Assessment In Engineering Courses: An Exploratory Approach

Farnaz Ganjeizadeh[1] and Julia A. Norton[2]
[1] Department of Engineering [2]Department of Statistics and Biostatistics,
California State University, East Bay, Hayward, CA 94542

Abstract
Educators are faced with constructing and evaluating student learning in engineering courses by asking students how confident they are in the domain knowledge that they have obtained. For self-assessment of individual exam questions, students assign values to each question indicating their degree of correctness. This study is based on the responses from engineering students at three levels: Introductory, undergraduate junior or senior level, and a graduate level course. All courses were taught as lectures which included analytic problem solving and case studies; hand-on Labs and an application paper solving engineering problems using qualitative and quantitative methods presented as a team. Students were asked to respond on a scale of 1 to 10 (1 not correct and 10 certain correct) to selected problems in the exams. The students' self- assessment scores were compared against their actual performance graded by the instructor. Additionally, a comparative analysis was conducted to measure the predictive correlation among the student self-assessment and actual grading based on the above three categories.
Key Words: Engineering, Educational assessment; Learning; Teaching

1. Introduction

Trends in assessment have turned lately to asking students how well they learned the material in a course. As with most publicly funded state colleges and universities, there is a continued effort to do more with less funding, to combine, reduce or eliminate small programs, and to demand that all programs demonstrate that they provide something for the public good. Ganjeizadeh and Norton have collaborated on a number of issues in educational assessment over the past ten years including learning in introductory courses. A Brazilian study of post graduate engineering students by Luis Roberto C. Ribeiro Maria Da Graça N. Mizukami (2005), found that a problem-based learning (PBL) implementation in the curriculum used was very satisfactory and may have promoted the acquisition of knowledge as well as the development of some desirable skills and attitudes, such as teamwork and communication skills and respect for divergent ideas.

Summaries of our previous studies appear in several *Proceedings of the American Statistical Association Section on Statistical Education* as our data increased and the questions became more varied (Lovell, Dietz, Eudey and Norton with others between 2000 and 2006). These papers consider assessments in introductory courses and our statistics degree programs. The ideas discussed are consistent with the fundamental learning goals outlined in Garfield and Ben-Zvi (2007) and Norton and Lovell (1981). In 2006-2007 Norton served as Interim Director of Institutional Research, writing a broader survey of the assessment at the University (Norton 2007). Returning to teaching in 2007-2008 Norton collaborated with many faculty from all areas of the university in supporting assessment attempts (Norton, Zhou, and Ganjeizadeh 2008 and Eudey, Anand, Norton and Coulman 2009).

Seeking less controversial means of evaluation among university faculty and ones perhaps less intrusive into the classroom, some suggest asking students directly about their learning experience in terms of what they had learned using a consumer model of assessment. Since our previous research concluded that common finals written by committee or by outside evaluators gave results that satisfied us, we wondered how a version of these new methods might work. We are not in favour of using student evaluations as assessments, however it might generate some self-evaluation that might be useful for the students as a lifelong technique. Therefore, in 2011 we decided to associate the question of learning with the twenty questions already being used in the introductory statistics final. Achieving some indication of student ability to determine correctly how difficult a particular question was for an individual to answer, we decided to extend our investigation to introductory psychology courses as well in 2012 and 2013. The results were mixed. In 2014 we included similar results from a number of engineering courses to see how the results varied.

2. Relationship Between Faculty Evaluation And Student Evaluation

We wondered whether the quality of work as assessed by the faculty member is related to a similar assessment on a scale of 0 to 10 by the students. Similar studies related correct response to level of certainly about a student’s answers (Dietz, Lovell, Norton and Norton 2011. 2012, 2013).

In five Engineering courses all required, Ganjeizadeh included two questions designed to assess the learning of engineering students on two of the ABET accreditation measures. Specifically, whether students could adequately solve an engineering problem mathematically and whether students could communicate engineering ideas. Students were assigned the task of solving all problems on an exam with these two problems included, and for these two problems were asked on a scale of 0 to 10 how well they thought they had accomplished the goal of the problem. An engineering faculty member evaluated the actual response to those questions for some varying number of pre-assigned points, depending on the course and exam. The points were rescaled from 1 to 10 so that means would compare equivalent scales.

Our subjects are 265 students in six engineering classes at California State University East Bay, all taught by the same instructor and given credit for carefully completing the self-assessment as they completed one exam. We use the measures average and correlation between faculty assessment and the perceived quality as reported by each student. To ensure some degree of similarity, the exam scores are compared across the level of the courses. Looking at studentized residuals and deleted residuals for the variables shows markedly uninteresting patterns.

Table 1 below indicates a map and the level of difficulty required in the Engineering Curriculum by these two ABET assessments. Assessment questions were matched to appropriate difficulty level for each course. While evaluating student achievement is better accomplished by one or more outside evaluators, we were not in a position to do this. Courses in bold were included in the study as well as one master’s level course title Engineering Management, ENG 6200. Questions were keyed to the desired level of proficiency in the course.

Table 1: Mapping of Outcomes for Industrial Engineering by Course
I=Introduction, P=Practice, M=Mastering

	(a) Apply math/science engineering knowledge knowledge	**(g) Communicate**
General Education		I
Mathematics	I	
Natural Sciences	I	
Engineering Core		
ENGR 1011 Engr. an Intro.		
ENGR 2010 Electric Circuits	P	
ENGR 2060 Material Sceince	P	
ENGR 3101 Statics and Dyns.	P	
Program Required Courses		
CS 1160 Intro. to CS & Prog.	I	
ECON 2301 Microeconomics	I	
ENGR 1420 Engr. Graphics	I	
ENGR 2070 Fund. Of Manuf.	I	
ENGR 3020 Wk. Dsgn., Meas.	P	
ENGR 3140 Engr. Economy	P	P
ENGR 3190 Human Factors	P	
ENGR 3601 Stat. for CS/Engr. I	P	
ENGR 3602 Stat. for CS/Engr. II	P	
ENGR 3841 Operations Res.	P	
ENGR 4100 Prod. Planning	P	
ENGR 4200 Simulations	M	
ENGR 4280 Design and Mgmt. Human Work Systems		
ENGR 4300 Quality Engr.	M	
ENGR 4350 Reliability Engr.	M	P
ENGR 4400 Manufacturing Systems Engineering		P
ENGR 4430 Facilities Plan.	M	P
ENGR 4440 CIM	M	
ENGR 4610, 4620, Senior Design I, II		M

Tables 2 and 3 below show the correlations for the two separate groups of graduate and undergraduate students, for the two ABET questions. The graduate students are more tentative in assessing their engineering prowess than the undergraduates, but the direction is positive and moderate in all cases between the student evaluation and the faculty evaluation, with the graduate students' assessment of their ability to solve the engineering problem correlating on 0.264 with the faculty member's assessment. All the remaining correlations are on the order of 0.5. Since many of the graduate students come from a diversified background such as Construction Management, Information Technology

(IT),Business, etc., rather than a purely scientific one, one explanation we will explore in the future is level of the application of the mathematical modeling concepts or something similar.

Table 2: Correlation Matrix Graduate Student Self-Assessment with Faculty Evaluation:

		Student Rating Problem Solving	Student Rating Communication
Student Rating Problem Solving	Pearson Correlation	1	.350**
	Sig. (2-tailed)		.000
	N	122	122
Instructor Rating Problem Solving	Pearson Correlation	**.291****	.047
	Sig. (2-tailed)	.001	.605
	N	122	122
Student Rating Communicate	Pearson Correlation	.350**	1
	Sig. (2-tailed)	.000	
	N	122	122
Instructor Rating Communicate	Pearson Correlation	.174	**.567****
	Sig. (2-tailed)	.055	.000
	N	122	122

Values in bold are the ones of interest. The other correlation values support the notion that students are trying to give a true assessment of their ability rather than randomly marking an answer. Instructor rating of opposite question is low further verifying consistent results.

3. Mean differences Between Faculty Evaluation And Student Evaluation

In order to understand some of the underlying issues, we used a repeated measures, within subject analysis with student level as a between variable. Our goal in this analysis is to explore effects, not to test them. Table 4 below includes the main multivariate model for the four measurements on each subject made by the two raters, instructor and student, for each of the two ABET skills or questions, one on communication of ideas in engineering and the other the ability to solve a mathematical engineering problem.

SPSS was used to analyze the data and all of the usual test results are equivalent because of the two valued nature of the model under consideration (question: ability to solve and communication; rater: instructor and student). Although statistical tests are not really appropriate here, we can look at the information with an eye to thinking about the possibilities in an exploratory manner. No experimentation or long term study has been carried out.

We have asked students to rate how well they think that they did on a particular question and then asked an instructor to rate how well the student actually did on the same problem for two problems: one designed to assess the ability to solve an industrial engineering problem and the second to communicate issues in industrial engineering. Students are undergraduate or graduate level (masters) and course level is beginning or intermediate. Course level is not considered in this initial study.

Table 3: Correlation Matrix Undergraduate Engineering Student Self-Assessment with Faculty Evaluation.

		Student Rating Problem Solving	Student Rating Communicate
Student Rating` Problem Solving	Pearson Correlation	1	.370**
	Sig. (2-tailed)		.000
	N	123	121
Instructor Rating Problem Solving	Pearson Correlation	**.564****	.124
	Sig. (2-tailed)	.000	.174
	N	123	121
Student Rating Communicate	Pearson Correlation	.370**	1
	Sig. (2-tailed)	.000	
	N	121	121
Instructor Rating Communicate	Pearson Correlation	.172	**.546****
	Sig. (2-tailed)	.058	.000
	N	122	120

Values in bold are the ones of interest. The other correlation values support the notion that students are trying to give a true assessment of their ability rather than randomly marking an answer. Instructor rating of opposite question is low further verifying consistent results.

The biggest effect is that students consistently rate themselves more highly on average than the instructor rates the student ability to solve problems or communicate. There is some small rater interaction with student level. Graduate student ratings are more like instructor ratings than undergraduate student ratings are like instructor ratings. Seemingly this is consistent with what one would expect from a graduate student at the master's level, to be more self-aware. Figures 1 and 2 below show the average fits for solving engineering problems (Figure 1) and communication (Figure 2).

Table 4: Multivariate Tests Exploratory Use only

Effect		Value	F	Hypothesis df	Error df	Sig.
question	Pillai's Trace	.013	3.082[a]	1	240	.080
question * grad undergrad	Pillai's Trace	.020	4.875[a]	1	240	.028
	Wilks' Lambda	.980	4.875[a]	1	240	.028
	Hotelling's Trace	.020	4.875[a]	1	240	.028
	Roy's Largest Root	.020	4.875[a]	1	240	.028
rater	Pillai's Trace	.097	25.648[a]	1	240	.000
rater * grad undergrad	Pillai's Trace	.014	3.508[a]	1	240	.062
	Wilks' Lambda	.986	3.508[a]	1	240	.062
question * rater	Pillai's Trace	.002	.552[a]	1	240	.458
	Wilks' Lambda	.998	.552[a]	1	240	.458
question * rater * grad undergrad	Pillai's Trace	.005	1.302[a]	1	240	.255
	Wilks' Lambda	.995	1.302[a]	1	240	.255
	Hotelling's Trace	.005	1.302[a]	1	240	.255
	Roy's Largest Root	.005	1.302[a]	1	240	.255

a. Exact statistic
Note Wilks' Lambda, Hotelling Trace and Roy's Largest Root removed when space would be conserved. Since exact statistics, no information is lost.

b. Design: Intercept + gradundergrad
Within Subjects Design: question + rater + question * rater

Tests of Between-Subjects Effects Exploratory Use only

Measure: Rating by student and by instructor
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	54923.307	1	54923.307	4993.12	.000
Grad Undergrad	704.896	1	704.896	64.083	.000
Error	2639.948	240	11.000		

Continuation of Table 4 showing between subject effects

In Figure 1, if we interpret the average achievement level as a percentage, graduate student ratings indicate that the students believe their achievement for solving engineering problems at about 87%. The instructor rating for graduate students solving engineering problems is about 82% Undergraduate ratings place their abilities to solve problems at about 72%, while the instructor indicates the undergraduate ability to solve problems averages only about 60%.

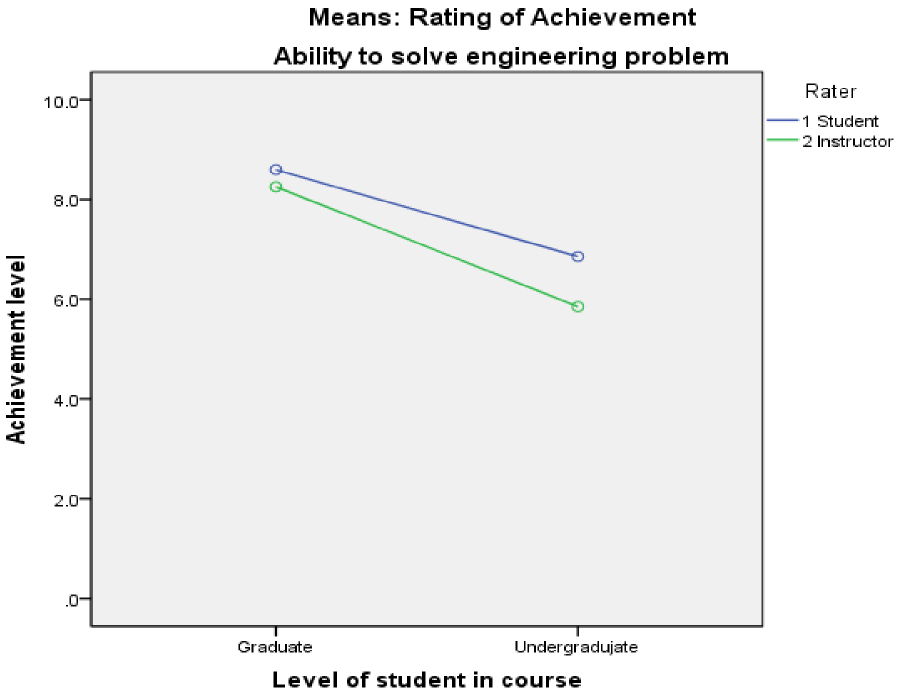

Figure 1. Average fit for ratings on ability to solve engineering problems, top line is student rating. Bottom line is instructor rating. There is a larger difference between undergraduate student ratings and instructor ratings as indicated in the interaction terms.

In Figure 2below, if we interpret the average achievement level as a percentage, graduate student ratings indicate that the students believe their achievement for communicating about engineering problems at about 85%. The instructor rating for graduate students communicating about engineering problems is about 80% Undergraduate ratings place their abilities to communicate about engineering problems at about 77%, while the instructor indicates the undergraduate ability to solve problems averages only about 70%. The small interaction between rater and type of ability is best shown in another set of graphs (See Figures 3 and 4.)

Figures 3 and 4 show the same information, but in these two figures, the type of ABET ability is shown on the small plot with graduate and undergraduate levels shown on separate plots.

The third way to view these averages is to view the raters separately (Figures 5 and 6.) These various graphs show the same information, but from a different orientation.

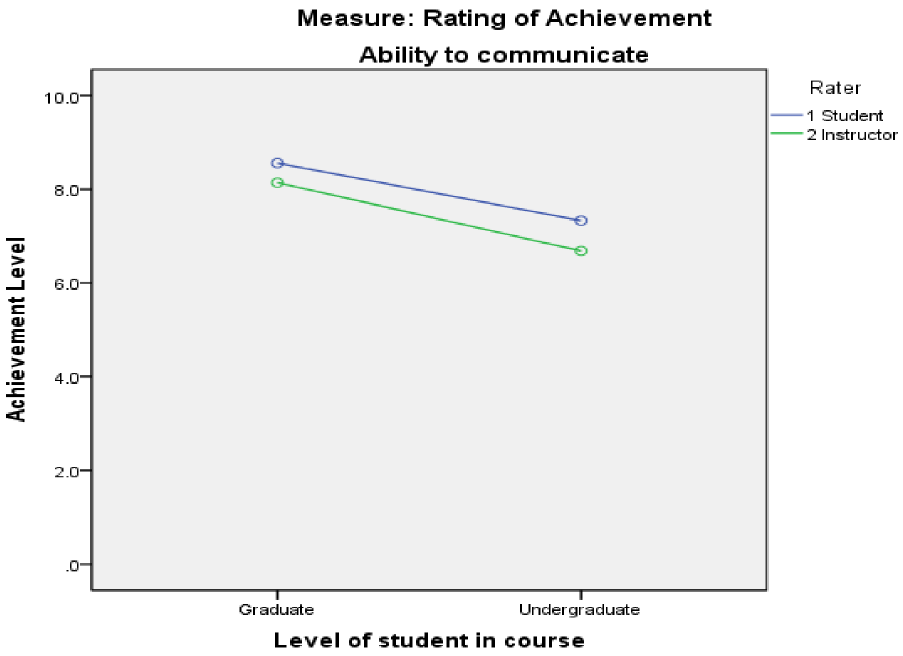

Figure 2. Average fit for ratings on ability to communicate about engineering problems, top line is student rating. Bottom line is instructor rating. There is a larger difference between undergraduate student ratings and instructor ratings as indicated in the interaction terms.

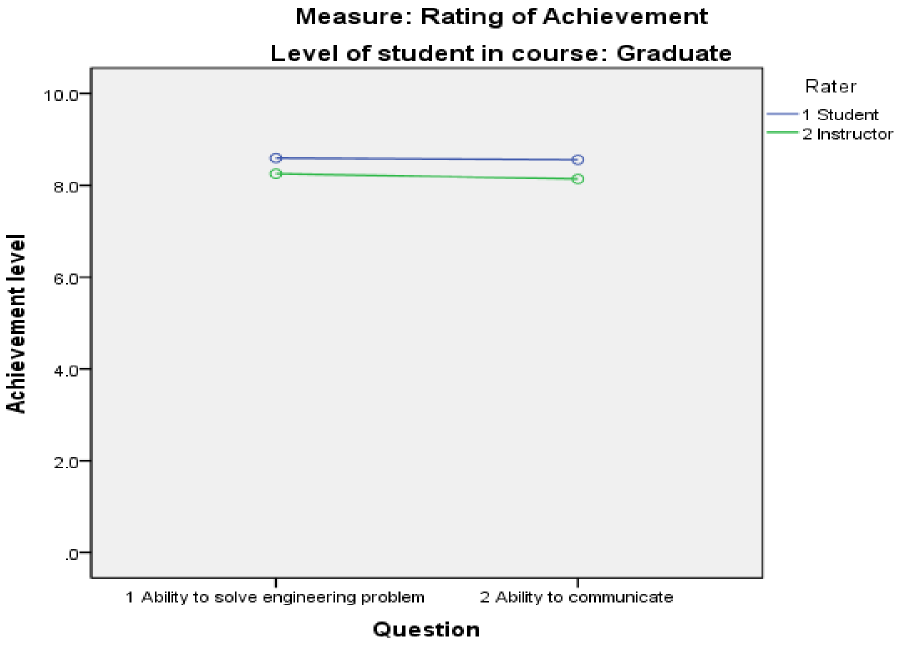

Figure 3: Graduate students only as rated on ability to either solve or communicate about a specific engineering problem, first by the student (bottom line) and second by the instructor (bottom line).

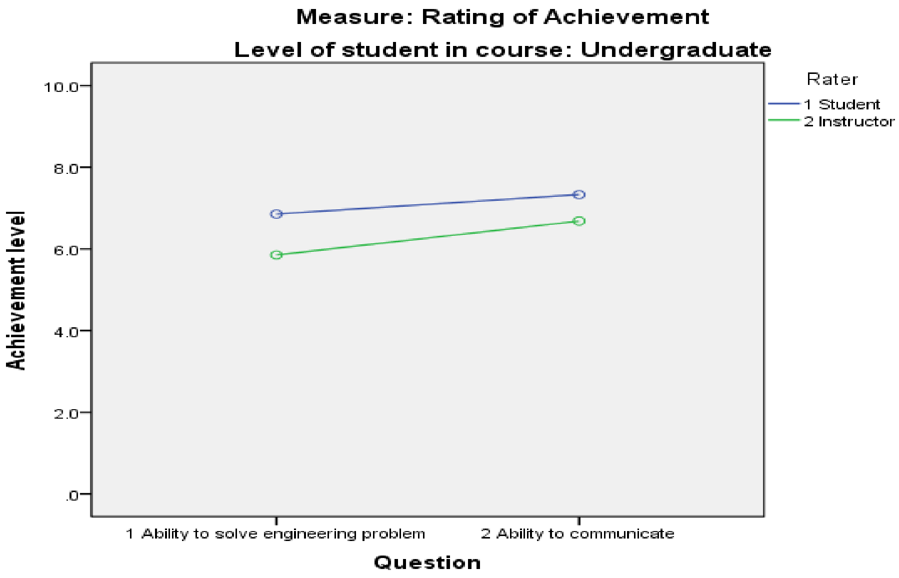

Figure 4: Undergraduate students only as rated on ability to either solve or communicate about a specific engineering problem, first by the student (top or blue line) and second by the instructor (bottom or green line).

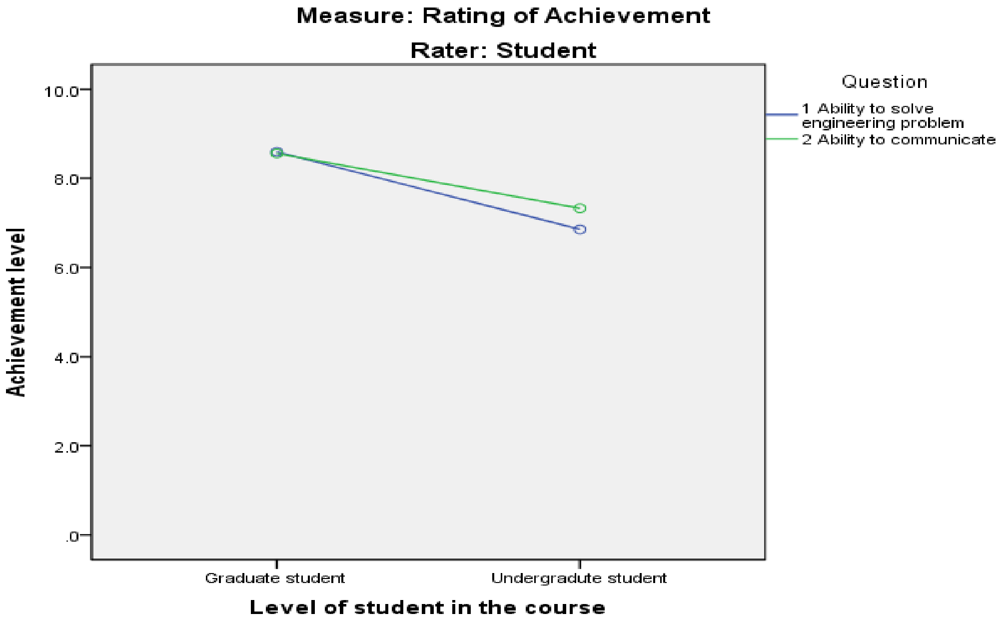

Figure 5: Considering the students alone, graduate students assess their abilities to solve and communicate about engineering problems as the same, but undergraduates feel that they are better able to communicate about engineering problems than to solve them.

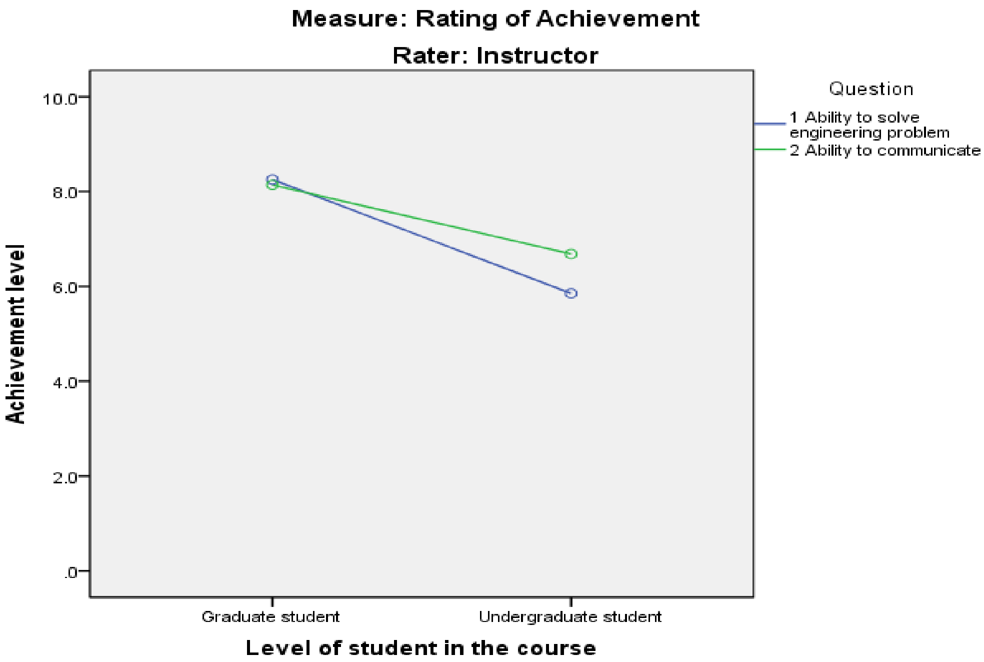

Figure 6: Considering the instructors alone, instructors rate graduate student abilities to solve and communicate about engineering problems as approximately the same, and the instructor also rates undergraduate as better able to communicate about engineering problems than to solve them.

Finally, we can ask about introductory version intermediate courses by creating a 4 level variable indicating both the course and student level simultaneously. Figures 7 through 10 include an indication of course level as well as student level. We expect the difficulty of the problems to increase from level to level within grade and more advanced students to assess better than beginning students. This trend is generally true and is easily visualized, especially for the graduate students. There is some downward evaluation of the undergraduate students at the intermediate level both by the instructor and by the students on the ability to solve problems. For the communication problems the instructor evaluations turns downward for the graduate students at the most advanced course level.

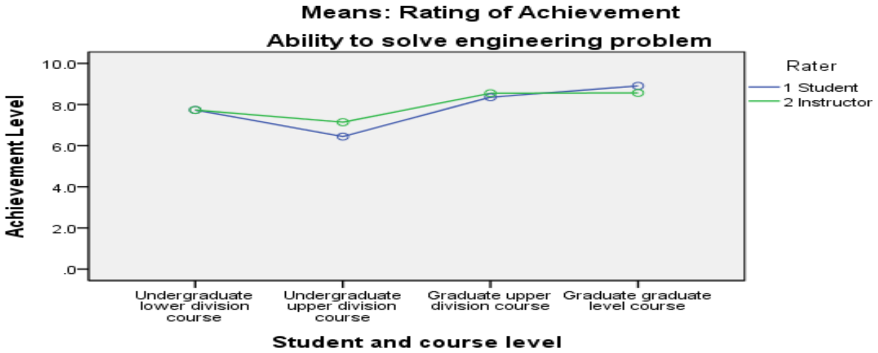

Figure 7: Ability to solve engineering problems as rated by (1 blue) Student and (2 green) Instructor for progressively advancing students in progressively advancing courses.

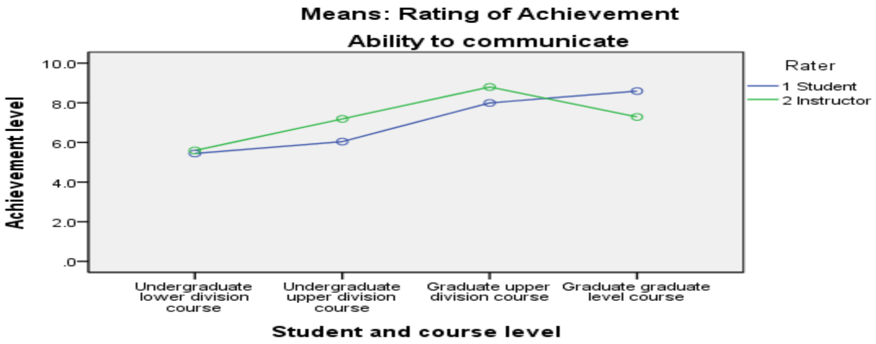

Figure 8: Ability to communicate about engineering problems as rated by (1 blue) Student and (2 green) Instructor for progressively advancing students in progressively advancing courses.

Student assessment of communicating the required engineering ideas is lower than that of instructor assessment until the highest course level. The task at the graduate level has not kept pace in the eyes of the instructor. More work is necessary either by the instructor at setting the standards or by the students in assessing limitations.

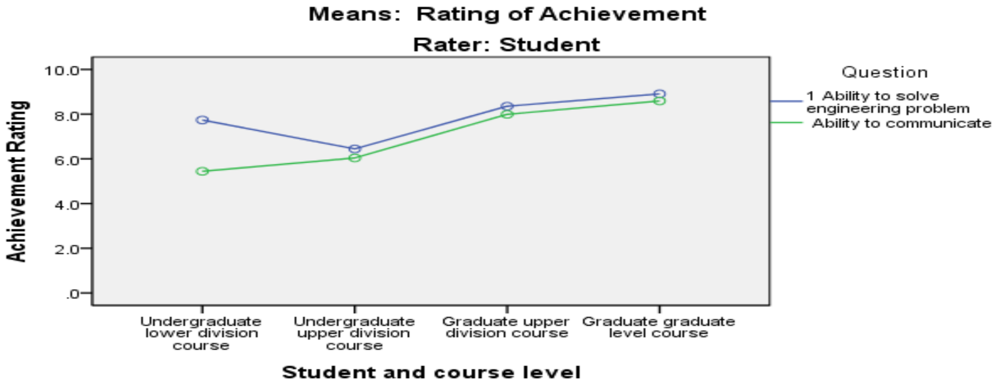

Figure 9: Differentiating types of ABET assessment items, (1 blue) Ability to solve engineering problems and (2 green) Ability to communicate about engineering problems as rated by Student for progressively advancing students in progressively advancing courses.

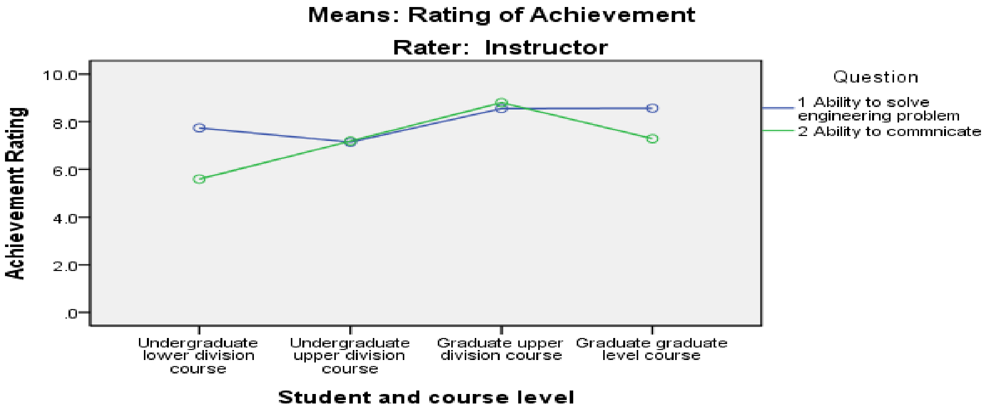

Figure 10: Differentiating types of ABET assessment items, (1 blue) Ability to solve engineering problems and (2 green) Ability to communicate about engineering problems as rated by Instructor for progressively advancing students in progressively advancing courses.

At the introductory level students assess their ability to solve real problems higher than their ability to communicate, but show expected progress as the level increases. The student communications ratings show steady progress as course level increases. Beginning students perceive their abilities to solve problems as considerably higher than they assess communication. Instructors also consider problem solving as generally as good or better than communicating about engineering for both problem solving and communication.

4. Summary

If educators are considering the self-study model of asking students how sure they are of the knowledge that they have obtained, at least in this setting of assigned surety to individual problems, we found that there is a similar pattern and association between faculty evaluation and student confidence in a particular answer as we did last year introductory psychology courses and previously in introductory statistics courses. In statistics courses, student's ability to self-evaluate was related to grade earned in the course. We did not have similar data in this year's study of engineering students.

References

1. Dietz, Zachariah, Lovell, J. D., and Norton, J. A. (2011) " Including Student Ability to Assess Learning with Other Assessment Tools", *American Statistical Association 2011 Proceedings of the Section on Statistical Education.*
2. Dietz, Zachariah, Lovell, J. D., Norton, Julia A. and Norton, John A (2012) "Student Self-Evaluation of Learning with Other Assessment Tools, Part II", *American Statistical Association 2012 Proceedings of the Section on Statistical Education.*

3. Eudey, L., Anand A., Norton J.A., and Coulman, L. (2009) “Assessment in Student Services”, *Proceedings of the American Statistical Association Section on Statistical Education.*
4. Garfield, J. B. and D. Ben-Zvi (2007). *Developing students' statistical reasoning: connecting research and teaching practice.* Emeryville, CA: Key College Publishing.
5. Lovell, J. D. and Norton, J. A. (2002) "Percent Content Mastery Testing Of Content In College Courses", *Proceedings of the American Statistical Association Section on Statistical Education.*
6. Norton, J.A. and L. Eudey (2006) "Assessment in Graduate Statistics Programs", *Proceedings of the American Statistical Association.*
7. Norton, J. A. and Lovell, J. D. (2000) "Repeated Measures Design In Assessment: Added Value of Instruction", *American Statistical Association 2000 Proceedings of the Section on Government Statistics and Social Science*, pp. 282-283.
8. Norton, J. A., Zhou, Y., and Ganjeizadeh, F. (2008). “Better Features in Teaching Introductory Statistics”, *American Statistical Association 2008 Proceedings of the Section on Statistical Education.*
9. Norton, J. A. (2001) "Assessment of Introductory Statistics: Phase II", *Proceedings of the American Statistical Association Section on Statistical Education.*
10. Norton, J., J. Lovell, E. Suess, B. Trumbo, C. Sugahara, W. Rodriguez, J. Fowler, E. Worth, R. Young, T. Grube, V. Sue. (1999). "Statistics Program Assessment," *American Statistical Association 1999 Proceedings of the Section on Statistical Education, pp.* 298-303.
11. Norton, J. A., and Lovell, J. D. (1981). “Statistics as an appropriate subject for satellite-based distance teaching”, *Education for the Eighties: The Impact of the New Communications Technology*. Deakin University Press, 151-154.
12. Ribeiro, and Maria Da Graça N. Mizukami (2005). “Problem-based learning: a student evaluation of an implementation in postgraduate engineering education”, *European Journal of Engineering Education, Volume 30.*
13. Wong, Y., Lovell, J. D., Norton, J. A., and Norton J. A. (2013). “Measurement Usage in Self-Assessment,” *American Statistical Association 2013 Proceedings of the Section on Statistical Education.*