ASSESSMENT OF INSTRUCTOR EFFECTIVENESS BASED ON STUDENT PERFORMANCE IN SUBSEQUENT COURSES

Randall D. Manteufel, Amir Karimi
University of Texas at San Antonio
rmanteufel@utsa.edu

Abstract

This paper describes one way to assess an instructor's effectiveness in a foundational engineering course based on grades earned in the foundational course as well as grades earned in subsequent engineering courses. This approach is evaluated using five years of data covering about 850 engineering students who enrolled in the engineering mechanics: statics course. If they passed statics, students then proceed into dynamics, solid mechanics and thermodynamics courses. The statics course was taught by ten different instructors. A statistical analysis does identify significant differences in grade distributions. This qualitatively correlates with the reputations of instructors. Student grades in statics are correlated in subsequent courses and significant correlations exist. Results show that students who do well in one class tend to do well in other classes, yet grade correlation does vary based on the instructor for the foundational course. Some instructors have a surprisingly low grade correlation while the highest are about 0.6. This study does illustrate how statistical grade-based measures can be used in the important task of discerning the effectiveness of instruction in foundational courses.

1. 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 a variety of sources both inside and outside the university. Feedback from employers, national associations and leaders from the community/city/state/nation frequently provide high-level guidance to improve engineering programs. One consistent theme is that each course needs to prepare students with the right skills and capabilities to succeed in their future endeavors. It appears logical that a foundational engineering course should focus on student mastery of fundamental skills needed in engineering. There are a number of proposed mechanisms to measure student mastery of the subject [2] yet the authors found little use of grade correlations. A survey of strategies to measure teaching effectiveness [3] 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 subsequent student success in follow on courses doesn't appear to be as fully evaluated as it could be.

In the first few semesters of an engineering program, there is a widespread expectation that foundational engineering courses prepare student for other engineering courses. The idea of prerequisite courses is deeply rooted into programs where students start and proceed through a sequence of courses learning and gaining new capabilities. Foundational courses are often the focus of intense debates during curriculum revision. It is common to hear that students “aren’t coming into my class prepared.” Sometimes, student don’t gain needed perspectives or skills in foundational courses. This is part of the feedback process which plays an important role in course redesign. Yet how does one quantitatively assess the skills mastered by students in foundational courses?

If a student earns an “A” in a foundational engineering class, they should be well-prepared to succeed in subsequent courses. Likewise, if the grade earns a “C”, 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 occasion, one detects the effect of instructors who teach foundational courses by having students either ill-prepared or well-prepared for more advanced engineering courses.

There can be significant variation between instructors in both rigor and coverage of material in foundational courses. It is common to hear students say, “We didn’t cover that in the other class” or “we ran out of time” or “the instructor said it wasn’t important”. Feedback like this is often informal. It appears worthwhile to attempt to perform a quantitative statistical assessment of the instructor’s effectiveness in a class. The purpose of this paper is to assess the instructor’s effectiveness using grades earned in the foundational course and grades earned in subsequent engineering courses.

2. 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, selection for a scholarship, admission to a program major-sequence of courses, or admission to graduate schools. 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 it does indicate unreasonable variability.

Table 1 provides a summary of the final course grades for 24 classes of engineering statics taught by 10 different instructors at the University of Texas at San Antonio (UTSA) taught from the fall 2004 to summer 2009. Over the 5 year period, 860 students enrolled or attempted statics.

The data includes those students who withdrew from the class after census date which is typically in the second week of class. At UTSA 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 in the table.

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

Fig. 1 shows 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, etc. The mean GPA is calculated on a 4.0 scale with A-4, B-3, C-2, D-1, and F-0. UTSA has a generous withdrawal policy and many students withdraw because of poor academic performance after the first or second mid-term exam. So in the computation of the GPA, the withdrawals are treated as "F" grades, hence the designation of GPAW.

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 to be linked to some of the instructors, where some have a significant difference in both GPAW and DFW rate.

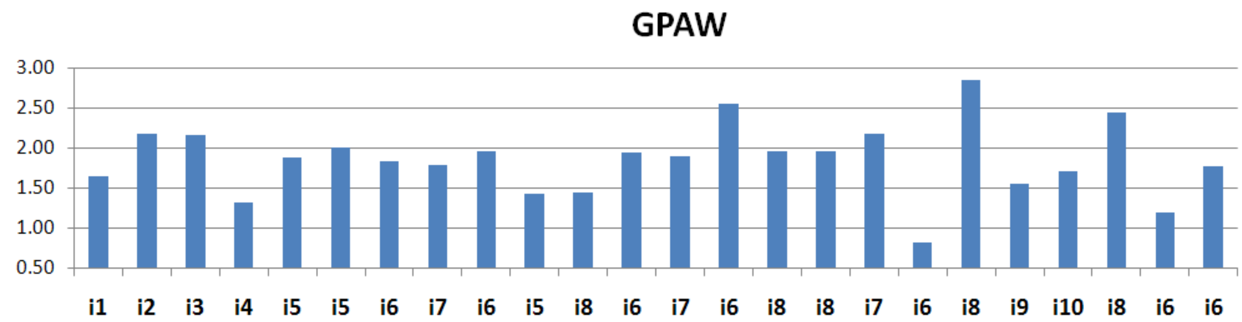

Fig. 1. Class Grade Point Average (GPAW) treating students who withdrawal from the course the same as earning "F" grades.

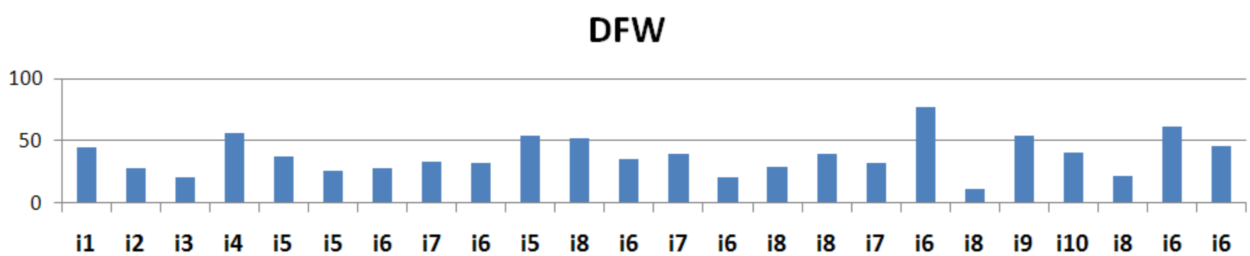

Fig. 2. Percentage of students not passing the class (DFW rate).

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. The goal of each student is to pass the class and an increasingly important metric for institutions are student retention rates and time to graduation (or graduation rate within 4 or 5 years). Those instructors with low GPA's and high DFW rates can easily be viewed as not being effective instructors. Likewise, those with high GPA and low DFW can be viewed as effective instructors and may be rewarded by annual assessment, raises, promotions, retention, etc.

A comparison is made of the GPAW for each class using the Z statistic. The mean GPA for each class compared with the 5-year course average (excluding that from the instructor teaching the class) is shown in Fig. 3 for each class. A value of Z greater than about 1.96 indicates the class has a significantly higher GPAW while those below -1.96 indicate significantly lower GPAW. The level of statistical significance is 0.05 or 5%. The largest deviation is for instructor-8 teaching the 19th class which had a high GPAW and instructor-6 teaching the 18th class with a low GPAW. Overall, three classes standout 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. Differences may be due to the arrival of a group of students who didn't master the material in a preceding course, typically math or physics. Or, the difference may be due to changes in the textbook or changes in the course (instructor trying new technologies or pedagogical strategies). Or, the difference may be attributed to 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). None of these other affects were investigated, although they are recognized as possibly being significant.

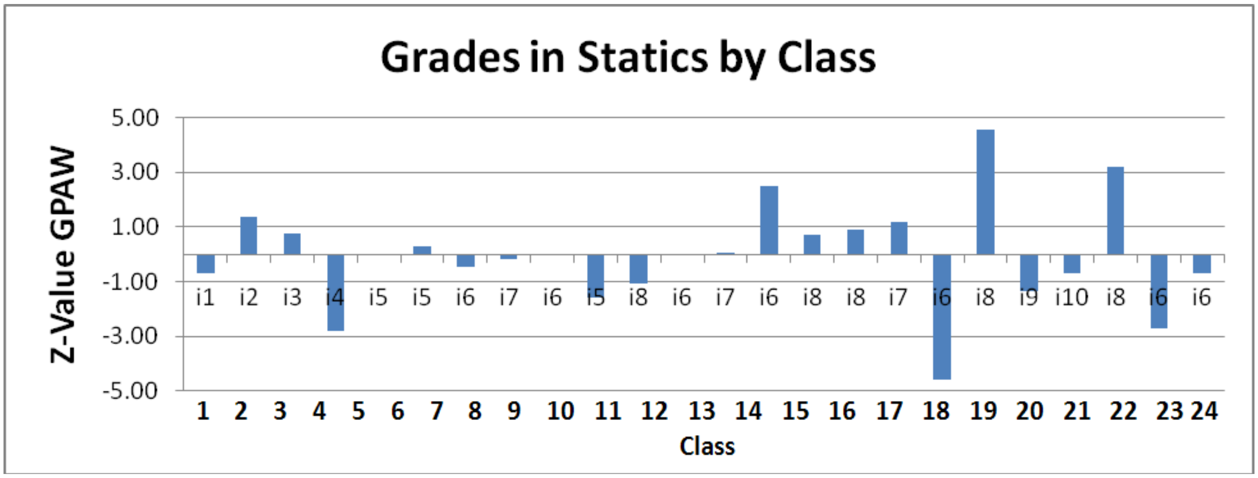

Fig. 3. Comparison of the mean GPAW for each class. Z values above 2 indicate the course has significant higher GPAW and values below -2 indicate significantly lower GPAW.

Fig. 4 shows the cumulative GPAW for each instructor compared with the average of all other statics instructors. When the two GPAWs are evaluated and compared, one can look for a statistical difference. From the data, instructor-8 is has significantly higher grades (Z = +3.85) while instructor-4 had significantly lower grades (Z = -2.99). All other instructors are within the 95% confidence (or not significant at the 0.05 level).

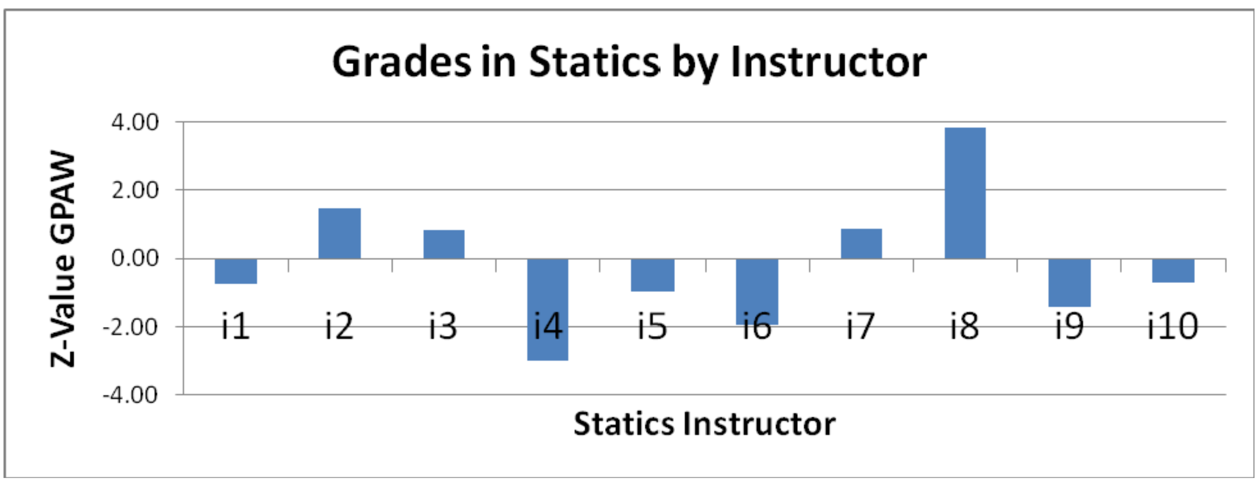

Fig. 4. Comparison of the mean GPAW on an instructor basis. Z values above 2 indicate the course has statistically significant higher mean GPAW. Z values below -2 indicate the class has a significantly lower GPAW.

The Z-statistic is often what students intuitively use when evaluating which instructor to take a course from. Students often describe instructors as being easy or hard. Many students want to enroll in a course taught by an instructor where the historical data shows that previous students succeeded in the course. One does conclude that there are significant variations between instructors and classes where the rate of success varies from one class to another as shown in Fig 4. Yet it is difficult to ascertain if student learning is significantly better from one instructor than another. Just looking at Fig. 4, instructor-8 appears to be the best instructor.

To evaluate learning, students who pass the foundational statics class are tracked into subsequent engineering classes. 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).

3. Subsequent Courses

The students who passed statics (EGR 2103) were tracked into a dynamics (EGR 2513), solid mechanics (ME 3813) and thermodynamics (ME 3293). An example of student performance is summarized in table 2 for instructor-8 from students taking dynamics.

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**
Sum	**38**	**42**	**39**	**119**

A total of 119 students passes statics with instructor-8 and attempted dynamics at UTSA. Of these, 39 earned A grades in statics, 42-B and 38-C. Of the 39 who earned A's, they went forward and 18 earned A's in dynamics, 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 subsequent classes. Likewise, those who earned C's in statics tend to earn C's in dynamics. This trend was observed in almost all of the data. In some regards, student grade performance is irrespective of the instructor where good students tend to thrive with the best instructor or survive the worst instructor.

Using the raw data, one computes the correlation coefficient, r, for the grades earned in both classes. For the data in table 2, one computes $r = 0.368$, showing that there is a positive correlation between statics and dynamics grades. Given the number of students tracked, one can compute an estimate of an upper and lower 95% confidence interval for r using a Z-Fisher transformation [4,5]. For the data shown, one concludes $0.20 < r < 0.51$, indicating that r is above zero with significant certainty (beyond the 0.05 confidence level).

A high value of r is desirable, because those students who earned A's in statics go forward and do well in dynamics. Likewise, those who earned C's in statics go forward and may pass dynamics, but do not do as well as those who earned A's in statics. This is based on a basic assumption that students who are well prepared for a course (earn high grades in prerequisite courses), do well in the follow on course. 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 used 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 in Table 3 (117) is nearly the same as those trackec in Table 2 (119). For instructor-6, the grade correlation is r=0.518 which is 0.15 greater than that for instructor-8. The computed 95% confidence interval range is: $0.37 < r < 0.64$. Because the ranges 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. Using the data in Table 4, for instructor-9, r=0.323 with a range of $0.03 < r < 0.56$. When one compares these three instructors, it appears the ranking of effectiveness goes from instructor -6 (best with r=0.518), instructor-8 (r=0.368) and then instructor-9 (r=0.323).

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	**19**
B	5	7	3	**26**
C	2	6	0	**33**
D	0	1	1	**12**
F	2	0	2	**12**
W	1	3	1	**15**

Fig. 5 is used to visually 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 a dot and the estimated 95% confidence interval as vertical lines. For instructor-9, there were only 14 students to track and this instructor has a negative correlation of -0.05 yet a large range. The diameter of the dot as well as the range of the r is controlled by the number of students. The larger diameter dot and smaller ranges in r are due to higher number of students tracked.

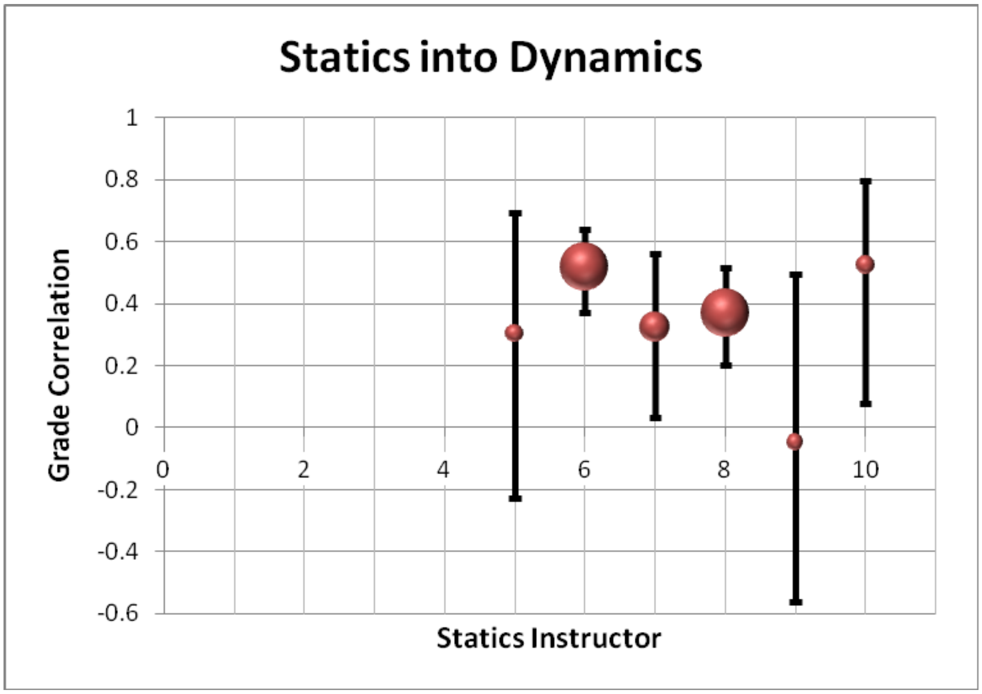

Fig 5. Grade correlation from statics into dynamics.

From Fig. 5, one discerns that some instructors have stronger grade correlations between the statics and dynamics courses. Strong correlations are expected when two things occur: (1) students who don't master the material in statics don't pass the course and (2) those who pass the course with good grades (A) achieved a greater mastery than those who pass with average grades (C). It appears that instructor-6 has a high DFW which means many students fail to learn the material and subsequently don't pass statics and don't proceed to the next class. One important part of being an instructor is qualifying students to progress through the program. Alternatively, 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 as strong as it could be because the uncertainty bands overlap.

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 in Figs. 6 and 7. Again, data is not plotted for those with less than 10 students to track from static or solids or statics to thermo.

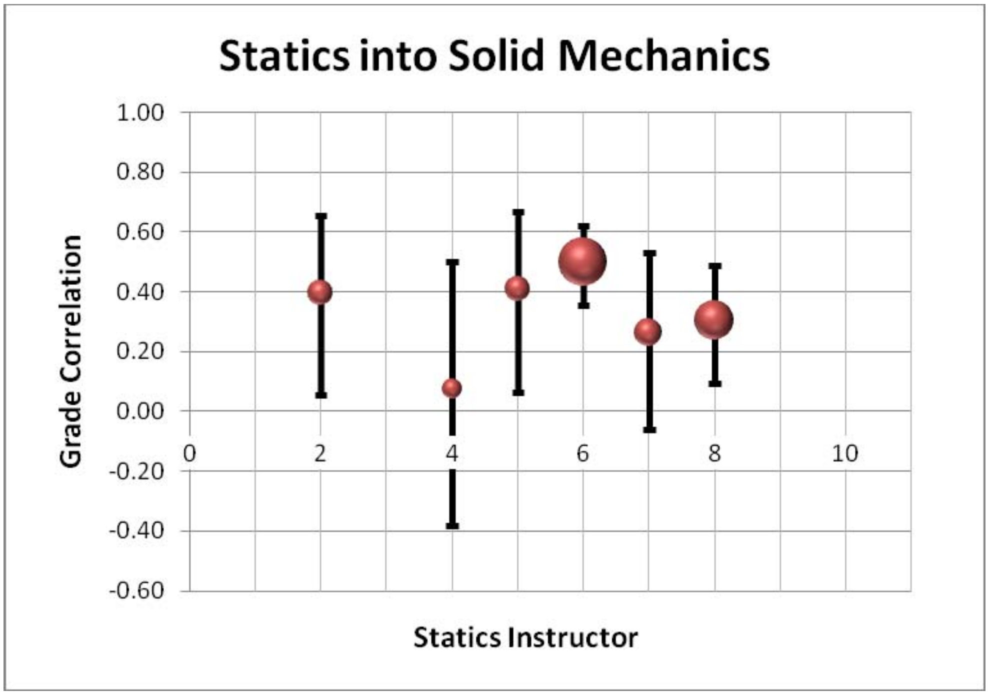

Fig 6. Grade correlation from statics into solids.

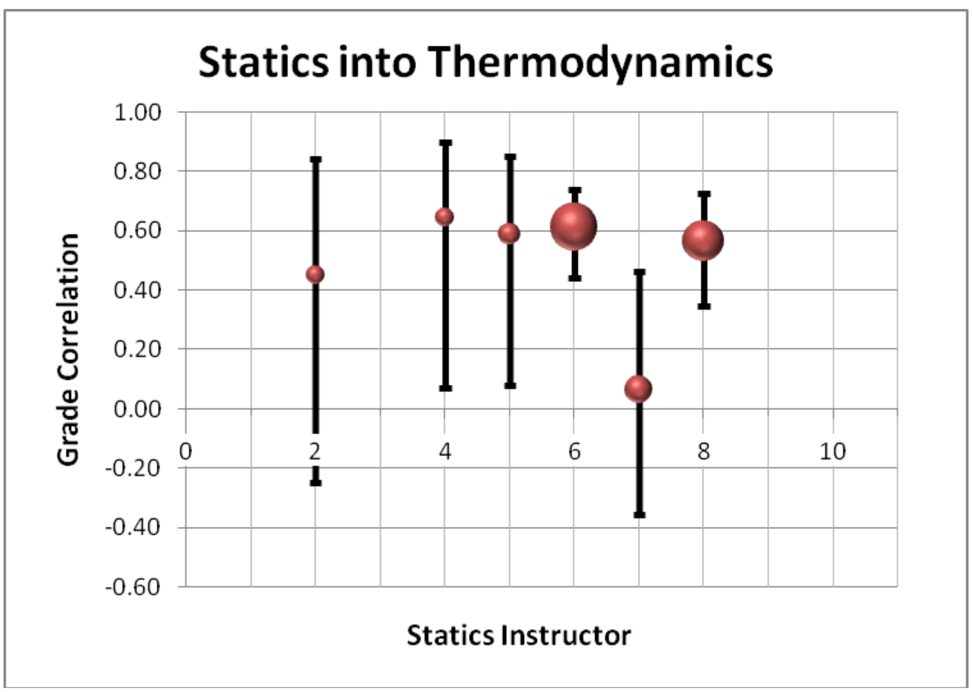

Fig 7. Grade correlation from statics into thermodynamics.

As one looks at the entirety of the data for the statics-dynamics, statics-solids, and statics-thermodynamics sequence, one sees that it is difficult to identify a clearly significant difference between statics instructors. If one compares the two instructors who have taught the largest number of students in statics (instructors -6 and -8) it appears that instructor-6 has larger positive grade correlations indicating a more effective instructor. Yet the use of statistics is such that other factors should be considered.

Some of the most logical appear the preparation of students in earlier foundational math and physics courses. It appears significant fluctuations in class GPAW and DFW rates may be due to the influence of earlier instructors because in many cases students progress through a program as cohorts. When talking with students, they often describe how their cohort group was together for important math and physics classes, and they often give frank assessments of the effectiveness of other instructors. Likewise, the data doesn't take into account the leniency or harshness of the instructor who taught the follow on engineering course (dynamics, solids or thermo). Just as there are significant class GPAW and DFW rate variations in statics, the same is true for dynamics, solids and thermodynamics. These factors, as well as many other possibilities, are not taken into account in this work.

4. Summary

The goal of this paper was to investigate a grade-based measure to assess the effectiveness of different instructors in the foundational engineering class of statics. Typically, student class surveys and qualitative feedback are used to assess the effectiveness of individual instructors. It is proposed that a good indicator of an instructor's effectiveness can be based on the correlation of grades earned in the foundational class and grades earned in the subsequent engineering courses. If the students are well prepared in the foundational class, they have a high probability of earning good grades in subsequent courses. Likewise, if students are only adequately prepared and earn a "C" in a foundational class, they will tend to earn lower grades in subsequent classes.

This paper traces students who pass statics and progress into three other courses: dynamics, solids, and thermodynamics. The proposed indicator is the correlation between grade earned in statics and that in the subsequent course. This appears to be a good indication of a student's mastery of material in statics which prepares them for success in the next class.

This paper tracks over 860 students who attempted statics in 24 different classes having 10 different instructors. Of the students enrolled, the overall statics 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 identify significant variations, yet subsequent academic success of students is a good measure of an instructor's effectiveness.

It is noted that there are two major drawback of this approach: a large number of students needs to be tracked in order to develop meaningful statistics, and instructors vary for courses both before and after the statics course. Regardless, the grade correlation, r, was computed and data shows a significant range of “r” especially for instructors who don’t teach the course often so there are only a few students to track. A significant r-value require large numbers of students. With limited data, it does appear that one can detect a difference between instructors. The instructor having the highest GPA and lowest DFW is not the instructor with the highest grade correlations into subsequent classes.

One thing this study reinforced is that students tend to exhibit the same level of academic achievement in their classes. Some students have the goal of achieving the highest possible grades and do so in their classes, while some are content to simply pass the class. Regardless of instructor, students tend to achieve grades commensurate with their effort. Hence, it is difficult to have a negative grade correlation. Likewise, student performance in a single class is affected by other factors, not just rigor of preparation in prerequisite class. Overall, it is difficult to have a high positive grade correlation, and the highest observed in this work is about 0.6. This can be interpreted that preparation and strong academic performance in a prerequisite class explains up to about 40% of the subsequent success in a follow on class, while 60% or more is attributed to other factors.

References

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

[3] 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.

[4] 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

[5] 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.