A Pass/Fail Option for First-Semester Engineering Students: A Critical Evaluation

Ken Stanton, Tom Siller
Colorado State University

Abstract – **The first semester of a collegiate engineering program can be quite stressful for a student, resulting in anxiety and decreased academic performance. While there are various factors leading to these outcomes, the pressure of grades early in an engineering program can be a significant contributor. GPA is typically used to determine a student's academic success, but the reality is that the student often needs time to learn how to earn and keep good grades in college before they accurately reflect the student's performance and potential. Considering this, Colorado State University recently offered an option for incoming engineering students to take all of their first-semester classes as pass/fail. For the fall semester of 2010, 138 of 448 first-year students enrolled in the pass/fail option. In this first semester of the program, success was determined from a few measures by comparing performance of students in the pass/fail option to those in traditional A-F grading. In the first analysis described herein, GPA was reviewed for both groups of students, which revealed that students who selected the option received lower grades than those who did not. Authors review the program critically, often holding opposing viewpoints on whether outcomes have been and will be positive.**

Index Terms – pass/fail, grades, student stress.

INTRODUCTION

This paper explores an experimental program implemented in a first-year engineering program from a unique angle. The authors of this work disagreed on the expected outcomes of implementing this experimental program, as well as how to assess the outcomes, and continued to hold these differing perspectives while writing this paper. Summarily, the difference between these views stems from the authors' backgrounds and current perspectives. The first author recently received his Ph.D. in Engineering Education from Virginia Tech, and hence is viewing the pass-fail program from the angle of student learning and motivation. The second author is currently the Associate Dean for Academic and Student Affairs, and therefore is most highly concerned with student well-being. As such, readers will hear differing and sometimes dissenting opinions in the background and analysis of this program, and hence can consider for themselves the information from different angles. Authors are referred to herein as "first" and "second" author for consistency.

During the fall 2010 semester, Colorado State University implemented an optional experimental program for all first-year students in its College of Engineering. A new pass/fail (P/F) project was designed that allows students to take their entire first semester academic load as P/F. This was motivated by a growing concern that first-year engineering students were suffering emotionally due to the rigor of the engineering curriculum. The premise was that relieving some of the pressure to perform, in terms of GPA, while also providing more contact and support with engineering faculty/staff, could enhance student performance and relieve some related stress.

The goals of this paper were to:

- Present a preliminary analysis of the pass/fail program
- Present diverse perspectives on expectations for and results from the pass/fail program

BACKGROUND

I. Student Stress

First-year college students are expected to make a very challenging transition from high school with relatively minimal help, as they progress to the early stages of adulthood. The typical, traditional, first-year student is about 18 years old, is living away from home for the first time, has to manage their own diet for the first time, has to build a new social group, has to navigate a new geographical location, and has to find classes around their new campus. On top of all this, college courses are typically faster paced, more rigorous, and more competitive than classes students experienced in high school. The challenge of collegiate courses is notably greater for students entering the field of engineering, as well.

Considering all of this collectively, the typical first-year engineering student experiences a great deal of stress. While this stress is important to motivate personal growth and learning in the student, administrators are often concerned with how some students handle the stress. A percentage of students drop out of classes, withdraw from college, resort to alcohol and drug use, get in trouble with the law, and, in the worst case, commit suicide in response to stresses, and administrators don't want the school to be the cause of any of these.

Therefore, schools have sought ways to lessen the initial stress of the college transition with orientation programs, mentorships, residence life programs, and academic assistance. Still, first-semester grades are often low for engineering students, which can negatively affect future academics, scholarships, and the ability to finish their program on time, and hence, undesired stress levels may remain.

In response, administrators at some engineering schools, such as MIT and CalTech [1], have taken steps that attempt to mitigate the impact of first-semester grades on students' stress levels and future academics. The same concern has been expressed for first-semester engineering students at Colorado State University. In response, the second author of this paper suggested that the university try out a P/F program option for first-semester engineering students, to see if impacts were beneficial.

II. The Pass-Fail Program

Colorado State University is a public university with approximately 1,800 undergraduate engineering students and 500 graduate engineering students. High school students may apply directly into any of the majors in the college, or enter as undecided engineering students. A typical first-semester course load includes a department-specific introduction to engineering course or a college-wide engineering course based on the NAE's Grand Challenges, calculus, and a science course – often physics or chemistry.

The decision was made during the summer of 2010 to implement the new P/F program for the Fall 2010 semester, under the supervision of the second author. Administrative-level support for this effort was attained based on the knowledge that similar systems are in place at other successful engineering schools, such as MIT and CalTech. At CalTech, a no-grade policy was first implemented in 1964, and was quite successful [1]. More recently, P/F systems have been successfully implemented in medical schools [2] with the intent of reducing student stress. With this as a background, the Colorado State University (CSU) College of Engineering was given permission to move forward with an experiment during the Fall 2010 semester, despite the fact that this would violate several established university policies. As this was considered an experiment, exceptions to the pertinent policies were given by upper administration with the proviso that a comprehensive assessment plan would be implemented and results shared with the administration. During the first week of classes, first-year engineering students were recruited into the program on a purely volunteer basis; 138 out of 448 students chose this option. The program included the following elements:

- A new, required, 1-credit seminar course was implemented to increase student-faculty contact. A senior faculty member coordinated multiple sections of this seminar and developed curricula that focused on connecting engineering concepts with math and sciences courses being taken by the students.
- Passing was defined by the university as S for a grade of D and above, and failing as a U for a grade of F.
- Students could choose to either take all courses S/U or all courses traditional grading. This decision had to be made by the university-established deadline for choosing the S/U option.
- Faculty members were unaware of students' choice to participate. Faculty graded using typical A-F grading, submitted the grades, and the Registrar's Office converted the grades to S/U for students in the project.
- Students were allowed to recover grades with certain restrictions, including: a student must recover all grades and the decision was then irreversible.
- The program was only available for first-time, entering freshman in their first semester.
- Students who chose to participate in this project were allowed to maintain all University and College of Engineering sponsored scholarships and Honors standing.

III. Pass-Fail Grading Options

The first author of this paper had concerns about the impact a P/F grading option would have on student learning, motivation, and academic performance, and hence reviewed literature on such options; a significant number of works were found, and a few reviewed, for this paper [3-7].

A Princeton study conducted in [3] was the earliest P/F study reviewed, and like [4], the authors argue that grades are a constant extrinsic motivator that interferes with attainment of intrinsic motivation for learning that continues after graduation; on a side note, this connects well to ABET's criterion related to lifelong learning [5]. As well, the authors in [4] assert that grade pressure is highest for students who are motivated to succeed but don't have a history of high performance – a group that is already stressed by the challenges of higher education. In both studies [3, 4], they did not inform professors which students would be getting P/F grades, to avoid any grade bias; this was repeated in the CSU program discussed in this paper. Both studies tested a P/F system with longitudinal assessment under the assumption that initial grades would be lower but subsequent grades should be higher due to the opportunity for students to focus on learning over grades. Results for both studies, however, showed that those who entered the P/F program earned lower grades throughout their studies, even in subsequent courses. Notwithstanding, in [4], students who took only one course P/F outperformed all other students throughout their studies, suggesting that P/F courses and programs can be done right, but perhaps require a certain balance and/or support system. Additionally, the researchers in [3] solicited student attitudes from a questionnaire, which revealed that students preferred the traditional, competitive grading system and were less motivated to learn in P/F courses. If student attitudes are different today, such a program may have better results.

In [6], the authors' review of the literature on P/F programs revealed that a primary motivation for instituting them was to encourage exploration outside the student's academic major by reducing the risk to GPA in taking such classes; however, few students actually used it for this. This raised concern by the first author of this paper that students may simply be putting off the necessary learning of "getting good grades" in their major until their second semester, where they will then face the same problems that first-semester students in traditional grading do. The outcomes of the study in [6] also showed that students who had participated in the P/F option attained lower grades in classes both in-major and out-of-major in subsequent courses, another concern for CSU's program. However, methods of their study raised more questions, as students were stratified by their GPAs, implying that students can be grouped into "typical" grade levels; the authors of this paper question the validity of such methodology.

Finally, two very recent studies of implementing P/F grading in medical programs found that gains were made in students' psychological well-being, including stress levels, with no adverse effects on academic performance or licensing exam scores [7, 8]. However, both studies implemented the option for all incoming students in the *second* year of the study, which raises two questions: 1) might the second group of students (all P/F) have been better academically than first group (all traditional), and 2) might there have been bias from faculty grading knowing that all students were in this P/F program? These questions and others need to be answered by new and well-planned studies, which this work intends to initiate.

Overall, these studies show potential for a P/F program to improve psychological factors for students, as the second author is interested in attaining, but raises concern for academic performance, motivation, and learning, as the first author is apprehensive about. Underlying all of these studies, however, is the assumption that grades are a valid and reliable measure of student's learning and/or academic performance; a brief discussion of this assumption follows.

IV. A Brief History of Grades

The act of assigning and reporting grades is a standard of higher education; however, it is questionable if many people give thought to the history of grading, the role of grades, or the importance of grades. In the early years of higher education in the U.S.A., the common way to evaluate if a student should be conferred a degree was by *viva voce* oral examination, which often came in the form of open questioning or debate. Around the mid- to late-18th century, written examinations entered the scene, but were still largely evaluated on a pass/fail basis, only. In 1792 William Farish is said to have introduced numerical grading, while working at the mechanical engineering department at Cambridge, that was used to total points earned on the exam and then normalize and rank students into tiers. Around the late-19th century, the *viva voce* examination had all but been replaced by written exams evaluated with numerical grades, as is commonly used today. It was said that this change was supported highly by those who empathized with "timid and different" students (p. 34) but opposed by others who noted written exams could be easily nitpicked; further, the increase in the number of students necessitated this move as faculty were overwhelmed and backlogged with oral exams. However, some historical records also indicate that numerical and alphanumerical grading was used in the 18th and 19th centuries, but only on small-scale assignments. As well, the move to numerical and alphanumeric grading coincided with all students receiving the same exam questions, in efforts to ensure fairness, and a push by faculty to return to pass/fail oral examinations fell short in 1860 [9].

Today, grades are used to make inferences and decisions by many different stakeholders and for various purposes: employers to filter job applicants, colleges and universities for admission to graduate programs, honors and scholarship programs to select nominees, parents to evaluate their kids' performance, professors to track and compare students, and students themselves to evaluate their performance comparatively with peers. Douglas B. Reeves [10] notes that grades can be biased, can be based on inaccurate grading policies, and may or may not motivate students to improve their learning. Stuart Rojstaczer [11] studied the trend of grade inflation at various schools and programs over the years, and notes that it is nearly epidemic across higher education institutions.

Conducting an intensive study would be beneficial to better understand the different philosophies and applications of grades and grading, but will not be undertaken here. However, in this study, the authors were particularly concerned with the effect that the P/F option would have on students, with the basic understanding that graded classes may affect the orientation, magnitude, and prioritization of motivations. Since internalized, or intrinsic, forms of motivation are optimal for cognitive tasks like learning and performing in the classroom [12], it is fair to be concerned that student learning and performance will be affected by changes in grading policy.

ASSESSMENT PLAN

From an administrative viewpoint, the goal of the assessment plan was to measure the relative success of the two groups of students: those in the P/F system and those in the traditional grading system. Several metrics are being used to assess this project, to be conducted in three phases: 1) performance during the fall semester when the P/F system is active; 2) the following semester; and then 3) long-term, through graduation. Phase 2 and 3 are in progress, so this paper focuses on a portion of the Phase 1 assessment. During Phase 1, several measurements were compared, including: performance in an early grade feedback systems implemented at the university level, participation in several student support activities, and end-of-semester performance. Student performance was defined as student grade point average (GPA) at the end of the fall semester.

This was one of the points of contention between the two authors of this paper: the use of grades and GPA as a measure of student performance. The second author gravitated towards GPA, as it is a well-established metric of academic performance that would be readily understood by the administrators reviewing the project. The first author, on the other-hand, felt that, as discussed above, grades have many confounding variables attached to their creation and use, and hence may not be good indicators of program success. For example, Reeves [10] found that when a set of grades from a student's work (homework, quizzes, tests) were given to faculty that they all came up with different final grades for the student.

Again, the authors of this paper believe that this warrants future work researching grades in higher education, but for purposes of this work, grades and GPA were used.

RESULTS

I. Fall 2010 Enrollment

In the Fall 2010 semester, 138 of 448 first-year students enrolled in the program. Demographics of the two student populations – those in the P/F system and those in the traditional grading system – were collected and reviewed. The results were a surprise to both the authors and other administrators. The gender breakdown in the P/F group was 77.0% male and 23.0% female, while the entire first-year population distribution was 77.5% male and 22.5% female – remarkably similar for a purely voluntary project. A second comparison was made based on students' choice of major. Again, the result was great similarity between groups as the percentages were within 2-3% points. The one exception occurred with the open-option (undeclared) students; they made up 19.0% of the P/F group but only 13.0% of the entire population. Finally, the groups were compared based on ethnicity; based on the demographics, the two groups appear to be quite similar.

II. Overall GPA Comparisons

When the initial comparison of student performance was conducted, the average GPA for students in the P/F option was 2.53[1] and those in traditional grading was 2.55, a statistically-insignificant difference[2]. However, upon further review, it was discovered that the cohort of students not in the P/F program included some students retaking classes (due to poor performance the previous year) who maintained their status as first-year students and hence were not newly-entered students; as well, one student was in the project by error. Therefore, these students were removed and the cohort was re-evaluated. The GPA comparisons for those in the project to those not in the project changed to 2.55 and 2.82, respectively, resulting in statistically-significant differences in the two groups of students.

At this point, the disagreement between the authors led to a discussion of the literature on the use of GPA. An important paper was found that was quite critical of using GPAs for comparison between students [13]. The basic complaint by Soh is that grades are really classifications and as such are ordinal numbers, not cardinal numbers. So grades might be appropriate for ranking students but combining to calculate a composite GPA implies they are, in Soh's view, cardinal numbers. At this point, the second author acknowledged the problems with GPAs but was unwilling to abandon their use completely. After further discussion between the authors, however, it was decided that it would be better to limit comparisons of grades between the two cohorts on a course-by-course basis. The rationale was that if grades really do represent a ranking system (something still under debate), then limiting comparisons to individual courses results in a less-confounded analysis. It is also important to remember at this point that faculty were unaware of whether students were in the P/F program, so grading was not biased by knowledge of students' choices.

III. Engineering Courses

The first courses reviewed were the first-year engineering courses. The departments provide four introductory courses: CBE 101-Chemical and Biological; CIVE 102-Civil and Environmental; ECE 102-Electrical and Computer; and MECH 100-Mechanical Engineering. The fifth course, ENGR 180, is for open-option students and taught for students in the college who have not chosen a major.

In Table 1, the mean GPAs for both groups of students are shown for the introduction to engineering courses. In all courses (except MECH 100, where all students received A's) the students in the P/F program consistently scored lower than students not in it. This data was preliminary evidence that students in the project are not performing as well as students in the program.

If we take Soh's [13] comment that grades are ordinal then an alternative perspective is warranted; see Figure 1. Here, the distribution of grades is shown for the engineering courses (excluding MECH 100). The results in Figure 1 show that the distribution of students into the traditional grade classifications indicates an important difference between the two groups of students (Note that the grading system at our university does not include grades of C-, D+, or D-). The students in the P/F program appear more normally distributed, with the average near a grade of B. Students *not* in the P/F program, on the other hand, are more evenly distributed across the range of grades, with a greater number of students in the A-range.

IV. Math and Science Courses

A similar analysis of the students' GPA performance was conducted for the common math and science courses taken during the Fall 2010 semester. Table 2 shows results

Table 1: Mean GPAs for introduction to engineering course

Course	Pass-Fail	Traditional	Overall
CBE 101	2.58	3.10	2.98
CIVE 102	2.76	3.04	2.96
ECE 102	2.51	3.07	2.88
ENGR 180	3.65	3.78	3.72
MECH 100	4.00	4.00	4.00
Overall	3.19	3.41	3.34

[1] All grades and GPAs are on a 4.0 scale.
[2] Means were compared with t-tests assuming unequal variances. Statistical significance was evaluated at the $p < 0.05$ level.

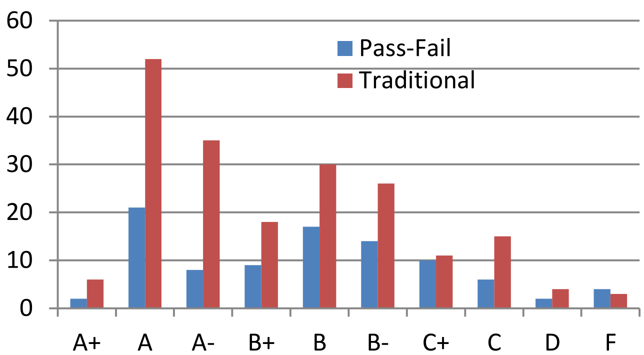

Figure 1: Grade distributions in engineering courses

for four courses: Calculus I and II (MATH 160,161), Intro to Chemistry (CHEM 111), and Physics (PH 141). The trend of students in the P/F project performing at a lower GPA level continues for all of these courses.

Following, grade distributions for each of the science classes were reviewed; Figure 2 shows the grade distribution for CHEM 111. This graph is very different from the grade distribution for the engineering courses. Both groups of students show very similar grade distributions: few "A" grades and a tendency towards the lower end of the grade scales. Figures 3-5 give the distributions for MATH 160, MATH 161, and PH 141, respectively. Here, the distributions appear similar to the engineering grades: P/F students tend to be normally distributed while non P/F students are more evenly distributed, with more "A" grades.

Discussion of Results

Although GPA data represents only a small component of the evaluation for this project, it has already revealed several issues that will require deeper investigation.

First, what do the differences in GPA scores for the two groups mean about the success of this project? As mentioned above, the use of GPA scores was a topic of disagreement between the authors of this paper. The results presented herein clearly indicate that the two groups of students are performing at different levels, as measured by GPA. However, an important next step still needs to take place. Although the two student populations are nearly identical across several demographics, there are important demographics that have not been evaluated in time to publish in this paper. For example, the level of student preparation has not been compared. Since this was a voluntary project, the question arose: Did the students self-select in a manner such that those with weaker academic records participated while students with stronger academic records did not feel the need to? Although this cannot definitively be answered, data are being collected and analyzed in an attempt to answer this question as best as possible.

Recognizing the potential issues with GPA, the data showing the distribution of grades, as opposed to average GPA scores, resulted in other intriguing questions: Why are the grades for the P/F students consistently shaped close to a normal bell curve while the non P/F students tend to have much more uniform distributions? Why are the grade distributions similar in CHEM 111 but not in the other classes? Again, these questions will be difficult to answer.

Table 2: Mean GPAs for engineering courses

Course	Pass-Fail Mean	Traditional Mean	t-test p-value
CHEM 111	1.93	2.21	0.06
MATH 160	2.24	2.44	0.18
MATH 161	1.86	2.75	0.04*
PH 141	2.24	2.52	0.30
Overall Mean	2.09	2.40	<0.01**

* $p < 0.05$ ** $p < 0.01$
Note: Did not assume equal variances across groups.

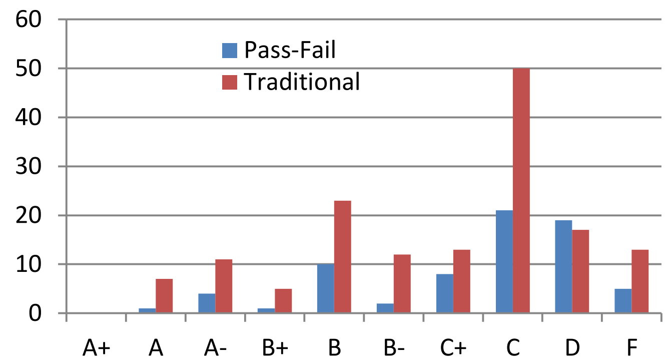

Figure 2: Grade distribution for CHEM 111

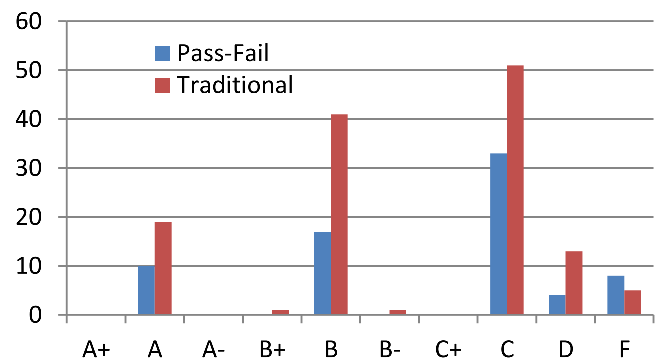

Figure 3: Grade distribution for MATH 160

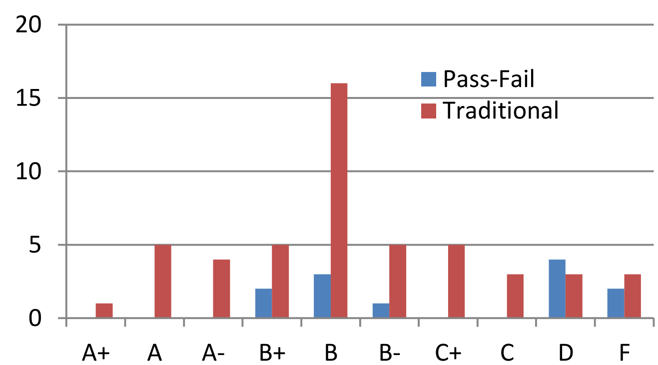

Figure 4: Grade distribution for MATH 161

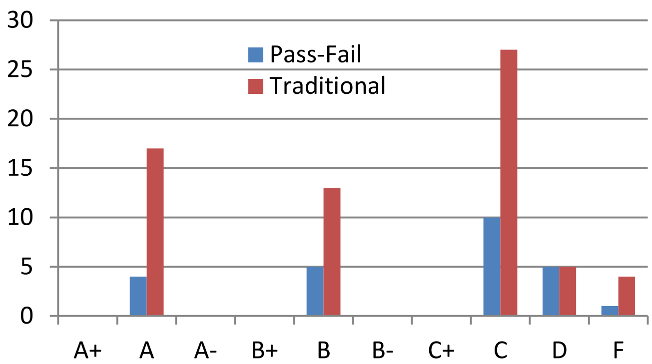

Figure 5: Grade distribution for PH 141

CONCLUSIONS

The genesis of this paper came from a discussion between the two authors concerning a new experimental project underway at Colorado State University. The first author expressed misgivings about the project based on his background in engineering education. The second author felt this was the right type of experiment to try, as an administrator, in an effort to assist first-year engineering students with their transition into the rigorous curriculum.

As discussions between the authors continued, they started to evaluate the assessment data from the project. Due to space limitations of this paper, focus has remained on the portion of the assessment that used student GPAs. This was determined to be an important metric at the beginning of the project and represents data most administrators were interested in seeing. Because this was such a point of disagreement, the authors reviewed literature that led to them rethinking their analysis. Hence, analysis switched from solely looking at mean values to look more closely at individual course grade distributions. As the data show, differences exist between students in and not in the program.

The differences in student performance bear semblance to other schools' efforts to implement P/F grade options, as discussed in the background section, which raises concern for the success of this program. This emphasizes the need to evaluate the other variables involved, including the additional contact time with faculty and the seminar class.

Major future efforts are focused on two concepts related to this project. First, the efforts around student performance continue, including development of a better characterization of student academic preparation heading into the Fall semester, based on metrics such as high school records and standardized test scores.

A second focus is on students' attitudes and perceived stress levels during the fall semester. Although the original plan for a national, mental well-being survey being available did not occur, a couple of surveys were taken that address these student issues. Analysis of these data is in progress.

Long-term, there are numerous assessments in progress. They include, for example, student persistence in the major, performance in sequential courses in the Spring semester, and GPAs earned in subsequent semesters.

The ultimate success of this project will not be known for some time, as the three phases of assessment, discussed at the beginning of the paper, are not yet complete. However, one very valuable lesson was learned: the open discussion and disagreement between the authors has led to a more in-depth evaluation of the assessment data collected that will be reported to administrators. It is unfortunate that these types of disagreements are rarely openly discussed in the literature, as they can lead to improvements in engineering education. Neither author can claim to be exclusively right about this project, but both feel that this experimental project will benefit from the contention.

REFERENCES

[1] R. Flammang, "No-Grades Policy to go to Committee Tomorrow," in *California Tech* vol. LXVII, 24 ed. Pasadena, CA: Associate Students of the California Institute of Technology, 1966, p. 2.

[2] D. E. Rohe*, et al.*, "The Benefits of Pass-Fail Grading on Stress, Mood, and Group Cohesion in Medical Students," *Mayo Clinic Proceedings,* vol. 81, pp. 1443-1448, Nov 01 2006.

[3] M. Karlins*, et al.*, "Academic Attitudes and Performance as a Function of Differential Grading Systems: An Evaluation of Princeton's Pass-Fail System," *The Journal of Experimental Education,* vol. 37, pp. 38-50, 1969.

[4] R. M. Gold*, et al.*, "Academic Achievement Declines under Pass-Fail Grading," *The Journal of Experimental Education,* vol. 39, pp. 17-21, 1971.

[5] ABET, Inc.. (2009). *Criteria for Accrediting Engineering Programs: Effective for Evaluations During the 2009-2010 Accreditation Cycle*. Available: http://abet.org/forms.shtml

[6] P. T. Bain*, et al.*, "An Investigation of Some Assumptions and Characteristics of the Pass-Fail Grading System," *The Journal of Educational Research,* vol. 67, pp. 134-136, 1973.

[7] R. A. Bloodgood*, et al.*, "A Change to Pass/Fail Grading in the First Two Years at One Medical School Results in Improved Psychological Well-Being," *Academic Medicine,* vol. 84, pp. 655-662 10.1097/ACM.0b013e31819f6d78, 2009. *Proceedings,* vol. 81, pp. 1443-1448, November 1, 2006.

[8] D. E. Rohe*, et al.*, "The Benefits of Pass-Fail Grading on Stress, Mood, and Group Cohesion in Medical Students," *Mayo Clinic*

[9] C. Stray, "From oral to written examinations: Cambridge, Oxford and Dublin 1700-1914," *History of Universities,* pp. 76-130, 2005.

[10] D. B. Reeves. (2009, January 16, 2011). Remaking the Grade, From A to D. *The Chronicle of Higher Education*. Available: http://chronicle.com/article/Remaking-the-Grade-From-A-to/48352/

[11] S. Rojstaczer. (2009, Jan 16). *National Trends in Grade Inflation, American Colleges and Universities.* Available: http://www.gradeinflation.com/

[12] J. D. Bransford*, et al.*, *How People Learn*, Expanded ed.: National Academy Press Washington, DC, 2000.

[13] K. C. Soh, "Grade point average: what's wrong and what's the alternative?," *Journal of Higher Education Policy and Management,* vol. 33, pp. 27-36, Mar 2011.

AUTHOR INFORMATION

Ken Stanton, Postdoctoral Fellow, Mechanical Engineering, Colorado State University, ken.stanton@colostate.edu.

Tom Siller, Associate Dean for Academic and Student Affairs, Colorado State University, thomas.siller@colostate.edu