AC 2010-2189: GRADE-BASED CORRELATION METRIC TO IDENTIFY EFFECTIVE STATICS INSTRUCTORS

Randall Manteufel, University of Texas, San Antonio

Dr. Randall D. Manteufel is Associate Professor of Mechanical Engineering at The University of Texas at San Antonio where he has taught since 1997. He received his Ph.D. degree in Mechanical Engineering from the Massachusetts Institute of Technology in 1991. His teaching and research interests are in the thermal sciences. He is the faculty advisor for ASHRAE at UTSA. Manteufel is a fellow of ASME and a registered Professional Engineer (PE) in the state of Texas.

Amir Karimi, University of Texas, San Antonio

Amir Karimi is a Professor of Mechanical Engineering and the Associate Dean of Undergraduate Studies at The University of Texas at San Antonio (UTSA). He received his Ph.D. degree in Mechanical Engineering from the University of Kentucky in 1982. His teaching and research interests are in thermal sciences. He has served as the Chair of Mechanical Engineering (1987 to 1992 and September 1998 to January of 2003), College of Engineering Associate Dean of Academic Affairs (Jan. 2003-April 2006), and the Associate Dean of Undergraduate Studies (April 2006-present). Dr. Karimi is a Fellow of ASME, senior member of AIAA, and holds membership in ASEE, ASHRAE, and Sigma Xi. He is the ASEE Campus Representative at UTSA, ASEE-GSW Section Campus Representative, and served as the Chair of ASEE Zone III (2005-07). He chaired the ASEE-GSW section during the 1996-97 academic year.

Metrics for Instructor Effectiveness
Based on Student Success in Courses

Abstract

Grade-based metrics are used to gauge instructor effectiveness. The final grade distributions for 24 classes of engineering statics, taught by 10 instructors over a five-year period are evaluated. A null hypothesis is that the grade point average (GPA) is no different than that issued by other instructors for the same course. In two cases, the null hypothesis is rejected, showing that one instructor is distinctly more lenient and one is harsher in their grade distributions. Data shows there can be significant class to class GPA variation for the same instructor, so class GPA is not proposed as a sufficient metric of an instructor's effectiveness. Students passing statics are tracked into three follow-on engineering courses: dynamics, solid mechanics, and thermodynamics. A correlation coefficient based on the statics grade and follow-on grade is proposed as a better measure of the statics instructor's effectiveness. The null hypothesis is that there is no difference between grade correlations for the statics instructors. The null hypothesis can't be rejected in most cases, implying that this metric doesn't identify which statics instructor is better at preparing students for subsequent courses. Although the correlations are weak, trends are discernable where students who succeed in passing statics taught by an instructor who has a reputation of being more rigorous, do better in the follow-on courses. At best, the grade-based correlation metric explains up to 25% of the future grade success in follow-on engineering courses for the most effective statics instructors.

Introduction

There is much discussion of the need to continuously improve our programs, curriculum, and courses[1]. The improvement is driven by assessments, evaluations, and feedback from both inside and outside the college. Feedback from employers, national associations[2] and leaders from the community frequently provide high-level guidance to improve engineering programs. One consistent theme is that the program and course needs to be preparing students with the right skills and capabilities to succeed in their future endeavors. It appears logical that foundational engineering courses prepare students with the fundamentals needed to succeed in subsequent courses. End of semester grades are the ultimate measure of a student's success in a class, which is assumed to be highly correlated with the learning (defined as the knowledge, skills, abilities and attitudes[2]) achieved by the student by the end of the course.

Although grades are used to assess student performance, there appears to be little use of grade-based correlations to identify instructors that do a better job of instruction in fundamental courses[3]. A survey of strategies to measure teaching effectiveness[4] lists 12 possibilities: student ratings, peer ratings, self-evaluation, videos, student interviews, alumni ratings, employer ratings, administrator ratings, teaching scholarships, teaching awards, learning outcomes, and teaching portfolio. Of these, the tracking of student grades in subsequent engineering courses doesn't appear to be used as a metric, so it is investigated in this paper.

In a recent National Academy of Engineering report, the recommendation is to "use of multidimensional metrics that draw upon different constituencies to evaluate the content, organization, and delivery of course material and the assessment of student learning." In addition, the consensus[2] is that "metrics for assessing teaching, learning and instructional effectiveness are not well defined or well established." We agree with this finding and offer this study looking at a quantitative grade-based metric that attempts to get at the core result: did the students learn anything in the course. If they learned much in the prerequisite course, they should be well-prepare for success in subsequent follow on courses.

In the first few semesters of an engineering program, it is expected that foundational engineering courses prepare students with basic engineering skills needed for other engineering courses. The idea of prerequisite courses is throughout engineering curriculum where students start and proceed through a sequence of courses learning and gaining new capabilities. If a student earns an "A" in a foundational engineering class, they should be well-prepared to succeed in subsequent courses. Likewise, a grade of "C", indicates that the student is adequately prepared for subsequent courses. But experience often contradicts this notion that grades are a good indicator of preparedness for subsequent courses. On many occasions, instructors detect the effect of earlier instructors who teach foundational prerequisite courses. It is common for instructors to complain that "students aren't coming into my class prepared." It is a common observation that students have passed important foundational courses, yet lack perspectives or skills they should have acquired in the foundational courses. This is part of the feedback process and plays an important role in the assessment of an instructor's teaching effectiveness.

There can be significant variation between instructors in both rigor and coverage of material in foundational courses. It appears worthwhile to perform a quantitative statistical assessment of a teacher's effectiveness using grades as the primary measure. The purpose of this paper is to assess the instructor's effectiveness using grades earned in the foundational course and grades earned in the follow on engineering courses.

Grades

Grades are the greatest single indicator used to measure student success in a class. There appears to be overwhelming consensus that grades are a reasonable indicator of student mastery of the material. A student's grade point average (GPA) is often a key factor in determine admission to an institution, admission to a program major-sequence of courses, or admission to graduate school. There appears to be reasonable consensus that grades from all classes and all instructors give an overall assessment of the student's performance.

Instructors assign grades based on student mastery of the material. Often there are significant differences between instructors for the grades assigned in the same course. One can detect if an instructor gives high or low grades compared with historical data for the course. Statistical difference doesn't establish causality, but do indicate unreasonable variability.

Table 1 provides a summary of the final grades for 24 classes of engineering statics taught by 10 different instructors from the fall 2004 to summer 2009. The class size format varied from 16-week long semesters with two- or three-meeting per week, to 10-week summer sessions meeting twice per week. Over the 5 year period, 860 students enrolled or attempted statics. Of these students, 535 passed the statics course and were tracked into subsequent engineering courses.

The data includes those students who withdrew from the class after census date which is typically in the second week of class. At the University of Texas at San Antonio, students can withdraw from a class up to the tenth week in a 16 week semester. The typical class size is less than 70 students. Only 6 of the 24 classes had more than 50 students. The total number of student is N.

Table 1. Grades for ten instructors teaching 24 classes of engineering statics.

Instructor	**Class**	**A**	**B**	**C**	**D**	**F**	**W**	**N**	**GPAW**	**stdGPAW**
i1	1	3	4	4	1	5	3	20	1.65	1.66
i2	2	10	13	13	4	6	4	50	2.18	1.45
i3	3	3	5	7	0	2	2	19	2.16	1.42
i4	4	4	9	10	5	14	10	52	1.31	1.48
i5	5	1	2	2	1	2	0	8	1.88	1.46
i5	6	3	4	10	2	2	2	23	2.00	1.30
i6	7	3	5	18	3	6	1	36	1.83	1.17
i7	8	0	2	4	2	1	0	9	1.78	0.97
i6	9	6	11	13	3	3	8	44	1.95	1.52
i5	10	2	3	8	7	6	2	28	1.43	1.25
i8	11	3	3	6	3	4	6	25	1.44	1.60
i6	12	9	11	15	6	4	9	54	1.94	1.54
i7	13	4	7	6	4	5	2	28	1.89	1.46
i6	14	13	11	11	5	2	2	44	2.55	1.33
i8	15	4	12	13	2	5	5	41	1.95	1.39
i8	16	10	15	9	7	5	10	56	1.96	1.64
i7	17	7	13	7	6	4	3	40	2.18	1.43
i6	18	5	4	1	1	7	25	43	0.81	2.05
i8	19	17	14	9	0	2	3	45	2.84	1.34
i9	20	8	4	4	2	14	3	35	1.54	1.76
i10	21	5	5	11	3	8	3	35	1.71	1.47
i8	22	19	10	17	4	4	5	59	2.44	1.50
i6	23	2	5	5	4	5	10	31	1.19	1.56
i6	24	4	8	7	8	5	3	35	1.77	1.39
Sum		145	180	210	83	121	121	860	1.89	1.45

Table 1 and Fig. 1 show the class GPAW (mean and standard deviation) for each of the 24 classes in chronological order. The instructor who taught the class is indicated as i-1, i-2, i-3, etc. The mean GPA was calculated on a 4.0 scale with A-4, B-3, C-2, D-1, and F-0. Because of the generous withdrawal policy, many students who withdraw do so because of poor academic performance after the first or second mid-term exam. The exact date of when a student withdraws was not analyzed, because some student may withdraw for a variety of nonacademic reasons. But this was not evaluated in the study. In the computation of the GPA, the withdrawals are treated as "F" grades, hence the designation of GPAW. The equation for the mean GPAW is:

$$M_{GPAW} = \frac{4n_A + 3n_B + 2n_C + 1n_D + 0n_F + 0n_W}{n_A + n_B + n_C + n_D + n_F + n_W}$$

Fig. 2 shows the percent of students having a final grade of D or F, or having withdrawn from the class. This is listed as the "DFW" rate which can be considered as the failure rate. This shows that classes have a significant variability in DFW rate. It appears that some instructors have relatively low DFW rates while others high. Some have a significant difference in both GPAW and DFW rate.

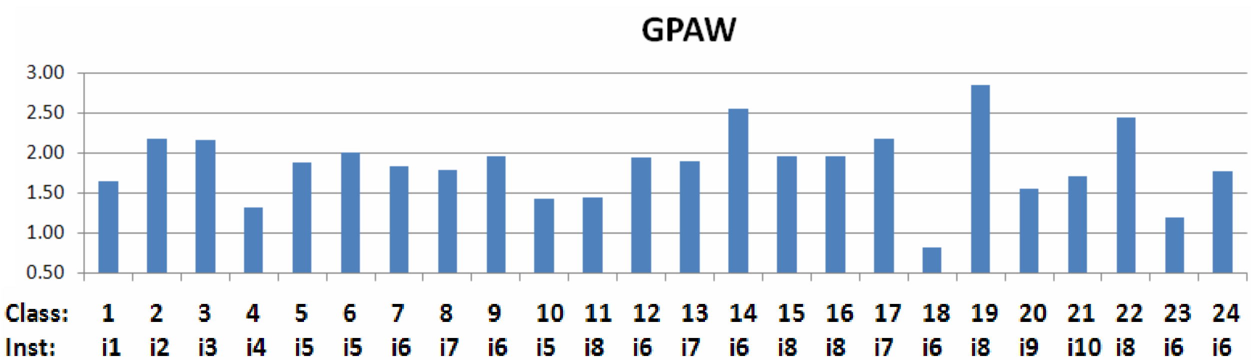

Fig. 1. Class Grade Point Average (GPAW) treating students who withdrawal from the course the same as earning "F" grades.

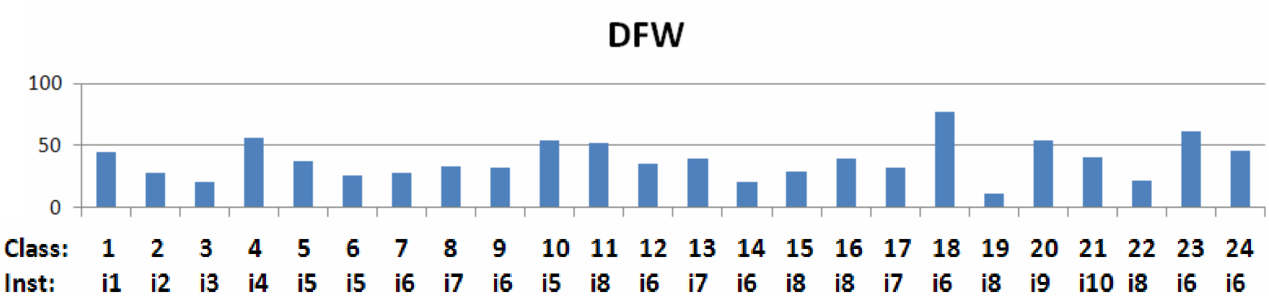

Fig. 2. Percentage of students not passing the class (DFW rate).

From Figs. 1 and 2, some trends are distinguishable, especially for classes 3, 14, 19 and 22 which have low DFW rates while high GPAWs. Given the importance of student retention and graduation rates, this type of data is increasingly being used by administrators to identify the most effective instructors. Since the goal of each student is to pass the class, instructors with low GPA's and high DFW rates can be viewed as not being effective instructors. Likewise, those with high GPA and low DFW are often viewed as effective instructors. Although the GPA and DFW rates are simple metrics to obtain, there should be much caution in their use since it is readily apparent that the simplest way for an instructor to manipulate these types of metrics is to be more lenient in grading or less rigorous in course content to boost GPA.

Knowing that GPA and DFW are metrics, an analysis is continued to compare instructors and classes. When comparing the GPAW between classes, one needs to calculate the standard deviation of the GPAW:

$$S_{GPAW} = sqrt\left(\frac{n_A(4-M_{GPA})^2 + n_B(3-M_{GPA})^2 + n_C(2-M_{GPA})^2 + n_D(1-M_{GPA})^2 + (n_F + n_W)(0-M_{GPA})^2}{n_A + n_B + n_C + n_D + n_F + n_W - 2}\right)$$

To compare either a class GPA or instructor GPA with the average for the course (over other classes and other instructors), one uses the pooled error:

$$PE = \sqrt{\frac{S_1^2 n_1 + S_2^2 n_2}{n_1 + n_2 - 1}}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

Here, n_1 is the total number of students in the class being evaluated and n_2 is the total number of students having taking the course in the past 5 years excluding those who took the course with the same instructor. A number of different ways were evaluated before deciding to exclude the same instructor from the course average. One instructor could possibly dominate the average if they taught the class more times with more students compared with other instructors. Therefore, when evaluating a course, all students taught by the same instructor were excluded in the computation of S_2 and n_2. For the data analyzed here, the instructor-6 (i6) taught 287 students in 7 classes and instructor-8 (i8) taught 226 students in 5 classes. These two are the only instructors who could possibly dominate the course 5-year statistics.

The Z statistic is used to determine if there is a significant difference in the mean GPAW for each class compared with the 5-year course average (excluding that from the instructor teaching the class).

$$Z = \frac{M_{GPAW,class} - M_{GPAW,otherclasses}}{PE}$$

Fig. 3 shows the Z-statistic for each class. A Z greater than about 2.0 indicates the class has a significantly higher GPA while those below -2.0 indicate significantly lower GPA. The criteria for determining significance is at the 0.05 level using 2-tailed criteria. The largest deviation is for instructor-8 teaching the 19^{th} class which had a high GPAW and instructor-6 teaching the 18^{th} class with a low GPAW. Overall, three classes stand out as having high student success: instructor-6/class-14, instructor-8/class-19, instructor-8/class-22. Likewise, three standout as having low student success: instructor-4/class-4, instructor-6/class-18, and instructor-6/class-23. It is interesting that the same instructor may have a class with either high or low GPAW, indicating other possible affects are present. Difference may be due to the arrival of a group of students who were poorly prepared in a preceding course, typically math or physics. Or, the difference may be due to changes in the textbook or changes in the presentation of the course. If an instructor tries a new instructional technology or alternative pedagogical strategy, this can have an impact. Or, the difference may be due when the class was offered (16-week spring/fall, 10-week summer, 5-week summer), or the time of day (8 am Tuesday-Thursday class, etc). Another significant difference is due to the individual student's attitude and peer influences in the class. Having taught a number of classes, the attitude of the class can be affected by a few students who influence other students with either a positive or negative attitude toward the university, college, department, course, or instructor. In this work, these influences are recognized but no attempt is taken to quantify them. Overall, these influences may be just as significant as the influence of the instructor. At the very least, these other factors make it more difficult to develop robust quantitative metrics.

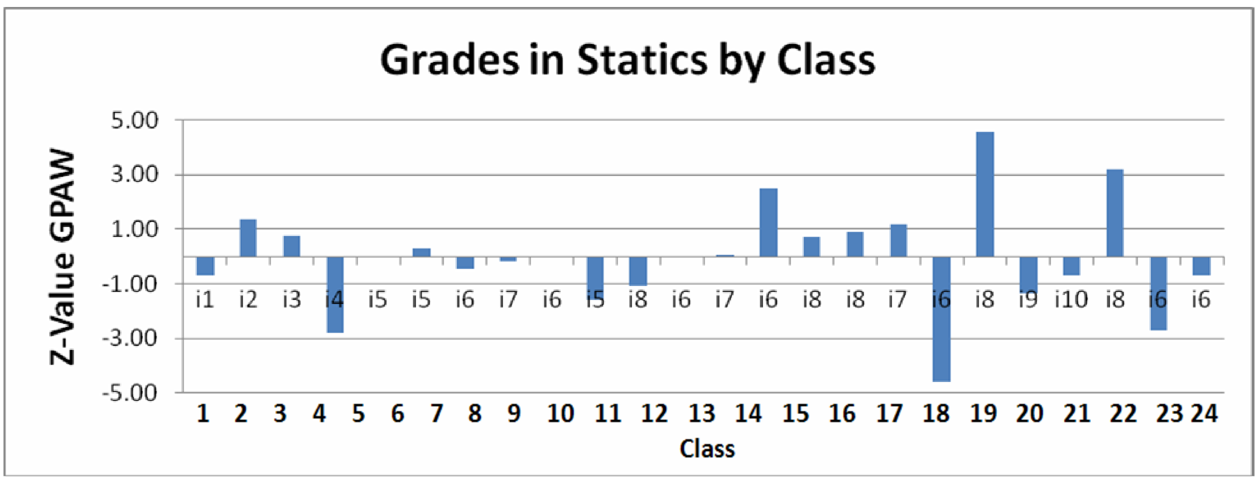

Fig. 3. Comparison of the mean GPAW for each class. Z values above 2 indicate the course has significant higher GPAW and values below -2 indicate significantly lower GPAW.

Fig. 4 shows the cumulative GPAW for each instructor compared with the average of all other instructors. Using a 0.05 significance level, GPAW's are compared, and instructor-8 issues significantly higher grades ($Z = +3.85$) while instructor-4 issues significantly lower grades ($Z = -2.99$). Other instructors are not statistically significant.

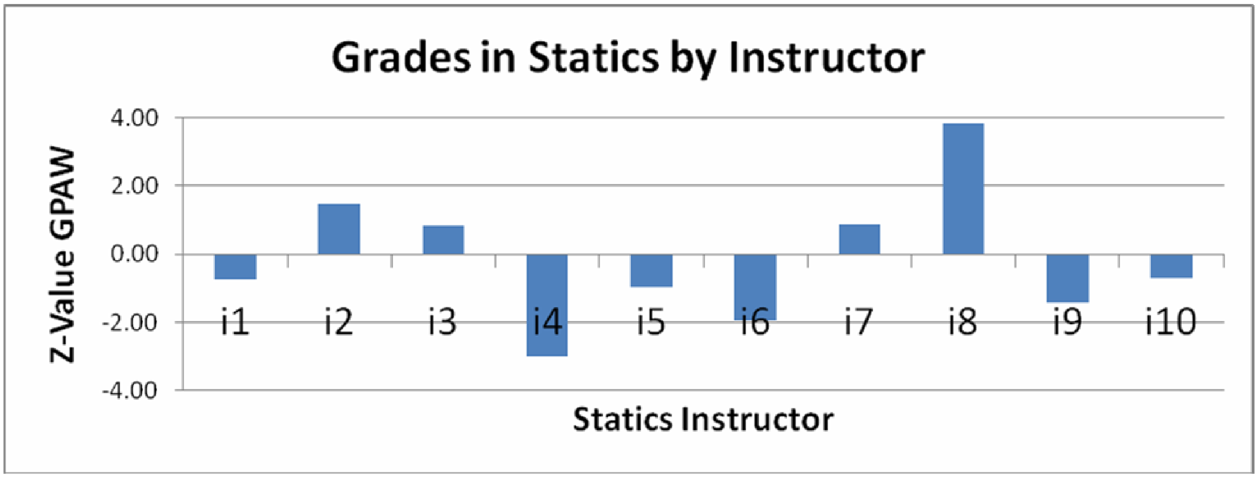

Fig. 4. Comparison of the mean GPAW on an instructor basis. Z values above 2 indicate the course has statistically significant higher mean GPAW. Z values below -2 indicate the class has a significantly lower GPAW.

The Z-statistic is often what students intuitively use when enrolling in a class. Students talk about instructors and share their perspectives and experiences. Students describe instructors as being easy or hard, or some similar descriptors that captures the idea contained in Figure 4. Students want to enroll in a course taught by an instructor where the historical data shows that previous students succeeded in the course.

Another interesting point which is only noted here but not explored further, is that for the same instructor there can be significant variations from class to class. Again, this points to the many other factors which influence student learning which were not quantitatively considered in this work. From the author's perspective, the two most important other factors are: (1) rigor of preparation in the calculus-based prerequisite physics course, and (2) peer influence where as few as one student can disseminate either a positive or negative attitude toward the instructor/course/textbook/room/etc. Although important, this paper focuses on grades earned in the class. To evaluate student learning in statics, those who passed the statics class (with A, B or C) were tracked into the follow on engineering classes where subsequent performance (A, B, C, D F or W) is compared with performance in the statics class. Because of the scarcity of the data, the students are tracked by the statics instructor (of which there are 10), and not necessarily by the statics class (of which there are 24).

Subsequent Courses

The students who passed statics were tracked into three engineering courses: dynamics, solid mechanics and thermodynamics. It was decided to track students into these three courses because the only prerequisite for each course is the statics course. A fourth course in materials engineering was considered, but not tracked because of its requirement of prerequisite chemistry course. All other engineering courses require either the dynamics, solid mechanics, thermodynamics or materials course as prerequisite(s).

An example of student performance data collected and used in this work is summarized in table 2 for instructor-8 from students having passed dynamics and the result of their first attempt at taking dynamics. Only the first attempt of the follow on course was considered. Students who earned D/F grades or withdrew from the course were not considered even if they reenrolled in the course.

Table 2. Student grade performance from Statics to Dynamics for Instructor-8.

Grade Earned in Dynamics	**Grades Earned in Statics**			Sum
	C	**B**	**A**	
	38	42	39	**119**
A	5	0	18	**23**
B	8	16	10	**34**
C	12	16	7	**35**
D	2	4	2	**8**
F	7	3	0	**10**
W	4	3	2	**9**

A total of 119 students passed statics with instructor-8 and attempted dynamics. Of these, 39 earned A grades in statics, 42-B and 38-C. Of the 39 who earned A's, these students proceeded in dynamics and 18 earned A's, 10-B's, 7-C's, 2-D's, 0-F's, and 2-W's. Overall, the raw data confirms that good students who earn good grades in one class (like statics) typically earn good grades in other classes (like dynamics). Likewise, those who earn C's in statics tend to earn C's in dynamics. This general trend was observed in all of the data. In some ways it is both humbling and reassuring that student grade performance is irrespective of the instructor. It is humbling since some of the best instructional strategies will fail to motivate some of the students who at times appear content with C's or content with failing. This is qualitatively confirmed by student's who often admit that their poor performance was due to their immaturity or lack of diligence. The result is also reassuring because good students tend to thrive with the best instructor or survive the worst. Student learning is primarily the function of the student.

Using the raw data, one computes the correlation coefficient, r, for the grades earned in both classes. To do this, one computes the mean GPA earned from those passing statics:

$$GPA_1 = \frac{4n_{A1} + 3n_{B1} + 2n_{C1}}{n_{A1} + n_{B1} + n_{C1}}$$

And the mean GPA earned in dynamics. Here again, a withdrawal is viewed as earning an F. There are many legitimate reasons to withdraw from a class, but this study lacked the ability to discern why the individual student withdrew. Based on experience, many students withdraw after failing a major exam so the W's are treated as F's in this study. Also, we only track the first attempt of the student in dynamics, since some students do repeat dynamics before earning a passing grade.

$$GPA_2 = \frac{4n_{A2} + 3n_{B2} + 2n_{C2} + 1n_{D2} + 0n_{F2} + 0n_{W2}}{n_{A2} + n_{B2} + n_{C2} + n_{D2} + n_{F2} + n_{W2}}$$

The covariances are computed as:

$$\mathrm{cov}_{11} = n_{A1}(4 - GPA_1)^2 + n_{B1}(3 - GPA_1)^2 + n_{C1}(2 - GPA_1)^2$$

$$\mathrm{cov}_{22} = n_{A2}(4 - GPA_2)^2 + n_{B2}(3 - GPA_2)^2 + n_{C2}(2 - GPA_2)^2 + \\ n_{D2}(2 - GPA_2)^2 + n_{F2}(2 - GPA_2)^2 + n_{W2}(2 - GPA_2)^2$$

$$\mathrm{cov}_{12} = n_{A1A2}(4 - GPA_1)(4 - GPA_2) + n_{A1B2}(4 - GPA_1)(3 - GPA_2) + n_{A1C2}(4 - GPA_1)(2 - GPA_2) \\ + n_{A1D2}(4 - GPA_1)(1 - GPA_2) + n_{A1F2}(4 - GPA_1)(0 - GPA_2) + n_{A1W2}(4 - GPA_1)(0 - GPA_2) + \ldots$$

The correlation coefficient is then computed as:

$$r = \frac{\mathrm{cov}_{12}}{sqrt(\mathrm{cov}_{11}\,\mathrm{cov}_{22})}$$

For the data in table 2, one computes a correlation coefficient r = 0.368. One observation is that there is a positive correlation between statics and dynamics grades, which should be expected. Given the number of students tracked, one estimates a 95% confidence interval for the true correlation using a Z-Fisher transformation[5,6]. For the data shown, the confidence interval is from 0.20 to 0.51. There is a strong confidence that r is above zero. A high value of r is desirable, because those students who earned A's in statics went forward and did well in dynamics. Likewise, those who earned C's in statics went forward and may have passed dynamics, but did not do as well as those who earned A's in statics. This is based on a basic assumption that students are well prepared for a follow on course if they earn high grades in prerequisite courses. So those who earn A's in statics are better prepared for dynamics than those who earned C's. The data confirms this and the grade correlation is proposed as an indicator of the static instructor's effectiveness.

For comparison, the data for instructor-6 are shown in Table 3.

Table 3. Student grade performance from Statics to Dynamics for Instructor-6.

Grade Earned in Dynamics	Grades Earned in Statics			Sum
	C	**B**	**A**	
	49	39	29	**117**
A	2	7	10	**19**
B	6	9	11	**26**
C	13	14	6	**33**
D	7	4	1	**12**
F	9	3	0	**12**
W	12	2	1	**15**

The total number of student tracked for both (Table 2 and 3) are nearly the same (119 versus 117 students). For instructor-6, the grade correlation is r=0.518, which is 0.15 greater than that for instructor-8. The 95% confidence interval is from 0.37 to 0.64. Because the intervals overlap, it is difficult to ascertain if one is significantly greater than the other. As a final example of the treatment of the raw data, the instructor with the next highest number of students that were tracked is instructor-9 with the data summarize in Table 4.

Table 4. Student grade performance from Statics to Dynamics for Instructor-9.

Grade Earned in Dynamics	Grades Earned in Statics			Sum
	C	**B**	**A**	
	14	20	11	**45**
A	4	3	8	**15**
B	5	7	3	**15**
C	2	6	0	**8**
D	0	1	1	**2**
F	2	0	2	**4**
W	1	3	1	**5**

Using the data in Table 4, for instructor-9, r=0.323 with the true value expected between 0.03 and 0.56. When one compares these three instructors, it appears the ranking of effectiveness goes from instructor -6 (best with r=0.518), -8 (r=0.368) and then -9 (r=0.323).

Fig. 5 allows one to determine if there is a significant difference between the grade correlation of the instructors for statics. Only 6 instructors had at least 10 students to track from statics into dynamics. The first 4 instructors have insufficient numbers of students so their correlations are not plotted. Fig. 5 shows both the computed r-value as well as the confidence interval. For instructor-9, there were only 14 students to track and this instructor has a negative correlation of -0.05 yet a large confidence interval. The diameter of the dot as well as the interval size is controlled by the number of students.

Figs 5 and 6 track the students from statics into solid mechanics and thermodynamics. Again, data is not plotted for those instructors with less than 10 students to track.

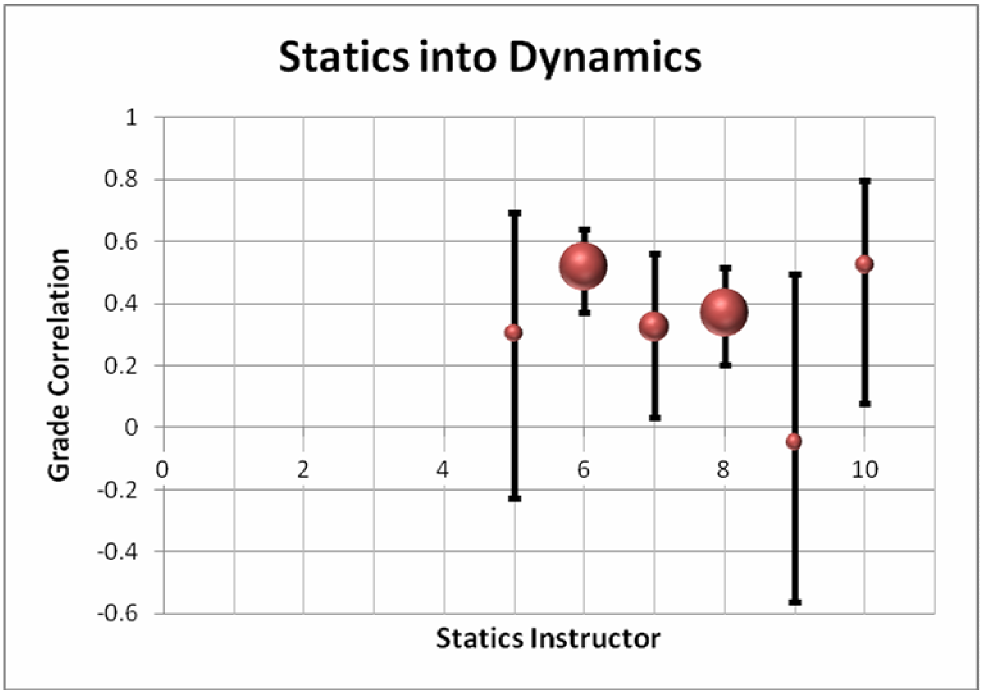

Fig 5. Grade correlation from statics into dynamics for the 10 statics instructors.

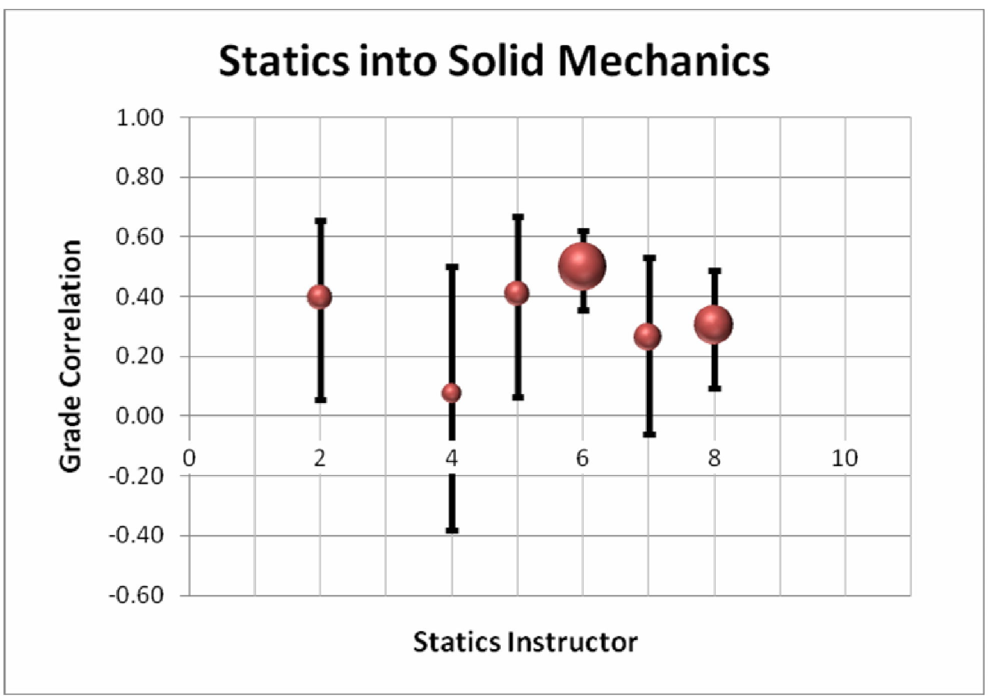

Fig 6. Grade correlation from statics into solids for the 10 statics instructors.

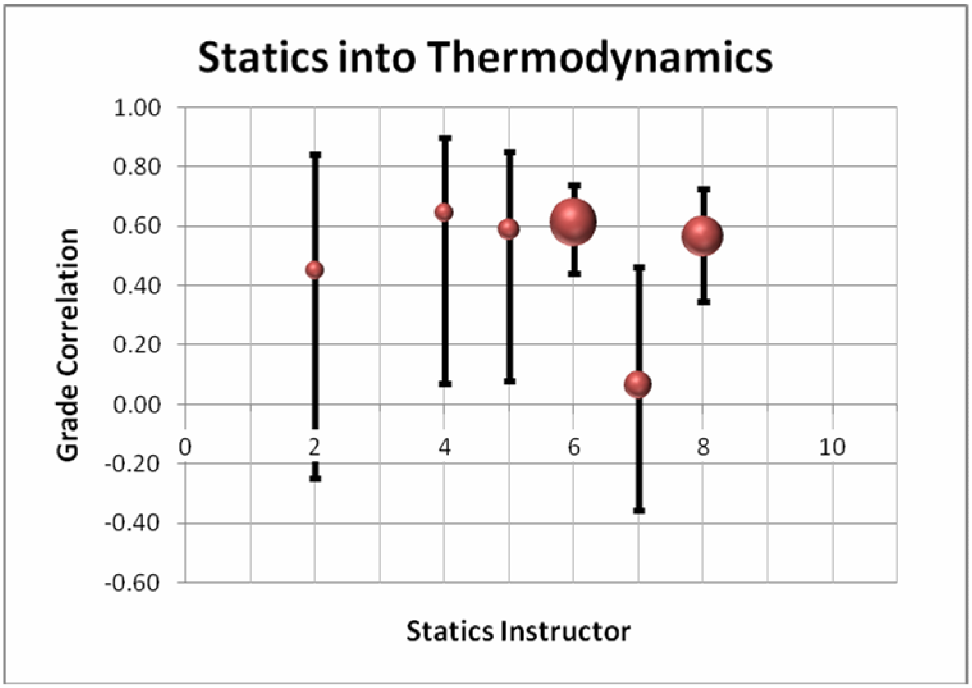

Fig 7. Grade correlation from statics into thermodynamics for the 10 statics instructors.

From Figs. 5-7, one discerns that some instructors have stronger grade correlations than other instructors. A more quantitative evaluation is performed to see if there is a significant difference between the two correlation values. The approach assumes both are from two independent samples[2,3], which is a reasonable assumption for this work. The comparison is based on transforming the grade correlation (r) for each instructor using the Fisher transformation.

$$F_1 = \frac{1}{2}\ln\left(\frac{r+1}{1-r}\right)$$

The grade correlation for all other instructions (excluding the instructor being evaluated) is also computed for the same sequence of courses.

$$F_2 = \frac{1}{2}\ln\left(\frac{r+1}{1-r}\right)$$

The standard error of the difference between the two transformed correlations is:

$$SE = sqrt\left(\frac{1}{n_1-3}+\frac{1}{n_2-3}\right)$$

Where n_1 is the number of students used to compute r_1 (or F_1) for the instructor and n_2 is the number of students used to compute r_2 (or F_2) for all other statics instructors. For example, the largest n_1=119 for instructor-8 and n_2=227 in this case. Instructor-9 had n_1 = 14 and n_2=332. In both, the n_1+n_2 = 346 which is the total number of students tracked from statics into dynamics. The transformed values are assumed to be normally distributed so the z-statistic is computed.

$$Z = \frac{F_1 - F_2}{SE}$$

If the magnitude of the Z value computed is greater than 1.96, then the instructor's grade correlation is significant at the 0.05 level (two tailed). Large positive values of Z indicate the instructor has a large correlation which indicates the instructor is more effective at preparing students for dynamics. Large negative values of Z are interpreted to indicate that the instructor is less effective.

Fig. 8 shows the Z value for the grade correlations going from statics into dynamics. There are no significant differences at the 0.05 level. The closest is instructor-6 which had a Z= 1.86. The statistics don't allow one to conclude why there are no significant differences. It appears natural to conclude that instructor-6 has a high DFW which means many students who fail to learn the material don't pass statics and don't proceed to the next class. This should be considered an important part of being a good instructor. If students don't master the concepts and material in statics, they shouldn't be passed through the course. This appears obvious, but one suspects that instructors with high GPAs and low DFW rates aren't doing a thorough job of qualifying students to progress through the program. For example, instructor-8 has some of the lowest DFW rates so that nearly everyone passes the course yet some students don't do well in the follow on course like dynamics. Yet the degree that this trend is detected is not a strong and remains below the 0.05 level.

Because statics is a foundational course for many other engineering courses, the same is repeated for the grade correlations into two other engineering courses: solid mechanics and thermodynamics. These are shown in Figs. 9 and 10 with minimal discussion because the discussion is similar to that already presented for dynamics.

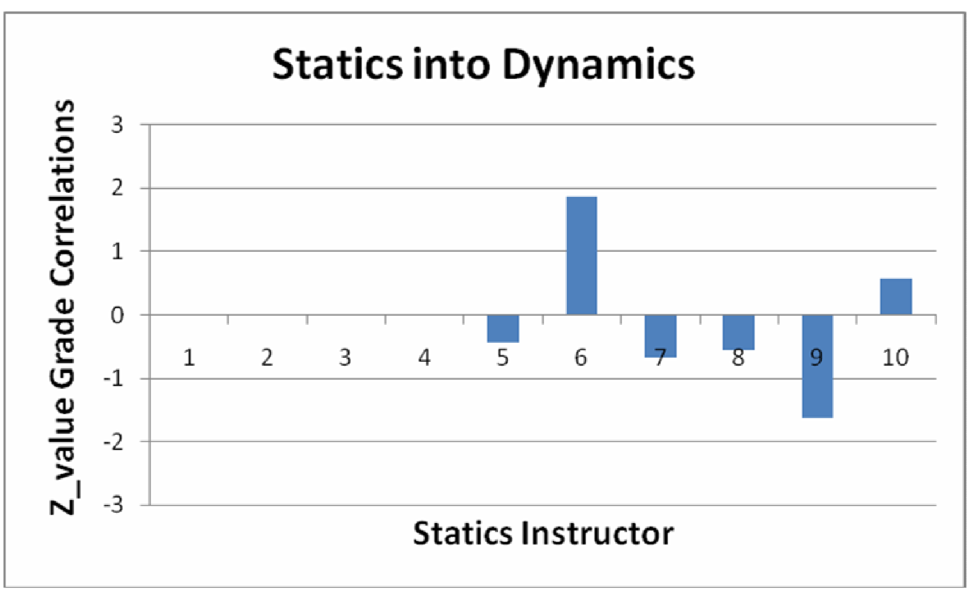

Fig 8. Significance of Grade Correlations from Statics into Dynamics.

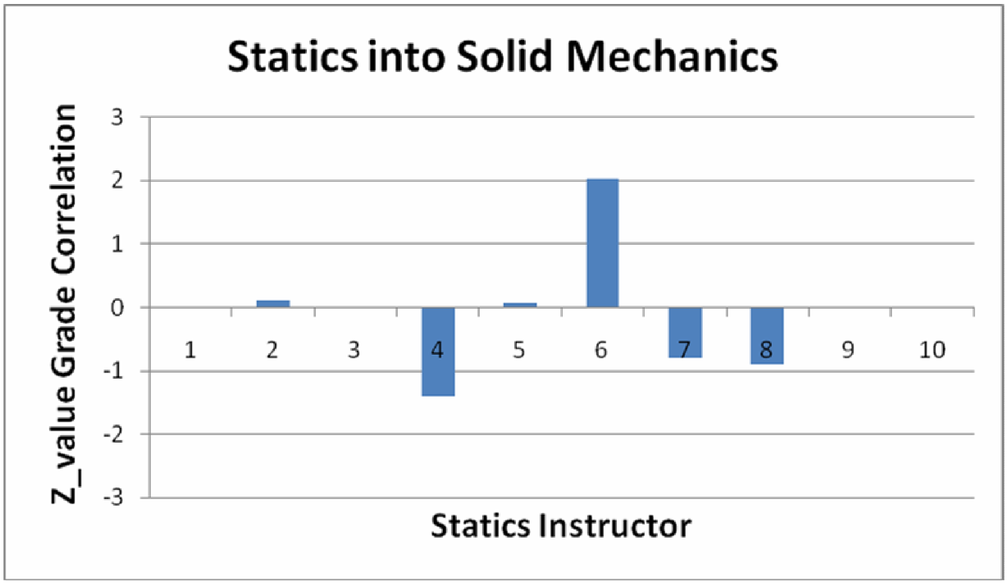

Fig 9. Significance of Grade Correlations from Statics into Solid Mechanics.

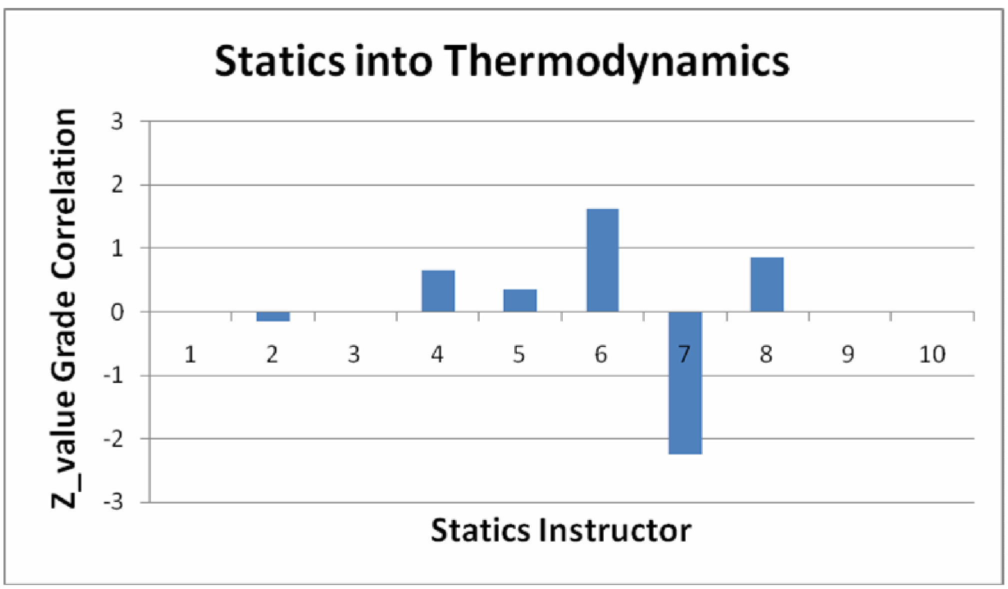

Fig 10. Significance of Grade Correlations from Statics into Thermodynamics.

As one looks at the entirety of the data for the statics-dynamics, statics-solids, and statics-thermodynamics sequence, it is difficult to identify a clearly significant difference between statics instructors. In two cases where was significance differences: instructor-6 for statics-solids (Z = +2.03) and instructor-7 for statics-thermo (Z= -2.25). Overall, the paucity of data limited the ability to make inference concerning grade correlations. Yet a stronger conclusion appears to be that the instructor has only a limited influence of student learning, both in the foundational course they may teach as well as subsequent learning achieved by the student in follow on courses.

Summary

The goal of this paper is to investigate quantitative grade-based measures to assess the effectiveness of different instructors that have taught statics, a foundational engineering class. Typically, many colleges use student surveys, grade distributions, and various sources of feedback to assess the effectiveness of individual instructors. These assessments are used in annual evaluations and tenure/promotion decisions. Given the importance of accurately assessing an instructor's effectiveness, it was decided to investigate grade-based correlations to discern if this is a meaningful metric.

Because the most important function of a foundational engineering course is to prepare students for subsequent higher-level courses, it is proposed that a quantitative metric of an instructor's effectiveness can be established on grades earned in both the foundational course and follow on courses. Grades are an indication of student learning and overall course performance. All students who pass the foundational class are considered prepared for subsequent courses. Students who earn high grades in the foundational class are considered better prepared for follow on courses, and are expected to achieve a higher level of accomplishment in subsequent courses. Likewise, those who are only adequately prepared and earn a "C" in a foundational class, they will tend to earn "C's in subsequent classes. Hence, this paper traces students who pass statics and then take three other courses: dynamics, solids, and thermodynamics.

A metric for the instructor's effectiveness is the correlation between grade earned in statics and that in the subsequent course. This is based on the understanding that grades are a measure of student's mastery of the course material.

In total, 860 students who attempted statics in 24 different classes having 10 unique instructors were studied. Of the students enrolled, the overall grade distribution is A-17%, B-21%, C-24%, D-10%, F-14%, W-14%, thus having an average GPA of 1.89 with a fail rate of 38%. The class to class statistics are compared to detect instructors with a significantly higher or lower GPA as well as higher or lower DFW rate. These statistics are useful to alert one to variations between classes and instructors.

Students who pass statics were tracked to determine their subsequent academic success in follow on courses. There are two drawbacks of this approach: a large number of students need to be tracked in order to develop meaningful statistics, and just as instructors vary for statics they also vary for subsequent courses. Acknowledging the reality of many other factors affecting grades, the grade correlation was computed and evaluated. It was determined that there is a significant range in the grade correlation, especially for instructors who don't teach the course often so there are only a few students to track. Statistically, the correlation has large uncertainty where there are a few students to track, yet one does detect a difference between instructors. In two cases, an instructor was identified as doing a better job (instructor-6 for statics into solid mechanics) while another instructor a distinctly poor job of preparing students (instructor-7 for statics into thermodynamics). In general the correlations are weak.

Overall, it is difficult to determine if one instructor is significantly better than another at helping student learn by only evaluating grades. Students tend to achieve consistent grades in their classes. Students who have the goal of achieving the highest possible grades often do so, while other students appear content to simply pass the class. Regardless of instructors, students tend to achieve grades commensurate with their effort. Hence, it is rare to see a negative grade correlation. Likewise, student performance is affected by many factors other than the instructor. In this study, the highest grade correlations are about 0.5 (instructor-6 averaged for dynamics, solids mechanics and thermodynamics). It appears that the best instructors using grade-based correlations can explain about 25% (0.5 squared) of the future grade success in follow on engineering courses.

Bibliography

1 C.M. Saviz, K.C. Schulz, W.J. King and R.H. Turpin, “Assessment Measures to Enhance Teaching and Curriculum Effectiveness” Proceedings of 31st ASEE/IEEE Frontiers in Education Conference, T1A-7, 2001.

2 National Academy of Engineering, “Developing Metrics for Assessing Engineering Instruction: What Gets Measured is What Gets Improved” Washington, DC: National Academies Press, 2009.

3 D.P. Hoyt, “Measurement of Instructional Effectiveness”, Research in Higher Education, Vol. 1, No. 4, 1973, pp. 367-378.

4 R.A. Berk, “Survey of 12 Strategies to Measure Teaching Effectiveness” International Journal of Teaching and Learning in Higher Education, Vol 17, No. 1, 2005, pp. 48-62.

5 F.A. Brandner, A Test of the Significance of the Difference of the Correlation Coefficients in Normal Bivariate Samples” Biometrika, Vol. 25, No. 1, May, 1933, pp. 102-109.

6 G.D. Garson, "Correlation", from Statnotes: Topics in Multivariate Analysis. Retrieved 12/20/2009 from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm. 3/24/08.