Assessment of Project-Based Learning Courses Using Crowd Signals

Mr. Georgios Georgalis, Purdue University-Main Campus, West Lafayette (College of Engineering)

Georgios is currently a Ph.D. candidate in the School of Aeronautics and Astronautics at Purdue University and has completed his undergraduate degree at the National Technical University of Athens (NTUA). His research concentrates on a new approach to project risk assessment that is human-centric and allows for prediction of upcoming failures, which gives practitioners the opportunity to prevent them.

Dr. Karen Marais, Purdue University-Main Campus, West Lafayette (College of Engineering)

Dr. Karen Marais' educational research focuses on improving systems engineering education. She is the author of several technical publications, including 17 journal papers and two book chapters. She received an NSF CAREER award in 2014. Dr. Marais has worked in engineering for two decades, first in industry and then in academia. She holds a B. Eng. in Electrical and Electronic Engineering from the University of Stellenbosch, a B.Sc. in Mathematics from the University of South Africa, and an S.M and Ph.D. from the Department of Aeronautics and Astronautics at MIT.

Assessment of Project-Based Learning Courses using Crowd Signals

Project-based learning (PBL) is a growing component of engineering education in the United States. Its perceived educational value is exemplified by its explicit mention in ABET's Criterion 5, which requires engineering programs to provide a culminating design experience that incorporates engineering standards and multiple constraints. Capstone courses and design-build-test projects allow students to synthesize and apply engineering knowledge, skills, and tools to open-ended design problems. Students work and communicate in teams to complete tasks like generating requirements, and testing and integrating equipment. There appears to be widespread consensus that project-based learning is valuable, but, how well do these projects prepare students for engineering challenges in professional practice?

We consider one aspect of professional practice—failure. Despite many improvement efforts by organizations, systems engineering failures continue to occur. Previous research identified a set of common causes for these project failures. Does PBL provide students with opportunities to fail safely, and thereby learn to avoid failure in professional practice? We present here an approach to compare the rates of occurrence of failure causes in student team projects with industry projects. By comparing the occurrence rates, we achieve our first research goal to evaluate whether PBL offers sufficient opportunities of failure to students. Out of the ten failure causes we examined, we found that four are underrepresented in PBL, two of which are fundamentally related to systems engineering practice.

Failure causes may be hard to identify without the benefit of hindsight, so, we developed a set of crowd signals that may point to their presence. We observed 18 different student design projects across two semesters. Each week, the students answered a set of questions we developed to measure these crowd signals, while the instructors directly pointed out any instances of the failure causes they observed during the students' efforts. With the availability of such data we built logistic regression models to find correlations between specific crowd signals and the occurrence of a particular failure cause. By interpreting the regression coefficients, we achieve our second research goal to suggest specific improvements that instructors can use to give their students more failure opportunities during PBL.

1 Introduction

ABET's Criterion 5 requires engineering programs to provide all undergraduate students a *major design experience* that entails technical knowledge and skills acquired through the curriculum and incorporates realistic standards and constraints. The major design experience mentioned in the criterion is an example of project-based learning (PBL): the theory and practice of using real-world projects that have time restrictions to achieve specific objectives and to facilitate individual and collective learning [1]. PBL is a learner-centered approach that allows students to engage with an ill-defined project to promote research, teamwork, critical thinking, and synthesis of multidisciplinary technical knowledge [2], [3]. The instructor usually acts as a facilitator who guides the students through the learning experience as necessary, while allowing them to take responsibility for their project decisions [4]. PBL is widely considered to be successful, with students positively evaluating the approach and suggesting that it helps them develop their engineering intuition, making them responsible for their decisions, and becoming flexible thinkers

[5], [6]. Researchers and faculty also consider PBL an integral part of education that teaches students to handle complex problems that require diverse thinking and integration skills [7].

Despite the abundance of literature about PBL (e.g., see [8] for a review), some claim (e.g., see [9] , [10]) that research has not influenced PBL in practice. One reason for this disparity is that research studies focus on "packaged" projects (i.e., theoretical problems designed by education professionals to exactly meet PBL criteria), while faculty create a design project based on subject-matter expertise without exposure to most recent PBL research on how to maximize students' conceptual understanding of engineering design [11]. Others also question whether we have been evaluating the effectiveness of PBL correctly, emphasizing that instead of evaluating PBL based on student exam scores, we need strategies to identify whether particular aspects of PBL help educators satisfy specific learning outcomes [12].

In this paper, we evaluate the effectiveness of PBL by questioning how well it prepares students for a frequent engineering phenomenon in professional practice—failure. Does PBL provide students with enough opportunities to fail safely during their projects, and thereby learn to avoid failure in professional practice? We investigate how often different types of causes that lead to systems engineering failures happen. Previous research [13] identified such failure causes. These causes stem from the daily activities of the project team and may be hard to identify. We therefore developed a set of crowd signals that may point to the presence of these causes and use regression models to determine which crowd signals correlate with a particular cause. By interpreting our models, we suggest improvements to PBL that instructors can use to offer students more failure opportunities and better prepare them for industry.

The remainder of this paper is laid out as follows: Section 2 presents the types of information we collect about each student project, as well as the development of the crowd signals. Section 3 provides the research method on how we evaluate whether the occurrence rate of failure causes in student projects is similar to industry, and also how we use the crowd signals to propose improvements to PBL. Section 4 shows the results of the statistical tests to show which failure causes are underrepresented in PBL, and the interpretation of the regression models to identify which crowd signals correlate with increasing occurrence of a particular failure cause. Section 5 summarizes our conclusions.

2 Development of Crowd Signals

Despite advances in systems engineering approaches, project failures continue to occur with high frequency. According to the Project Management Institute (PMI), in 2018 only 53% of projects were on time, 58% within budget, and 68% met their goals [14]. In the aerospace sector in particular, defense programs often fail to meet performance criteria or expectations [15].

A posteriori knowledge of the failures themselves does not always help us prevent them, for example the knowledge that a project is behind schedule on its own does not give us enough information to act on, we have to look deeper and find the underlying reasons for the delays. Sometimes there is the expectation that such failures occur due to a single cause or in a chain-of-events manner. In complex systems however, we usually encounter multiple and interrelated causes, many of which would not cause a failure on their own [16], [17]. For example, the Boeing 787 project was three years behind schedule and over budget because of improper testing that led to battery overheating, increasing costs due to ineffective out-sourcing, supply chain communication issues, and poor management decisions [18]. Previous research has extensively studied and analyzed failures like the aforementioned and identified a set of 21 causes of systems engineering failures [13]. Of the 21 causes, we consider in this work the 10 that apply to student projects, as shown in Table 1.

Table 1: Common causes of systems engineering failures. Adapted from [13].

Systems engineering failure causes	
Failed to consider design aspect	Actor(s) in the organization failed to consider an aspect in the system design. In many cases, this causal action describes a design flaw, such as a single-point failure or component compatibility.
Used inadequate justification	Actor(s) in the organization used inadequate justification for a decision.
Failed to form a contingency plan	Actor(s) in the organization failed to form a contingency plan to implement if an unplanned event occurred.
Lacked experience	Actor(s)' lack of experience or knowledge led to the failure.
Kept poor records	Actor(s) in the organization did not review documentation or other work sufficiently to capture errors and deficiencies.
Inadequately communicated	Actor(s) in the organization failed to communicate with each other such that personnel were confused with the information they were given, had to "fill in the gaps" in the information they were given, or not notified about important information at all.
Subjected to inadequate testing	One or more actors in the organization subjected a component or subsystem to inadequate testing. This causal action captures inadequate tests as well as adequate tests performed inadequately.
Managed risk poorly	Actor(s) in the organization failed to identify, assess, formulate, or implement a proper mitigation measure.
Violated procedures	Actor(s) in the organization violated a procedure pertaining to the system, such as a maintenance or operation procedure.
Did not allow system aspect to stabilize	Actor(s) in the organization did not allow a system aspect like personnel, design, or requirements to stabilize before moving forward with an action.

Identifying the causes of a failure is also often not enough to completely understand why failures occur. People do not willfully omit important design aspects or conduct inadequate testing, but can be careless or lose focus while performing a critical design or testing task. We know from literature that certain activities, behaviors, and even personality characteristics can lead to poor habits and performance. For example, coordination failures can lead to ineffective group interactions and poor group performance [19].

For understanding all aspects of a systems engineering failure, we should capture information on both levels: the failure causes and the human actions and behavior that lead to these causes. To collect both levels of information in the student projects, we collect data both from the instructor and the students, once per week. The instructor provides the failure-centric information: whether they witnessed any of the ten listed causes that lead to failure (Table 1). The instructor answers a total of ten questions (one for each failure cause) with yes/no answers. Table 2 shows a summary of the questions to the instructor.

Table 2: We ask the instructor ten questions to record occurrences of systems engineering failure causes in student projects.

	Failure causes
1	Indicate whether a team "Failed to consider an aspect in the system design" this past week.
2	Indicate whether a team "Made a decision or action that was not well justified" this past week.
3	Indicate whether a team "Did not consider redundant components or measures for their actions" this past week.
4	Indicate whether a team "Made a mistake because members lack experience" this past week.
5	Indicate whether a team "Did not properly document their progress" this past week.
6	Indicate whether a team "Run into communication issues" this past week.
7	Indicate whether a team "Did not run adequate tests for their equipment" this past week.
8	Indicate whether a team "Managed risk poorly" this past week.
9	Indicate whether a team "Violated rules or procedures" this past week.
10	Indicate whether a team "Rushed into action without fully understanding the impacts to the system" this past week.

The students provide the human-centric information: the individual actions that lead to the failure causes and eventually to failures. To capture the human-centric information, we developed a set of crowd signals. Each week, the students that participate in the project teams answer a set of questions to measure these crowd signals. We created the crowd signals using three different approaches: using factors that previous research found are related to project failures and team performance, using systems archetypes to capture dysfunctional team practices, and using cognitive biases to identify potentially irrational and destructive individual actions. To arrive at a successful set of crowd signals, first we started by surveying literature that includes factors that affect team or project performance. Each factor then led to one or more questions that we asked the students. We note that we used the literature as a guide to define an initial set of questions and our questions are just one way of identifying the presence of a corresponding factor. For an initial set of factors, we included a wide range of literature from the following research areas in the search: human factors, systems engineering, project management, engineering education literature, psychology, and social sciences. Then, we used the definitions of ten common cognitive biases to capture individual actions that may contribute to failure and four systems archetypes that correlate with poor safety practices.

In summary, we developed questions in four categories:

1. Cognitive biases (*Individual Actions & Decisions*)
2. Organizational safety archetypes (*Team Actions & Archetypes*)
3. Factors that are related to team performance and/or project success from human factors, engineering education, and systems engineering literature (*Performance*)
4. Critical success factors from project management literature (*CSF*)

To develop each crowd signal, we pose a question that is guided by the definitions from literature but applies specifically to the specialized context of student projects. When possible, we phrase the questions so they are hard to game, meaning they do not have obvious "correct" answers. For example, *proactivity* is a factor that is associated with project performance as proactive people are willing to take action to affect their environment, in contrast to non-proactive individuals who are less likely to take action. Rather than asking students directly whether they think they are proactive (where the answer would most likely be "yes"), we look for evidence of the student attempting to get involved with a task that was outside their immediate responsibility (Q13).

Following this process, we generated a set of 20 questions, as shown in Table 3.

Table 3: The twenty questions to the students. We generated each question based on the definitions of corresponding factors from literature.

		Individual Actions & Decisions [20], [21]	
1	Bandwagon effect	Tendency to do or believe what others do or believe. As more people come to believe in something, others do too, regardless of the underlying evidence.	During the past week, did you have any arguments with your team about the next project actions/tasks?
2	Not invented here	Aversion to contact with or use of products, research, standards, or knowledge developed outside a group.	During the past week, did you get any new ideas about your project from other teams or people?
3	Confirmation bias	The tendency to search for, interpret, focus on and remember information in a way that confirms one's preconceptions.	During the past week, did you learn any new things that surprised you, because of your involvement with this project?
4	Normalcy bias	The refusal to plan for, or react to, a disaster which has never happened before.	During the past week, did you spend any time thinking about how things might go wrong for this project?
5	Ambiguity effect	The tendency to avoid options for which missing information makes the probability seem "unknown".	During the past week, did you disagree with an idea or decision because you thought you did not understand all potential implications?
6	Focusing effect	The tendency to place too much importance on one aspect of an event.	During the past week, can you single out one project decision by your team as the most important?
7	Parkinson's Law of Triviality	The tendency to give disproportionate weight to trivial issues.	During the past week, did your team spend significant time discussing what you thought as trivial matters about the project?
8	Anchoring	The tendency to rely too heavily, or "anchor", on one trait or piece of information when making decisions (usually the first piece of information acquired on that subject).	For any new project decisions that you had to make this week, did you think through all viable solutions or go with the one that you thought of first?
		Team Actions & Archetypes [22]	
9	Unintended side effects of fixes	Poorly thought out fixes may have unintended side effects.	If new problems occurred this week, do you think they were handled appropriately?

10	Stagnant risk management	When technological advances are not accompanied by concomitant understanding of the associated risks, risk may increase.	During the past week, did your team consider new potential risks as a result of any new project tasks or updates?
11	Fixing symptoms rather than root causes	Fixes to problems that only address the symptoms may worsen or prolong the original problem.	During the past week, were you disappointed because a problem that your team thought had been fixed, had instead continued or gotten worse?
12	The vicious cycle of bureaucracy	When organizations respond to problems with more rules and bureaucracy, employees may become apathetic or alienated.	During the past week, were you frustrated about any rule or bureaucracy that was out of your control?
		Performance	
13	Proactivity	Proactive individuals show initiative, are willing to take action and affect their environment, and show perseverance [23].	During the past week, how many times did you attempt to get involved with a project-related task that was outside your immediate responsibility?
14	Coordination	The unification, integration, synchronization of the efforts of group members to provide unity of action in the pursuit of common goals [24].	During the past week, how often did you interact with your team members while completing separate project tasks?
15			During the past week, for roughly what percentage of your team do you know exactly what they worked on?
16	Standardized work	Standardized work practices detail how work should be performed [25]	During the past week, rate the level of freedom you felt you had on how to complete your project tasks.
		CSF (Critical Success Factors) [26], [27]	
17	Modularization	Modular design, or "modularity in design", is a design approach that subdivides a system into smaller parts called modules or skids, that can be independently created and then used in different systems.	During the past week, roughly what percentage of the tasks you performed could be done independently of the rest of the project?
18	Clear objectives	To have effective tasks, then it is important to plan and pen clearly defined objectives that can deliver desired results.	During the past week, how clearly defined were your team's objectives?
19	Commitment	The state of being dedicated to a cause.	If your team announced to you today that they all quit, would you be willing to continue working on the project with a completely new team?
20	Availability of resources	Availability means capable of being used or the extent to which resources are available to meet the project's needs.	During the past week, rate your team's availability of resources (tools/space/software/funds) for you to use.

Each of the questions provides different types of data. Some questions have binary answers (e.g., Q1: whether an individual had arguments or not), some are expressed on Likert-scale (e.g., Q20: ranging availability of resources from very low availability to very high availability), and some are continuous numbers (e.g., Q17: the percentage of tasks that can be performed independently).

3 Research Approach

For the first part of our research, we identify which of the ten failure causes are underrepresented in student projects by comparing the occurrence rate of the failure causes between two samples: the student (“PBL”) and industry (“IND”) projects. The null hypothesis is that the occurrence rate of the failure causes in student projects is greater than or equal to the equivalent rate for the industry projects. We make two assumptions in forming the null hypothesis. First, we assume that student PBL-inspired projects are successful from an educational perspective (i.e., offer sufficient opportunities for the students to experience a particular failure cause before they graduate). Second, we assume that failure causes are more likely to occur in amateur teams, such as student teams, compared to professional teams. To test our hypothesis, we use Barnard’s exact statistical test [28], that can handle sample proportions.

The “PBL” sample included the occurrences of failure causes from the 18 student projects we observed over two semesters. The student teams were enrolled in two different courses at Purdue University. The student teams were formed in senior-level design courses and allowed students to participate in small-scale engineering projects. Each course had one instructor, who supervised the projects in their course and guided the students. We used an online survey built in Qualtrics to collect the answers from the students and instructor at the end of each week. Student identities were anonymized by having them provide their answers with a username of their choice. To motivate students to participate, we entered participants into a weekly drawing for a $20 Amazon gift certificate. During the summer 2018 semester we observed 6 projects, 4 of which continued in the fall semester and 2 of which ended. During the fall 2018 semester we observed 8 new projects. For the purposes of this study we consider the 4 projects that continued info the fall semester separately from their summer versions, because the student teams changed completely and the instructor also evaluated the projects on a per-semester basis.

The “IND” sample included failure cause data from 32 industry project failures from previous work [13] that identified and categorized occurrences of failure causes on various non-accident project failures.

To begin our analysis, we create an estimated *occurrence* measure for both samples. The measure describes the proportion of student and industry projects that included a particular failure cause *i*. We compute the *occurrence* measure separately for the two samples. For the industry sample, the estimated occurrence is:

$$\hat{O}_{(IND)i} = \frac{\sum_{j=1}^{n_1} TRUE_{(IND)i,j}}{n_1} \quad (1)$$

Where *i* is one of the ten failure causes, n_1 is the number of industry projects equal to 32, and $TRUE_{(IND)i,j}$ is a binary variable that is equal to 1 if failure cause *i* occurred during industry project *j* or 0 if not.

For the “PBL” sample, since we collected data for each project over time, we first compute the average detection rate $\widetilde{D}_{i,j}$ of a cause *i*, during project *j* over the number of weeks T_j we observed it for:

$$\widetilde{D}_{i,j} = \frac{\sum_{k=1}^{T} TRUE_i}{T_j} \tag{2}$$

If the detection rate $\widetilde{D}_{i,j}$ exceeds a set threshold, we assume that cause *i* has occurred during project *j*. We set the threshold somewhat arbitrarily at 20%, meaning that we assume the failure cause did occur if the instructor noted it in at least 1 out of 6 of the summer project weeks, or at least 2 out of 8 of the fall project weeks.

$$TRUE_{(PBL)i,j} = \begin{cases} 1 & \widetilde{D}_{i,j} \geq 0.2 \\ 0 & else \end{cases} \tag{3}$$

Then, the estimated *occurrence* for PBL is:

$$\hat{O}_{(PBL)i} = \frac{\sum_{j=1}^{n_2} TRUE_{(PBL)i,j}}{n_2} \tag{4}$$

Where *i* is one of the ten failure causes, n_2 is the number of student projects equal to 18, and $TRUE_{(PBL)i,j}$ is a binary variable that is equal to 1 if failure cause *i* occurred during student project *j* or 0 if not.

With the estimated *occurrence* measures defined for the two binomial samples "PBL" and "industry", we proceed with Barnard's exact statistical test with the null hypothesis that the actual failure cause occurrence rate $O_{PBL,i}$ in student projects is equal to or greater than the actual occurrence rate $O_{IND,i}$ in industry projects. We select a significance level α equal to 0.05.

$$H_0: O_{PBL,i} \geq O_{IND,i}$$
$$H_a: O_{PBL,i} < O_{IND,i} \tag{5}$$

Barnard's statistical test is a proportion test that has more power than other exact tests and is more accurate for small sample sizes like in our case, than a chi-squared test [29, 30]. The test assumes that the two samples "PBL" and "IND" are binomial experiments, which means that each student project and industry project are independently equally likely to show signs of a failure cause. To conduct the test, we first create a 2x2 contingency table for each failure cause *i*:

Table 4: Generic contingency table for each of the ten failure causes *i*.

Failure cause *i*	**Occurrence**	**Not occurrence**	**Total**
"PBL" sample	$x_{1,i}$	$18 - x_{1,i}$	18
"IND" sample	$x_{2,i}$	$32 - x_{2,i}$	32
Total	$x_{1,i} + x_{2,i}$	$50 - x_{1,i} - x_{2,i}$	50

Based on the previous definitions, the estimated *occurrence* measure for both samples is:

$$\hat{O}_{(PBL)i} = \frac{x_{1,i}}{18} \tag{6}$$

$$\hat{O}_{(IND)i} = \frac{x_{2,i}}{32} \tag{7}$$

Under the null hypothesis, we posit that the common probability responding to the two groups is p. Then, the probability of obtaining the contingency table, M_0, is the product of two binomials:

$$P(M_0|p) = \binom{18}{x_{1,i}}\binom{32}{x_{2,i}} p^{x_{1,i}+x_{2,i}}(1-p)^{50-x_{1,i}-x_{2,i}} \tag{8}$$

To obtain the p-value for the test, we consider the critical region that contains all contingency tables that represent an outcome at least as extreme the observed table. Then, the significance level $\alpha(p)$ can be obtained by summing the above probabilities over all of the tables in the critical region (CR). The p-value is obtained by maximizing the significance level function $\alpha(p)$ [31]:

$$a(p) = \sum_{CR} P(M_0|p) \tag{9}$$

$$a^* = \max_{0<p<1}(a(p)) \tag{10}$$

After the completion of the statistical test for all ten failure causes, we identify in Section 4 which ones are underrepresented in PBL compared to industry projects (i.e., the ones that the null hypothesis is rejected with a p-value less than the 0.05).

We then find which student actions are more likely to increase the occurrence of the underrepresented failure causes from the previous step. The crowd signals assist in this process. We built mixed effects logistic regression models that correlate the crowd signals with the occurrence of a failure cause. The predictors X are the student responses to the crowd signals. The predicted variable Y is binary, such that when the particular failure cause i occurs, $Y_i = 1$, and when it does not, $Y_i = 0$. We used mixed effects regression here because we collect data over time and we need to account for subject non-independence, as suggested by [32]. We built a total of N models, one for each underrepresented failure cause. We computed the regression models as follows:

$$\hat{Y}_{i,t} = a + bX_{i,t}{}^T + \delta_t t + c_i + \varepsilon_{it} \tag{11}$$

where a is the intercept constant, b is a column vector of slopes for each predictor, $X_{i,t}{}^T$ is a row vector of the predictors for week t, $c_i \sim N(0, \sigma_i{}^2)$ are the i student-specific random effects, δ_t is the coefficient for week-specific fixed effects, and $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon{}^2)$ is the observation-specific random error.

By interpreting the regression coefficients, we determined which of the 20 crowd signals (shown before in Table 3) are strongly and positively correlated with an increase in the occurrence of each of the N underrepresented failure causes. With this information, we provide targeted suggestions to instructors that may increase the occurrences of underrepresented failure causes and therefore provide the students with more, and more appropriate, failure experiences before they join the workforce.

4 Results and Discussion

To complete the first research goal to identify which of the ten failure causes are underrepresented in PBL, we populate the ten 2x2 contingency tables and compute the estimated *occurrence* measures for each of the failure causes as described in Section 3, and proceed with the statistical test. Table 5 shows the results of the statistical test.

Table 5: Occurrence measures and Barnard's statistical test results for the failure causes. 4 out of 10 failure causes are underrepresented in the 18 student projects we studied compared to the 32 industry projects from [13].

Failure cause	$x_{1,i}$	$x_{2,i}$	$\widehat{O}_{(PBL)i}$	$\widehat{O}_{(IND)i}$	Barnard's test p-value	H0 rejected?
Failed to consider design aspect	**8**	**29**	**44.4%**	**90.6%**	**0.00035**	**Yes**
Used inadequate justification	6	11	33.3%	34.4%	0.49566	No
Failed to form a contingency plan	3	8	16.6%	25.0%	0.35212	No
Lacked experience	9	15	50.0%	46.9%	0.44882	No
Kept poor records	4	4	22.2%	12.5%	0.27398	No
Inadequately communicated	4	11	22.2%	34.3%	0.25507	No
Subjected to inadequate testing	**3**	**15**	**16.6%**	**46.9%**	**0.02299**	**Yes**
Managed risk poorly	**1**	**12**	**5.5%**	**37.5%**	**0.00965**	**Yes**
Violated procedures	1	5	5.5%	15.6%	0.21434	No
Did not allow system aspect to stabilize	**3**	**16**	**16.6%**	**50.0%**	**0.01391**	**Yes**

Any statistical test includes limitations and does not necessarily reflect truth. Here, the quality of the student project data depends on the ability of the instructor to identify the failure causes and report them accordingly.

Based on the statistical test, we identify 4 out of 10 failure causes that statistically appear less frequently in the student projects compared to industry. "Failed to consider a design aspect" and "Did not allow system aspect to stabilize" appear underrepresented in PBL, and are two failure causes that are fundamentally linked to proper systems engineering. "Subjected to inadequate testing" is less frequent in PBL perhaps because instructors tend to monitor equipment testing more than other team activities because mistakes can be expensive (CNC machine time, raw materials, etc.). For the student data, "managing risk poorly" and "violating procedures" have very low occurrence measures. The instructors observed these two failure causes, but not frequently enough in a project to count as an occurrence according to our threshold as discussed in the previous section. Figure 1 shows the four failure causes where there is sufficient evidence to reject the null hypothesis (i.e., their occurrence measure values are small enough to conclude that they are underrepresented in PBL compared to industry projects).

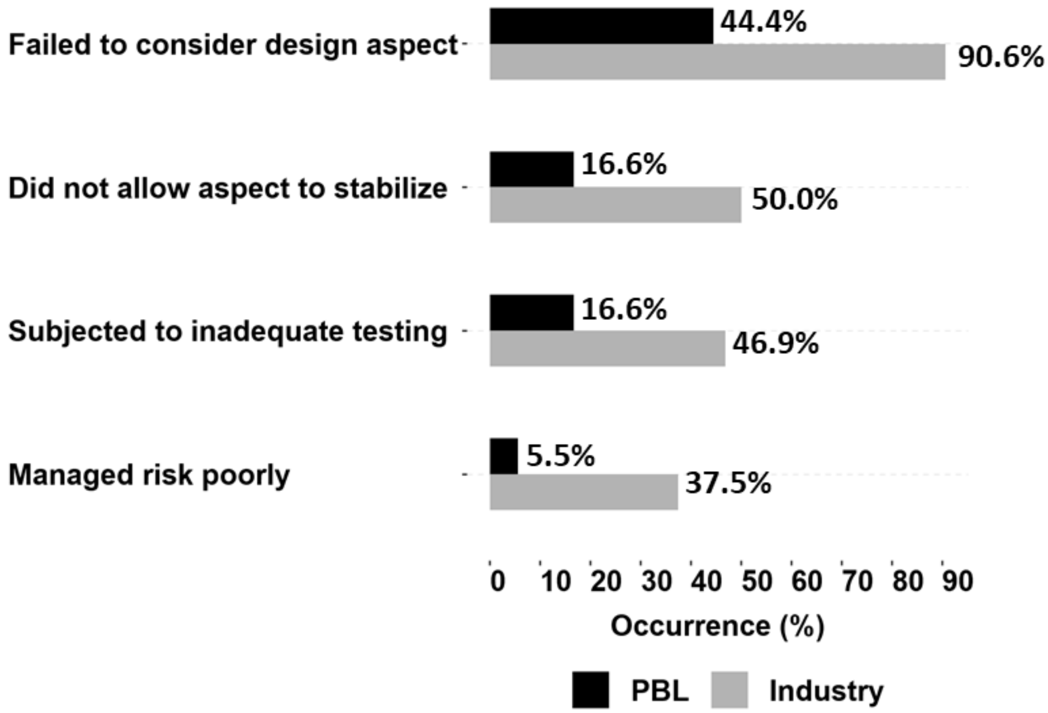

Figure 1: The four failure causes that statistically occur less frequently in PBL compared to industry. "Failing to consider a design aspect" and "not allowing a system aspect to stabilize" are fundamentally related to systems engineering. One reason for this result might be that the student projects are less complicated as systems, but as an educational community we should offer more opportunities for students to experience such failures.

Figure 2 shows the six failure causes for which the null hypothesis holds, that is, these causes statistically occur at least as frequently in student projects when compared to industry, which is encouraging for PBL. A perhaps unexpected finding is that student projects have lower occurrences of "failure to form contingency plans" and "inadequate communication" than professional teams. We justify this result on the notion that student teams are often not required to form contingency plans and are instead given a very specific task from the instructor. Because student teams tend to be smaller than professional teams, it is common for student projects to delegate tasks individually. In this case, communication may not be required as often and therefore is less likely to appear inadequate, compared to a large project with many systems engineers that communicate continuously.

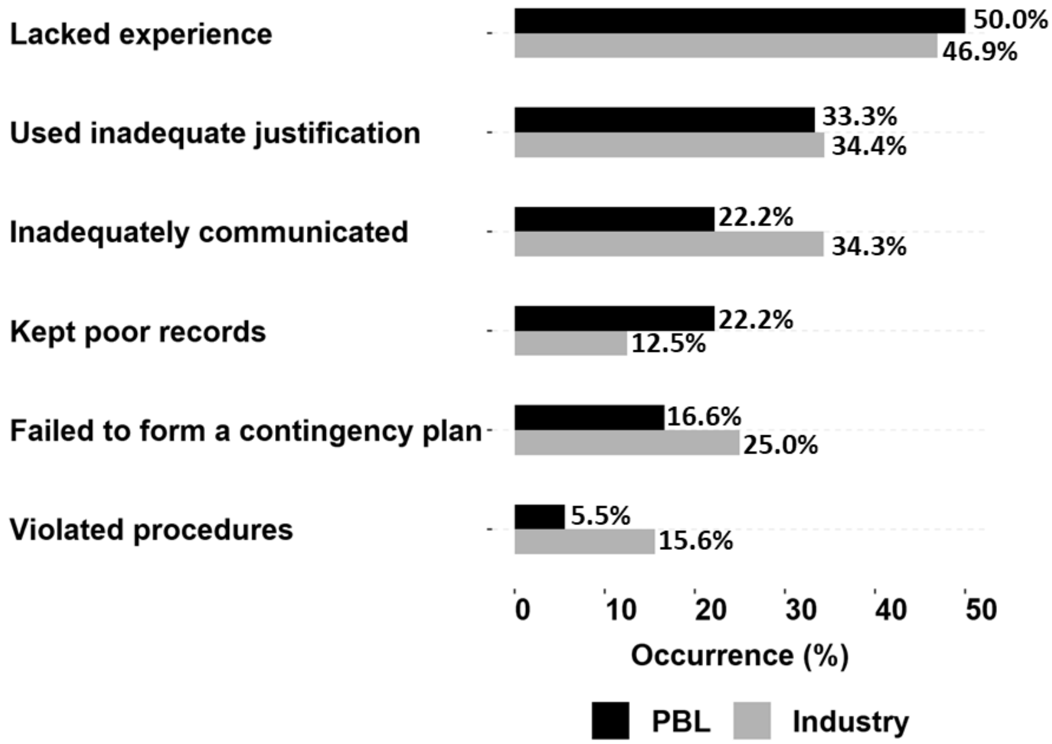

Figure 2: The six failure causes that statistically occur at least as frequently in PBL as in industry. "Violating procedures" has the lowest occurrence overall suggesting that perhaps it is harder to detect than other causes. On the contrary, "lacking experience" occurs more often and is likely easier to detect.

After we identified the $N=4$ underrepresented failure causes in PBL, we used the crowd signals to build mixed effects logistic regression models that relate the binary occurrence of each of these failure causes (Y) with the crowd signals collected form the students (X). We built one model for each of these 4 failure causes. The responses to questions 1 to 12 were binary (Yes/No), and the responses to questions 13 to 20 were either numerical or Likert-scale (and coded as numerical from 1-5). Defining an observation as a set of answers to the crowd signals from a student team member, we built the regression models with 198 observations.

Table 6 shows the resulting model for the failure cause "Failed to consider design aspect." The interpretation of the coefficients gives us the exact change in odds of the failure cause occurring over not occurring for each of the questions. For the scope of this work, we do not focus on these values themselves, but rather on the statistical significance and the sign of the coefficients. This first model shows with some statistical significance that when student teams are not having arguments (Q1) and are handling problems properly (Q9), the likelihood of failing to consider a design aspect of a system decreases. Also, the likelihood of the failure cause occurring increases with time, which is expected since a project becomes more complicated and more involved with progress. From this interpretation, if an instructor was to increase the occurrence of this failure cause, which lags significantly behind when compared to industry (Figure 1), they would have to find a way to increase (meaningful) arguments between the students. Perhaps encouraging students to take initiative in the project, share more ideas with each other, and have brainstorming sessions outside class time could help in this direction. Although handling problems improperly is hard and perhaps inappropriate to directly encourage, the instructor could give more responsibility to the students with tasks that are not very clear, and let them make mistakes and attempt to resolve them themselves.

Table 6: Logistic regression model for the failure cause *Failed to consider design aspect*.

Coefficient	Estimate (std. error)	Coefficient	Estimate (std. error)
a	0.717 (1.452)	b_{13}(Q13)	-0.016 (0.233)
b_1(Q1 = No)	**-0.975 (0.560)^**	b_{14}(Q14)	-0.140 (0.233)
b_2(Q2 = No)	-0.146 (0.536)	b_{15}(Q15)	-0.197 (0.217)
b_3(Q3 = No)	0.150 (0.473)	b_{16}(Q16)	-0.079 (0.185)
b_4(Q4 = No)	-0.257 (0.485)	b_{17}(Q17)	0.174 (0.226)
b_5(Q5 = Yes)	-0.422 (0.516)	b_{18}(Q18)	-0.041 (0.175)
b_6(Q6 = Yes)	-0.391 (0.549)	b_{19}(Q19)	-0.265 (0.201)
b_7(Q7 = Yes)	-0.072 (0.527)	b_{20}(Q20)	-0.084 (0.183)
b_8(Q8 = First thought)	0.450 (0.733)	**δ_t**	**0.303 (0.125)***
b_9(Q9 = No)	**1.523 (0.743)***	Random effects $c_i \sim N(0, e^{-14^2})$	
b_{10}(Q10 = No)	-0.741 (0.588)	^ $p < .1$ * $p < .05$	
b_{11}(Q11 = Yes)	-0.889 (0.587)	** $p < .01$	
b_{12}(Q12 = Yes)	0.602 (0.635)	*** $p < .001$	

Table 7 shows the resulting model for the failure cause "Subjected to inadequate testing." This second model shows with some statistical significance that when student teams discuss trivial matters (Q7), have increasing freedom (Q16) from the instructor, and have increasing number of tasks that can be completed independently (Q17), the likelihood of doing inadequate testing increases. The results from this model appear reasonable if we consider that discussing trivial matters perhaps indicates lack of attention to detail, which is important during testing. Having more than usual freedom from the instructor leaves more room for the student to err. When the number of tasks that require immediate attention increases for a project, there is often just one student responsible for the equipment testing, because student teams are generally small, and therefore the tasked student is in a more vulnerable position to make a mistake. We also find that the likelihood of this failure cause occurring increases with time, which may be because testing occurs at later stages in the project. The results from this second model suggest that instructors should give more freedom to students and allow them to responsibly complete a task on their own, which in turn offers them more opportunities to experience "safe" failure.

Table 7: Logistic regression model for the failure cause *Subjected to inadequate testing*.

Coefficient	Estimate (std. error)	Coefficient	Estimate (std. error)
$\boldsymbol{a}$	**-5.233 (2.393)***	b_{13}(Q13)	0.048 (0.372)
b_1(Q1 = No)	-1.08 (0.705)	b_{14}(Q14)	-0.002 (0.289)
b_2(Q2 = No)	0.194 (0.778)	b_{15}(Q15)	0.198 (0.314)
b_3(Q3 = No)	-0.339 (0.626)	$\boldsymbol{b_{16}}$**(Q16)**	**0.585 (0.265)***
b_4(Q4 = No)	0.638 (0.608)	$\boldsymbol{b_{17}}$**(Q17)**	**0.696 (0.348)***
b_5(Q5 = Yes)	-0.396 (0.647)	b_{18}(Q18)	-0.315 (0.238)
b_6(Q6 = Yes)	0.230 (0.708)	b_{19}(Q19)	0.028 (0.258)
$\boldsymbol{b_7}$**(Q7 = Yes)**	**1.202 (0.692)^**	b_{20}(Q20)	0.076 (0.245)
b_8(Q8 = First thought)	-0.004 (1.018)	$\boldsymbol{\delta_t}$	**0.440 (0.178)***
b_9(Q9 = No)	-0.560 (0.887)	Random effects $c_i \sim N(0, 1.027^2)$	
b_{10}(Q10 = No)	0.638 (0.723)	^ $p < .1$	
b_{11}(Q11 = Yes)	-1.074 (0.846)	$*$ $p < .05$ $**$ $p < .01$	
b_{12}(Q12 = Yes)	0.958 (0.853)	$***$ $p < .001$	

Table 8 shows the resulting model for the failure cause "Managed risk poorly." This third model shows with some statistical significance that when student teams do not communicate with other teams outside their own team (Q2) and when they have members argue about potential implications of decisions (Q5), they are more likely to manage risk poorly. Both situations involve some lack of perspective, either because they do not talk to other teams or because they do not understand all the potential implications. The model also shows that the occurrence of the failure cause decreases with increasing proactivity (Q13), clearly-defined objectives (Q18), and when students address symptoms rather than root causes of problems (Q11). Perhaps students who attempt to solve minor problems, even if they are not the perfect solution to the core problem, gain experience in dealing with problems, and are therefore less likely to manage risk poorly. Clearly-defined objectives from the instructor and proactivity are both positive aspects in terms of project success and, as expected, decrease the likelihood of managing risk poorly. Also, similarly to the previous two failures causes, managing risk poorly becomes more likely with project progress, perhaps because of increasing complexity and workload.

A recommendation to the instructor from this model is to once again promote arguments and discussion amongst the teams, possibly give objectives that are not as clear, and have teams work independently to give students more chances to experience the consequences of managing risk poorly.

Table 8: Logistic regression model for the failure cause *Managed risk poorly*.

Coefficient	Estimate (std. error)	Coefficient	Estimate (std. error)
a	-2.226 (1.418)	$\boldsymbol{b_{13}}$**(Q13)**	**-0.612 (0.359)^**
b_1(Q1 = No)	-0.026 (0.617)	b_{14}(Q14)	0.264 (0.227)
$\boldsymbol{b_2}$**(Q2 = No)**	**0.923 (0.551)^**	b_{15}(Q15)	0.022 (0.217)
b_3(Q3 = No)	0.264 (0.465)	b_{16}(Q16)	-0.180 (0.180)
b_4(Q4 = No)	0.095 (0.452)	b_{17}(Q17)	0.205 (0.215)
$\boldsymbol{b_5}$**(Q5 = Yes)**	**1.038 (0.528)***	$\boldsymbol{b_{18}}$**(Q18)**	**-0.296 (0.178)^**
b_6(Q6 = Yes)	0.138 (0.528)	b_{19}(Q19)	0.015 (0.188)
b_7(Q7 = Yes)	-0.963 (0.531)	b_{20}(Q20)	0.119 (0.171)
b_8(Q8 = First thought)	0.169 (0.729)	$\boldsymbol{\delta_t}$	**0.223 (0.124)^**
b_9(Q9 = No)	0.420 (0.696)	Random effects $c_i \sim N(0,\ e^{-14^2})$	
b_{10}(Q10 = No)	-0.919 (0.562)	^ $p < .1$ $*$ $p < .05$ $**$ $p < .01$ $***$ $p < .001$	
$\boldsymbol{b_{11}}$**(Q11 = Yes)**	**-1.192 (0.600)***		
b_{12}(Q12 = Yes)	-0.492 (0.700)		

Table 9 shows the resulting model for the failure cause "Did not allow system aspect to stabilize." This last model does not show significance in relation to any particular question or factor, and is therefore not as useful. We note that mixed effects logistic regression models use up a lot of degrees of freedom for the estimation of the parameters and due to the large number of features require strong evidence to show statistical significance.

Table 9: Logistic regression model for the failure cause *Did not allow system aspect to stabilize*.

Coefficient	Estimate (std. error)	Coefficient	Estimate (std. error)
a	-0.543 (1.187)	b_{13}(Q13)	-0.328 (0.274)
b_1(Q1 = No)	0.251 (0.506)	b_{14}(Q14)	0.041 (0.192)
b_2(Q2 = No)	0.244 (0.452)	b_{15}(Q15)	-0.165 (0.191)
b_3(Q3 = No)	-0.203 (0.409)	b_{16}(Q16)	-0.082 (0.163)
b_4(Q4 = No)	0.164 (0.398)	b_{17}(Q17)	0.008 (0.198)
b_5(Q5 = Yes)	-0.043 (0.447)	b_{18}(Q18)	-0.150 (0.152)
b_6(Q6 = Yes)	0.361 (0.448)	b_{19}(Q19)	-0.103 (0.159)
b_7(Q7 = Yes)	-0.150 (0.434)	b_{20}(Q20)	-0.034 (0.160)
b_8(Q8 = First thought)	-0.166 (0.640)	δ_t	-0.109 (0.111)
b_9(Q9 = No)	0.229 (0.580)	Random effects $c_i \sim N(0,\ e^{-14^2})$	
b_{10}(Q10 = No)	-0.399 (0.507)	^ $p < .1$ $*$ $p < .05$ $**$ $p < .01$ $***$ $p < .001$	
b_{11}(Q11 = Yes)	0.557 (0.502)		
b_{12}(Q12 = Yes)	-0.113 (0.545)		

5 Conclusions

We addressed two main research questions in this work. First, we evaluated whether a set of failure causes observed in industry projects appear in equivalent occurrence rates in project-based learning (PBL), which would indicate that PBL is helpful to prepare students with enough opportunities to experience failure before their graduation. We found that 4 out of the 10 failure causes we examined are not sufficiently represented in our student projects. Secondly, and after we identified the failure causes that are underrepresented in PBL, we built regression models using student crowd signals to arrive at useful recommendations for instructors. We developed the hard to game crowd signals starting from literature on factors that affect individual and team performance. By interpreting the model coefficients, we suggested recommendations to the instructors that may increase the occurrence of particular failure causes. Allowing students to take initiative, work with more freedom, share more ideas with each other, and assigning tasks with ill-defined objectives may help in this direction. The goal is for the recommendations to create an environment for the student teams that more closely resembles industry experience during failure. Some changes may appear as negatively impacting the performance of the teams by promoting more failure opportunities (e.g., ill-defined objectives). Despite the seemingly negative notion of failure, these changes may increase the effectiveness and educational value of PBL. Instructors can integrate such changes in a controlled manner to retain a fair grading scheme for their course (e.g., consistently applying changes across all projects or evaluating students based on effort). As an educational community, the goal of PBL is to prepare the students to the best of our ability, and the importance of preparing them for failure cannot be understated.

Acknowledgements

The material presented in this paper is based on work supported, in whole or in part, by the U.S. Department of Defense through the Systems Engineering Research Center (SERC) under Contract HQ00034-13-D-0004 RT #206. SERC is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology.

References

[1] R.J. DeFillippi, “Introduction: Project-based learning, reflective practices and learning.” *Management Learning*, Vol. 32, No. 1, pp. 5-10, 2001.

[2] J.E. Mills and D.F. Treagust, “Engineering education—Is problem-based or project-based learning the answer.” *Australasian Journal of Engineering Education*, Vol. 3, No. 2, pp.2-16, 2003.

[3] J.R. Savery, “Overview of problem-based learning: Definitions and distinctions.” *The Interdisciplinary Journal of Problem-based Learning*, Vol. 1, No. 1, pp. 9-20, 2006.

[4] C.J. Atman, R.S. Adams, M.E. Cardella, J. Turns, S. Mosborg, and J. Saleem, “Engineering design processes: A comparison of students and expert practitioners.” *Journal of Engineering Education*, Vol. 96, No. 44, pp. 359-379, 2007.

[5] W. Hall, S. Palmer, and M. Bennett, “A longitudinal evaluation of a project-based learning initiative in an engineering undergraduate programme.” *European Journal of Engineering Education*, Vol. 37, No. 2, pp. 155-165, 2012.

[6] M. Frank, I. Lavy, and D. Elata, “Implementing the project-based learning approach in an academic engineering course.” *International Journal of Technology and Design Education*, Vol. 13, No. 3, pp. 273-288, 2003.

[7] M. Lehmann, P. Christensen, X. Du, and M. Thrane, “Problem-oriented and project-based learning (POPBL) as an innovative learning strategy for sustainable development in engineering education.” *European Journal of Engineering Education*, Vol. 33, No. 3, pp. 283-295, 2008.

[8] D. Kokotsaki, V. Menzies, and A. Wiggins, “Project-based learning: A review of the literature.” *Improving Schools*, Vol. 19, No. 3, pp. 267-277, 2016.

[9] A.C. McCormick, J. Kinzie, and R.M. Gonyea, “Student engagement: Bridging research and practice to improve the quality of undergraduate education.” *In Higher education: Handbook of theory and research*, pp. 47-92, Springer, Dordrecht, 2013.

[10] S. Cottrell, “A matter of explanation: assessment, scholarship of teaching and their disconnect with theoretical development.” Medical Teacher, Vol. 28, No. 4, pp. 305-308, 2006.

[11] W.T. John, and W. Thomas, “A review of research on project-based learning.” The Autodesk Foundation, San Rafael, California, 2000.

[12] C.E. Hmelo-Silver, “Problem-based learning: What and how do students learn?” *Educational Psychology Review*, Vol. 16, No. 3, pp. 235-266, 2004.

[13] D. Sorenson, and K. Marais, “Patterns of causation in accidents and other systems engineering failures.”, In *Systems Conference (SysCon), 2016 Annual IEEE* (pp. 1-8), IEEE, 2016.

[14] Project Management Institute, “The Pulse of the Profession. 10th Global Management Survey”, Project Management Institute (PMI), 2018. [Accessed Nov. 13, 2018]

[15] R. Charette, "What's Wrong with Weapons Acquisitions?" IEEE Spectrum, 1 November 2008. [Accessed Jan. 15, 2019]

[16] V. Milch and K. Laumann, "Interorganizational complexity and organizational accident risk: A literature review." *Safety Science*, Vol. 82, pp. 9-17, 2016.

[17] M.E. Paté-Cornell, "Learning from the piper alpha accident: A postmortem analysis of technical and organizational factors", *Risk Analysis*, Vol. 13, No. 2, pp. 215-232, 1993.

[18] S. Denning, "What Went Wrong at Boeing?" *Forbes*, Jan. 21, 2013. [Online]. Available: https://www.forbes.com/sites/stevedenning/2013/01/21/what-went-wrong-at-boeing/#31cefa357b1b

[19] K. Srikanth, S. Harvey, and R. Peterson, "A dynamic perspective on diverse teams: Moving from the dual-process model to a dynamic coordination-based model of diverse team performance." *The Academy of Management Annals*, Vol. 10, No. 1, pp. 453-493, 2016.

[20] G. Montibeller and D. von Winterfeldt, "Cognitive and motivational biases in decision and risk analysis." *Risk Analysis*, Vol. 35, No. 7, pp. 1230-1251, 2015.

[21] P. Baybutt, "The validity of engineering judgment and expert opinion in hazard and risk analysis: The influence of cognitive biases", *Process Safety Progress*, Vol. 37, No. 2, pp. 205-210, 2018.

[22] K. Marais, J.H. Saleh, and N.G. Leveson, "Archetypes for organizational safety", *Safety science*, Vol. 44, No. 7, pp. 565-582, 2006.

[23] B.L. Kirkman and B. Rosen, "Beyond self-management: Antecedents and consequences of team empowerment", *Academy of Management journal*, Vol. 42, No. 1, pp. 58-74, 1999.

[24] E. Salas, N.J. Cooke, and M.A. Rosen, "On teams, teamwork, and team performance: Discoveries and developments", *Human Factors*, Vol. 50, No. 3, pp. 540-547, 2008.

[25] L.L. Gilson, J.E. Mathieu, C.E. Shalley, and T.M. Ruddy, "Creativity and standardization: complementary or conflicting drivers of team effectiveness?", *Academy of Management Journal*, Vol. 48, No. 3, pp. 521-531, 2005.

[26] J.K. Pinto and D.P. Slevin, "Critical factors in successful project implementation", *IEEE Transactions on Engineering Management*, Vol. 1, pp. 22-27, 1987.

[27] D.K.H. Chua, Y.C. Kog, and P.K. Loh, "Critical success factors for different project objectives." *Journal of Construction Engineering and Management*, Vol. 125, No. 3, pp. 142-150, 1999.

[28] G.A. Barnard, "A new test for 2×2 tables." *Nature*, Vol. 156, p.177, 1945.

[29] J. Röhmel and U. Mansmann, "Unconditional non-asymptotic one-sided tests for independent binomial proportions when the interest lies in showing non-inferiority and/or superiority." *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, Vol. 41, No. 2, pp. 149-170, 1999.

[30] S. Suissa and J.J. Shuster. "Exact unconditional sample sizes for the 2×2 binomial trial." *Journal of the Royal Statistical Society. Series A (General)*, pp. 317-327, 1985.

[31] A.S. Mato and A.M. Andrés, "Simplifying the calculation of the P-value for Barnard's test and its derivatives." *Statistics and Computing*, Vol. 7, No. 2, pp. 137-143, 1997.

[32] X.A. Harrison, L. Donaldson, M.E. Correa-Cano, J. Evans, D.N. Fisher, C.E. Goodwin, B.S. Robinson, D.J. Hodgson, and R. Inger, "A brief introduction to mixed effects modelling and multi-model inference in ecology", *PeerJ*, Vol. 6, p. e4794, [Online], Available: https://peerj.com/articles/4794, 2018.