype. Experts ranging from business game owners to board game developers assessed the students’ final gamma prototype. The external experts and teaching assistants graded on a 10-point scale; please see Table 1. The grade sheet consisted of 10
questions that were grouped into the three key topics; (1) innovation, (2) prototyping strategies, and (3) design.

Table 1: Questions used to evaluate student’s board games and grouping of questions into Innovation, Prototyping, and Design.

<table>
<thead>
<tr>
<th>Innovation</th>
<th>Prototyping</th>
<th>Design</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>What is the genre and targeted population of the game?</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td>X</td>
<td>What is the potential market size of the game?</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>Did the company plan ahead for success and future growth?</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>What are the manufacturing costs for the game? (How do these costs compare to market average)</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>What percentages of these costs are focused on the pieces, board, labor? (Do the costs adequately reflect the quality of the game?)</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td>What is the manufacturing suggested retail price, MSRP? (Is this price appropriate for the targeted consumer? Below/above price range?)</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>How easy are the instructions to comprehend? (How long are they and is there complicated language? If the targeted consumer is very young or very old, will they need frequent assistance in understanding how to play the game?)</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>On average, how many hours of game play can the consumer expect? (Is it likely to hold their interest throughout this duration, or will they quickly become bored?)</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td>What is the minimum, average, and maximum number of players who can play this game at one time?</td>
</tr>
</tbody>
</table>
| X          |             |        | Is the game content (pieces, story, duration) appropriate for the target audience? (Has the
company given thought to industry standards and taken steps to properly regulate themselves?)

**Statistical Approach**

The Kruskal Wallis, non-parametric, test was used to analyze the data. Excel was used to compile all the results and generate descriptive statistics and plot graphs. Statistical Package for the Social Sciences (SPSS) was used to analyze the inferential statistics. Post Hoc analyses were performed using Langley, R. *Practical Statistics Simply Explained* (p220)\(^4\) to test for all various combinations.

The significance analysis was performed using Table 2. Rank Total represents the mean rank value output from SPSS. The mean rank value is compared to the K value in Table 2. For the analysis performed, \(p=0.05\) in the first group was used. This effectively compared any groups in pairs versus comparing to a control.

Table 2: Analysis used to determine statistical significance with the equation used to calculate the K value based off the Mean rank (Rank Total).

\[
    K = \frac{\Delta (\text{Rank Total}) - 0.8}{N * \sqrt{N}}
\]

<table>
<thead>
<tr>
<th>Total number of groups in the analysis</th>
<th>When comparing any groups with each other in pairs</th>
<th>When comparing several groups with a control group but not with each other</th>
</tr>
</thead>
<tbody>
<tr>
<td>K Needs to be equal to or greater than the table value, for it to be significant at the level indicated by p</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>p = 0.05</th>
<th>p = 0.01</th>
<th>p = 0.05</th>
<th>p = 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.89</td>
<td>3.60</td>
<td>2.72</td>
</tr>
<tr>
<td>4</td>
<td>4.22</td>
<td>5.12</td>
<td>3.86</td>
</tr>
<tr>
<td>5</td>
<td>5.6</td>
<td>6.69</td>
<td>5.00</td>
</tr>
<tr>
<td>6</td>
<td>7.01</td>
<td>8.30</td>
<td>6.17</td>
</tr>
<tr>
<td>7</td>
<td>8.46</td>
<td>9.92</td>
<td>7.37</td>
</tr>
<tr>
<td>8</td>
<td>9.94</td>
<td>11.58</td>
<td>8.55</td>
</tr>
<tr>
<td>9</td>
<td>11.43</td>
<td>13.25</td>
<td>9.77</td>
</tr>
<tr>
<td>10</td>
<td>12.97</td>
<td>14.95</td>
<td>11.01</td>
</tr>
</tbody>
</table>
Results

Statistical Evaluation of Semester Scores by Validators and Design Challenge

Over the course of four semesters a total of 190 students completed both a design challenge and gamma prototype that was assessed by industry experts and teaching assistants. The following figures show semesters of scored data reports. It should be noted that there was 100% student involvement in the gamma prototype presentation each semester. The average from external experts, teaching assistants, and design challenge scores is presented in the figures.

Figure 1. Investigation of Spring Semester 2014: Kruskal-Wallis analysis of the three data groups (Innovation, Prototyping, Design) showed different statistical differences. Statistical (+ p<0.05, n=20, Error shown is Standard Error) significance was found in the innovation, prototyping, and design category for teaching assistants scores.

Figure 1 shows a comparison between scores generated by external experts, teaching assistants and the design challenge during the spring 2014 semester. There was statistical difference between the teaching assistants scores and external experts and design challenge scores. In the innovation category, teaching assistants scored students with a 9.68 while the average design challenge score was a 6.16. In the prototyping category teaching assistants scored students with an 8.69 while the design challenge score was a 6.05. Lastly, in the design category teaching assistants scored students with a 9.38 while the design challenge score was a 6.43. Additionally, the coefficient of variance for teaching assistants in each category respectfully was 3%, 9%, and 6%. In almost all of the cases the score difference between external experts and the design challenge scores was only 4.3%.
Figure 2. Investigation of Fall Semester 2014: Kruskal-Wallis analysis of the three data groups (Innovation, Prototyping, Design) showed statistical (+ p<0.05, n=20, Error shown is Standard Error) significance in the innovation, prototyping, and design category when comparing teaching assistant scores and external experts scores to the design challenge.

Figure 2 shows a comparison of external experts and teaching assistants to the design challenge in fall 2014 semester in the categories of innovation, prototyping, and design. There was statistical difference in all three categories between design challenge scores and both validator scores. In all three categories the average score of external experts and teaching varied no more then 28% from the design challenge and the coefficient of variance was less then 5%. For instance, the innovation category external experts ranked students with an 8.15 while teaching assistants gave a 7.94. The design challenge score was a 6.32 and was statistical different from both external experts and teaching assistants. Additionally, the coefficient of variance in the innovation group for external experts was 3% and teaching assistant 5%.
Figure 3. Investigation of Spring Semester 2015: Kruskal-Wallis analysis of the three data groups (Innovation, Prototyping, Design) showed no statistical (p>0.05, n=20, Error shown is Standard Error) significance in the innovation, prototyping, and design category.

Figure 3 shows a comparison of external experts and teaching assistants to the design challenge in the spring 2015 semester in the categories of innovation, prototyping, and design. This semester there was no statistical difference of the score between the three groups in any of the categories. However, it should be noted that the max difference between external experts and teaching assistants was 8% and was in the design category. Additionally, the coefficient of variance for external validators and teaching assistants was no more then 5%. The low scoring difference and coefficient of variance would imply that the scores given to students on their prototypes were closely similar to their actual performance.
Figure 4. Investigation of Fall Wednesday Semester 2015: Kruskal-Wallis analysis of the three data groups (Innovation, Prototyping, Design) showed statistical (+ p<0.05, n=20, Error shown is Standard Error) significance in the innovation, prototyping, and design category when comparing teaching assistant evaluation and external experts evaluation to the design challenge.

Figure 5. Investigation of Fall Friday Semester 2015: Kruskal-Wallis analysis of the three data groups (Innovation, Prototyping, Design) showed statistical (p<0.05, n=20, Error shown is Standard Error) significance in the innovation, prototyping, and design category when comparing teaching assistant evaluation and external experts evaluation to the design challenge.
Figures 4 and 5 show a comparison of external experts and teaching assistants to the design challenge in the fall 2015 Wednesday and Friday class. The fall 2015 semester overall observed statistical difference between the design challenge score and both validator scores. Additionally, it should be noted that the max score difference of external experts and teaching assistants was 6% and was in the fall 2015 Wednesday design category. However between groups the scoring from the external experts and teaching assistants was very similar as shown by the coefficient of variance being no higher then 5%.

Statistical Evaluation of Average Semester Scores by Validators and DC

Figure 6. Investigation of Semester Average Validator and DC Scores:
Kruskal-Wallis analysis of the three data groups (Innovation, Prototyping, Design) showed only statistical (\( p<0.05, n=100 \), Error shown is Standard Error) significance in the innovation, prototyping, and design category when considering teaching assistants.

Figure 6 takes into consideration the average score across all five semesters. The average of all semester scores for external experts and teaching assistants is compared to the average of all semester design challenge scores. The data shows the only statistical difference observed between teaching assistant scores and the design challenge score in all three categories: innovation, prototyping, and design. Furthermore, the difference in external experts scores to design challenge was no higher then 20% while the coefficient of variance was no higher then 4%. On average external experts seem to more accurately evaluate student performance in innovation, prototyping, and design then teaching assistants.
Statistical Evaluation of Longitudinal External Validation Scores

Figures 7 through 9 show longitudinal graphs of external experts scores on the gamma prototype. The three graphs show semester scores for innovation, prototyping, and design. Within the innovation group spring 2014 is statistically different from spring 2015 and fall Wednesday 2015 section is statistically different from fall 2014. The prototyping group had statistical difference in spring 2014 to fall 2014 and fall 2014 to spring 2015. Lastly, the design category did not show any statistical difference between the various semesters.

Figure 7. Investigation of Innovation Scores-Longitudinally: Kruskal Wallis analysis of the five semester groups showed statistical (p<0.05, n=20, Error shown is Standard Error) paired significance indicated by “x” and “+”.

Figure 8. Investigation of Prototyping Scores-Longitudinally: Kruskal Wallis analysis of the five semester groups showed statistical (p<0.05, n=15, Error shown is Standard Error) paired significance indicated by “x” and “+”.
The results from Figures 1-5 show a semester distribution assessing student scores in the three categories of innovation, prototyping, and design. Spring 2014 semester showed that teaching assistants ranked students higher in all three categories when compared to student actual performance and external validators’ scores. While, the highest marginal difference between external experts and the design challenge was 10% compared to 57%. While spring 2014 data showed high variance between the external experts and teaching assistants the following semesters (Fall semester 2014, fall Wednesday/Friday semester 2015)
experienced very similar results. The external validators and teaching assistants had at the most a 5% difference in scoring students on the three categories. Additionally, the design challenge was approximately 20% lower for all three metrics than the prototype validator scores. It should be pointed out that spring semester 2015 did not show any statistical significance. According to the data the external validators and teaching assistants were able to assess the students’ skills similarly however, there was a 20% difference of validator score to the actual performance, as noted by the DC, scores.

The overall average across five semesters did show an interesting trend. Student performance on the design challenge was aligned more with external experts’ grades than scores of the teaching assistants. Considering all students’ anecdotal reviews showed that students felt uneasy about being graded by external validators and attempted to perform much better during their presentation of their final prototype. Additionally as gathered from anecdotal reviews, students felt that teaching assistants were more of ‘friends’ then mentors and guides. Students teaching students are one of the most widely employed educational models, and one of the most helpful methods in learning new material\textsuperscript{1,15,16}. Utilizing teaching assistants in delivery of course material can help students better understand difficult concepts. Peer to peer teaching is viewed as a positive impact in educational understanding\textsuperscript{15}. However, the constant interaction in a casual manner may invalidate the scores from teaching assistants\textsuperscript{17}. This further emphasizes the need for external validators not only for their expertise, but unbiased opinion.

One of the key aspects of incorporating external experts and teaching assistants is to ensure that there would be no statistical difference in validator scores to actual performance scores. However, the results here showed that on average the external experts and teaching assistants showed very little differences in scoring the gamma prototype in any of the three categories: innovation, prototyping, and design. The main issue arose from design challenge scores being far lower than both external experts and teaching assistants. However, in considering the average across semester the external experts were closer to the average design challenge score. Additionally, There were only statistical differences in external expert scores across the first 3 semesters in innovation and prototyping. In the category of innovation spring 2014 to fall 2015 Wednesday scores showed statistical significance even with marginally small score differences. Prototyping showed the marginal score differences with statistical significance in spring 2014 to spring 2015, while the design category had not statistical differences.

The results show that there would need to be some form of standardization for the external validators or post score modification to account for the various differences. However, we can also see that the spring 2014 section was different in all three categories, as this was the first time the class was offered. When considering the courses excluding the spring 2014 sessions it appears there is some standardization among the scores. Additional analysis would be needed in
order to determine the most appropriate methodology to standardize and gage assessments by validators.

Addressing the need of accurate assessment for a project-based course by using external experts and teaching assistants is an effective strategy to ensure that students are retaining the information in this class for future courses\textsuperscript{13,18,12,11}. Additionally, the use of external validators to validate student performance is proving to be an effective strategy in assessing student performance in innovation, prototyping, and design. This is a key assessment for the future development of this course and for future development of engineering courses using project-based learning.

The results show key findings for a longitudinal semester assessment of project validation in a project-based course. In ensuring proper assessment of student achievement in vital engineering concepts, external experts and teaching assistants could be valuable resources to address the need. However, it should be noted that based off the results external experts could more accurately assess key concepts then teaching assistants. For this biomedical engineering course it is important to ensure innovation, prototyping, and design is learned so students can begin to use those skills in their senior design, workplace, or graduate school\textsuperscript{8,9}. Specifically, the work here has shown the comparison of a design challenge, an assessment of student performance, to scores presented by external experts. The overall findings show that external experts have the ability of accurately assessing student performance in a project-based prototyping. This rare need, in a junior biomedical engineering project-based course, has shown to be an effective strategy in assessing and ensuring that students have effectively learned innovation, prototyping, and design and are prepared to utilize these skills in future design tasks during their academic career.
Figure 10 Comparison of a Board Game and Medical Device: This figure shows the comparison of similar specifications that medical devices and board games have in common. The left column shows the board game process and the right column the medical device process. A diagnostic meter is on the right and a zombie board game on the left.
References


