AC 2011-66: INVESTIGATING THE VALIDITY OF STUDENTS' SELF-ASSESSMENTS OF THEIR ABILITY IN STATICS

Jeffrey L. Newcomer, Western Washington University

Jeffrey L. Newcomer is a Professor of Manufacturing Engineering Technology at Western Washington University. He received B.S. (1988) and M.Eng. (1989) degreesin Aeronautical Engineering, a M.S. in Science and Technology Studies (1993), and a Ph.D. in Mechanical Engineering (1994) from Rensselaer Polytechnic Institute. He is engaged in research to improve instruction and assessment in engineering, with an emphasis on engineering fundamentals such as mechanics

INVESTIGATING THE VALIDITY OF STUDENTS' SELF-ASSESSMENTS OF THEIR ABILITY IN STATICS

Abstract

In this paper students' self-assessment of three skills in Statics—drawing free body diagrams, writing equilibrium equations, and solving equilibrium equations—were compared to their performance in the same three areas across multiple problems on a final exam. Furthermore, additional cognitive data such as overall grade point average and grades in pre-requisite courses were also compared to performance in Statics. While there was correlation between students' self-assessment of their ability to draw free body diagrams and performance on the final exam in all three areas, the strongest correlations were with overall grade point average. There were also some correlations with grades in the pre-requisite Physics I class and whether that class was algebra or calculus based. The effects associated with these factors were small, and none of the models accounted for much more than a third of the variation in the data. As such, these data indicate that students' self-assessment scores should not be used for assessment in place of direct measures, but might still serve as supplemental information.

Introduction

Self-reported surveys are one of the simplest forms of assessment data to collect, but do self-reports have any correlation with actual student ability? If self-reports could be shown to correlate with performance on exams, then they would be valid substitutes for direct outcomes assessment measures, but without validation this form of indirect assessment should only be used to supplement more rigorous, direct assessment of student work. Previous studies in engineering education have examined controlled situations where one problem was directly compared to one survey question, and have shown limited correlation. This paper will present the results from a study that asked students to self-assess their fundamental skills in a Statics class compared to performance across multiple problems on the final exam. Students' self-assessments from a Statics class were compared to their performance on the final exam in three areas: the ability to draw free body diagrams (FBDs), the ability to correctly write equilibrium equations, and the ability to correctly solve these equilibrium equations. Due to the nature of Statics, students were required to complete multiple problems in each of the areas of interest, so student performance is based on aggregate data. In addition to their exam performance, additional cognitive data such as grade point average (GPA) and performance in pre-requisite courses were considered to determine if any are a more effective predictor of student performance than self-assessments.

This is not the first attempt to determine if self-assessments correlate with performance in engineering, although there are few examples. A previous study in industrial engineering by Sarin and Headley directly compared one exam problem to one self-assessment survey question.[1] This work showed some correlation between student self-assessments and objective performance, but small effects, implying that self-assessments are neither good predictors of performance nor justified for use in summative evaluation. Another student in a first-year course by Collura and Daniels looked at aggregate skills in spreadsheet usage and basic programming by comparing students' self-assessments to instructor assessments for the same areas.[2] The findings of this study were consistent with the previous one: there are some correlations, but the effects are small.

There is one important difference between the two studies. Despite finding $R^2 < 0.3$ in all cases, Collura and Daniels conclude that aggregate (not individual) self-assessments can be used for assessment purposes, but they did not study if any other factors correlate. With that difference noted, these findings are consistent with work in the social sciences that found that correlations are moderate when objective measures are compared to self-assessments for individuals.[3] Of course, there are many factors that affect students' ability to accurately assess their own performance and ability, including the experience of the students at assessing their own performance or ability, the level of the students, the complexity of the material, and even the measurement scale.[2,4] Preliminary work on this project, completed with Yokomoto, examined students' ability to assess their performance in Statics and Engineering Ethics.[5] In the case of the preliminary study, however, students were asked to rate their performance in Statics prior to taking the final exam and no other factors were considered. This study indicated that there were mild correlations between performance and self-assessment (enough to warrant further study).

The present study looks to see if comparing students' self-assessments to performance across multiple problems shows any more correlation than was found in the one question to one problem work of Sarin and Headley.[1] The analysis is based upon data collected in the Applied Engineering Statics course at Western Washington University (WWU) during the Winter and Spring quarters in 2007. The analysis includes multiple regression variables such as overall GPA and grades in Physics I and Calculus I.

Methodology

The Applied Engineering Statics course at WWU, which has a pre-requisite of one quarter of physics (mechanics) and one quarter of calculus, is organized around five topics: free body diagrams, equilibrium, equivalence, separation of rigid bodies, and friction, without differentiation between two vs. three dimensional cases, concurrent vs. non-concurrent force systems, and single bodies vs. frames and trusses.[6] These situations are all addressed in the course, but not in the order of traditional textbooks. Students also complete weekly Warm-up exercises for the first eight weeks of the quarter.[7] Otherwise the course is a standard lecture-based course with homework, midterms, projects, a final exam, and a limited number of think-pair-share exercises.

The goal of Statics is to determine all of the loads on, and sometimes in, a rigid body system that result in the bodies in the system having no acceleration; the goal of the structure of the Statics course at WWU is to get students to consistently apply a methodical approach to solving Statics problems. The general approach to solving problems is to first draw a free body diagram (FBD), to then use that FBD to write equilibrium equations, and then to solve said equations for the unknown load or loads. Due to the nature of the course, problems in Statics have a tendency to be very similar, even when the systems under study appear to be very different. This characteristic makes Statics a good candidate in which to test student self-assessments against actual performance on exams, for students must complete the same steps on almost every problem. Thus comparisons can be based upon aggregate data, rather than only one problem.

Preliminary work on this project was completed using data from the Winter 2006 section of the Applied Engineering Statics course at WWU.[6] At the end of the course, but before the final exam, students were asked to voluntarily complete a self-assessment survey. Among the questions

on the survey, students were asked to rate how well the course has helped them learn to: (1) draw a correct FBD, (2) write equilibrium equations, and (3) solve them. In Winter 2006, 66 of 69 students in the course completed the self-assessment survey.

Students self-assessment of their ability in these three areas was then compared to their performance on the final exam. The final exam included five problems that required students to draw FBDs, five problems that required students to write equilibrium equations, and three that required students to solve their equilibrium equations. Students were given separate scores for each part of each problem. The topic scores in each area were created by combining the individual topic scores from each of the problems; these were then recorded as percent of maximum possible score. The self-assessment scores were regressed against the overall percentage score in each topic area. The FBD score was found to be significant ($R = 0.417$, $p < 0.002$), while writing and solving equilibrium equations were not. Even in the case of FBDs, the linear model accounted for a small fraction, only about 17%, of the variation in student performance on the final exam. In practical terms, the regression model predicted that a student who rated his or her ability to develop FBDs as a 5 would score approximately 11% higher in this area than a student who rated her or his ability in this area as a 3.

This analysis raised several questions, for there are several reasons that the correlation coefficients may be low, including students' lack of experience at rating themselves. One question was whether or not asking students to assess their abilities after completing the problems would make a difference. Another question was whether or not the some other piece of information might serve as an effective predictor of students' ability to draw FBDs, write equilibrium equations, and solve them.

As a result, the basic structure of the analysis was repeated with the Winter 2007 and Spring 2007 Applied Engineering Statics classes at WWU, but with two changes. First, the self-assessment questions were moved to the end of the final exam, so that students assessed their abilities after completing the problems. Second, extra data was collected to be used in the regression. Students' GPAs at the beginning of the class were recorded. The type of Physics I course students completed—algebra based, calculus based, or at a community college (CC)—was recorded. Whether or not students had completed a second quarter of calculus before or concurrently with Statics was recorded. Finally, the specific grades in Physics I and Calculus I were recorded for students who took these courses at WWU (the information was not available for transfer students).

As with the previous study, students were asked to complete five problems on the final exam, all of which required FBDs and equilibrium equations, and three of which required solving the equilibrium equations. For continuity, these two sections were given the same problems, but the exams were carefully controlled, so students in the second class did not have access to exams from the first class.

Linear regressions were conducted to see if any of the factors had statistically significant correlations with the grades on the final exam in the three areas. The plan was to use the model shown below:

$$y = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \varepsilon$$

where the terms are:

μ is the grand mean
β_i are the regression coefficients
x_1 is the student GPA (normalized 0 to 1)
x_2 is WWU physics (0) or CC physics (1)
x_3 is completed Calculus II (0 or 1)
x_4 is WWU physics type (algebra-based = 0, calculus-based = 1)
x_5 is the grade in Physics I (normalized 0 to 1)
x_6 is the grade in Calculus I (normalized 0 to 1)
x_7 is FBD self-assessment score (normalized 0 to 1)
x_8 is the writing equilibrium equations self-assessment score (normalized 0 to 1)
x_9 is the solving equilibrium equations self-assessment score (normalized 0 to 1)
ε is random error

However, because grades in Calculus I and Physics I and information on the type of Physics I (algebra or calculus-based) were not available for the students who had gone to community college, and the number of students who had gone to community college in the group was small ($n = 6$), the community college variable x_2 was dropped from the regression. The regressions were therefore conducted on the entire group ($n = 67$) using the variables x_1, x_3, x_7, x_8, and x_9, and on the group who had taken Calculus I and Physics I at WWU ($n = 61$) using all variables except x_2. All three self-assessment scores were included in all of the regressions to see if there was any pattern to students' self-assessments. If self-assessments for one skill correlate with other skills then it is an indicator that students might not be truly considering the task involved when the rate their abilities. The linear regressions were conducted using Excel.

Findings and Implications

As with previous studies,[1-3,5] there were some statistically significant correlations, but the effects were small. For the entire group of students, GPA and the self-assessment score for drawing FBDs correlated with all three areas. The correlations with GPA and performance on the final exam were the strongest ($p < 0.001$ in all three cases). The correlation with FBD self-assessment and performance varied, with the $p < 0.001$ for FBD performance, $p < 0.005$ for writing equilibrium equations performance, and $p < 0.10$ for solving equilibrium equations performance. In none of these cases, however, was the effect large. For FBD performance adj. $R^2 < 0.35$, and it was smaller for the other two regressions. Figures 1 and 2 show the distributions for GPA and FBD self-assessment for FBD performance respectively. It is easy to see from these plots why the correlation coefficient is relatively low.

For the group of native students there were three other significant, but still small, effects. Taking calculus-based instead of algebra-based Physics I correlated with performance in writing equilibrium equations ($p < 0.02$) and solving equilibrium equations ($p < 0.05$), and the grade in Physics I correlated with solving equilibrium equations ($p < 0.10$).

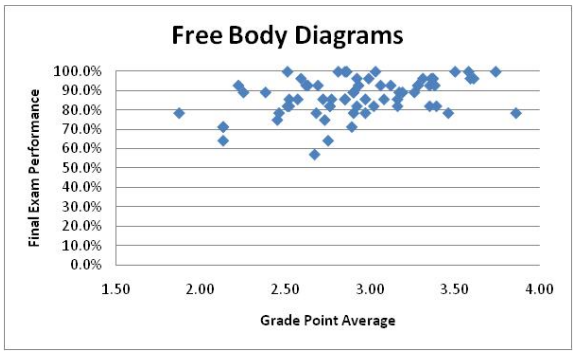

Figure 1:Grade Point Average v. FBD Final Exam Performance

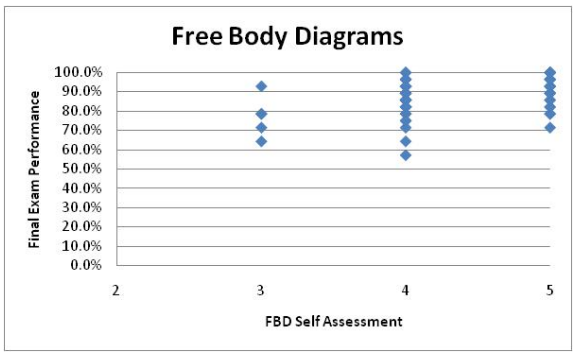

Figure 2: FBD Self-Assessment v. FBD Final Exam Performance

There are several interesting implications in these findings. First, these results agree with earlier findings that there are some correlations between students' self-assessments and their performance on exams, but the effects are not large, and here the correlations were limited to self-assessments for ability to draw FBDs, even for writing and solving equilibrium equations. Furthermore, even in the case of the model with the best fit to the data, drawing FBDs for the overall group, the model only accounted for less than 35% of the variation in the data, and GPA accounts for the largest share. Moreover, the FBD self assessment correlation is almost identical to what it was found to be in the preliminary study. This result implies that whether students self assess before or after the final exam is irrelevant. Second, GPA is a better predictor of exam performance in Statics than self-assessments. While this is not surprising in that past academic performance is often a predictor of future academic performance, one would hope that students' self-assessments on specific skills and topics would be more discerning than cumulative GPA. Third, that taking calculus-based instead of algebra-based Physics I and that the grade in the Physics I is a predictor of some performance in Statics, also is not a complete surprise given that an understanding of forces, vectors, and Newton's Laws is necessary to grasp the conceptual aspects of Statics. While a conceptual grasp is not strictly necessary to complete the methodology of Statics, Steif, Dollar, and Dantlzer showed that performance on conceptual and traditional Statics problems does correlate well.[8] Of course the correlation with performance in Statics and having taken calculus-based Physics I could also be a reflection of student preparation and self-confidence, but that is not discernable from these data.

Finally, possibly the most interesting finding here is that only the drawing FBD self-assessment correlates to performance, but it correlates to performance in all three of the areas examined: drawing FBDs, and writing and solving equilibrium equations. While most students gave themselves the same ratings in all three areas, students in this study who did change their rating scores in different areas were three times more likely to assess their ability to write equilibrium equations as lower than their ability to draw FBDs, so students may believe that the equations are the difficult part, but for many of them the equations follow from the FBDs more accurately and consistently than they give themselves credit for doing.

Conclusion

While there are several interesting things about this study, the bottom line is that students' self-assessments, even when performed after an exam and across multiple problems, are not good

predictors of exam performance. In fact, according to the analysis here, GPA shows a stronger correlation with performance in Statics than students' self-assessments. As such, students' self-assessments should not be used in program or course assessment, at least in Statics, as anything other than a supplement to direct assessment measures. It is possible that students' self-assessments do reflect something of value and interest, such as confidence, that could be useful in the improvement of instruction and course material, but they do not show a strong or consistent correlation with performance on actual problems in Statics.

Bibliography

1. Sarin, S., and Headley, D., "Validity of Student Self-Assessments," *Proc. of the ASEE Annual Conference and Exposition*, Montreal, Quebec, Canada, June, 2002, http://www.asee.org/acPapers/ 2002-831_Final.pdf.

2. Collura, M. and Daniels, S., "How Accurate is Students' Self-Assessment of Computer Skills?," *Proc. of the ASEE Annual Conference and Exposition*, Pittsgurgh, PA, June 2008, http://soa.asee.org/paper/conference/paper-view.cfm?id=9120

3. Volkwein, J. F., "On the Correspondence Between Objective and Self-Reported Measures of Student Learning Outcomes," 2005, http://www.ed.psu/chse/abet/pdf/Self_Reported.pdf.

4. Miertschin, S., Goodson, C., Faulkenberry, L., and Stewart, B., "Student Self-Assessment: Can It Be Used to Improve Instruction?," *Proc. of the ASEE Annual Conference and Exposition*, Chicago, IL, Jun., 2006, http://www.asee.org/acPapers/code/getPaper.cfm?paperID=10836&pdf=2006Full2235.pdf

5. Newcomer, J.L., and Yokomoto, C.F., 2007, "Do Students Know What They Know?: Comparing Self-Assessments to Exam Performance," *Proceedings of the 37th Annual ASEE/IEE Frontiers in Education Conference*, Milwaukee, WI, October, S4S-7-8.

6. Newcomer, J.L., 2006, "Many Problems, One Solution Method: Teaching Statics without 'Special Cases'," *Proceedings of the 36th Annual ASEE/IEE Frontiers in Education Conference*, San Diego, CA, October, S2D-7-12.

7. Patterson, E., and G. Novak, 2006, *Just-In-Time Teaching*, http://webphysics.iupui.edu/jitt/jitt.html.

8. Steif, P.S., Dollar, A., and Dantzler, J.A., 2005, "Results from a Statics Concept Inventory and their Relationship to other Measures of Performance in Statics," *Proceedings of the 35th Annual ASEE/IEE Frontiers in Education Conference*, Indianapolis, IN, October, T3C-5-10.