Joint Assessment and Evaluation of Senior Design Projects by Faculty and Industry

Samuel J. Dickerson[1], Steven P. Jacobs[1], Adrian M. Garcia[2] and David V.P. Sanchez[2]
[1]Department of Electrical and Computer Engineering
[2]Department of Civil and Environmental Engineering
Swanson School of Engineering
University of Pittsburgh
Pittsburgh, PA, USA

***Abstract*—Most engineering programs culminate in a capstone senior design project that are often evaluated only by faculty despite the fact that the majority of the graduating seniors will be transitioning to careers in industry. In this paper, we demonstrate an innovative educational assessment tool for simultaneously integrating industry and faculty in-person evaluations of senior design projects. Two rubrics were created and aligned to evaluate how well the students satisfied expected student outcomes for senior design: one for use by faculty members and the other for use by industry judges. These rubrics enable in-person in-depth evaluations of the senior design projects at the end-of-semester project showcase. These rubrics were aligned using ABET student learning outcomes. Roughly 30 senior design projects received a combined total of 200 evaluations from both faculty and industry to ensure that each project was evaluated across all criteria by both groups of evaluators. In our assessment tool, industry and faculty scores for each project were plotted in two dimensions to visualization of the degree of correspondence. The strength of alignment between faculty and industry is indicated by the slope of the mean line and the sample variance from this line. Additionally, the Cohen's kappa inter-rater reliability statistic is used. This assessment tool can be used to highlight areas of improvement for departments and senior design programs.**

Keywords—Capstone Design, Assessment, Industry, ABET

I. Introduction

Undergraduate engineering programs typically culminate in a capstone design project. In the department of Electrical and Computer Engineering at the University of Pittsburgh, as is the case with many other universities around the world, this involves a project in which students must:

1. Identify the problem and the design criteria for a suitable solution
2. Gather relevant information and constraints
3. Generate multiple solutions by synthesizing and applying appropriate engineering knowledge
4. Choose the best solution based on available and derived information while properly considering relevant constraints (economic, environmental, sustainability, ethical, health and safety, manufacturability, social and political)
5. Evaluate the chosen solution against the design criteria through experimentation and data analysis.

The capstone design experience is a critical checkpoint for programs, providing faculty and administrators an opportunity to routinely measure terminal student outcomes and overall program health. The majority of students seek employment in industry shortly after their senior capstone project. As such, the capstone project provides an ideal point in time to assess and evaluate the degree to which the department has prepared its students to begin their professional careers. Concomitantly, capstone design experiences are essential components in meeting standardized engineering education accreditation criteria, such as those set by the Accreditation Board for Engineering and Technology (ABET) [1]. As part of the current ABET engineering criteria (EC2000), the assessment of student outcomes within the context of a design project is a required procedure.

The use of industry to evaluate student projects is common practice, and their feedback is often used to inform faculty and administrators, and monitor student progress [2]. Hotaling et al. presented a sophisticated multi-variable regression analysis to demonstrate that industry scoring is useful in projecting long-term outcomes such as job placement [3]. Furthermore, researchers found it informative to use both alumni and industry partners in assessing the broad, higher-level objectives of their capstone course design.[4]

As a result of the various needs for capstone assessment, a wide variety of internal and external assessment strategies for capstone projects have been proposed and employed throughout the year. However, finding adequate summative assessments of open-ended student-driven design projects continues to be a challenge. That challenge is compounded for new faculty and external evaluators (i.e. industry) who are not in the field of education, nor experienced in assessing student work.

Previous reports of multi-rater capstone assessment strategies that include faculty have been reported in engineering education research literature [5-11]. While some programs have incorporated both faculty and industry feedback into their capstone assessment process, those results have been limited by the number of industry evaluators, use of indirect tools (i.e. satisfaction surveys), a lack of alignment with ABET student outcomes, and a non-integration of faculty and industry scores. In this regard, there has not been an assessment framework that aligns industry and faculty design evaluations using simultaneous, in-person, direct quantitative assessments to provide insights on the relative perspectives from both evaluator types and the overall health of the senior design program.

Finally, a national survey found that while most universities agree that the capstone project is an appropriate point in time to assess student outcomes, they also indicated that the outcomes are not being assessed to the degree they should be [2], highlighting the need for new innovative capstone assessment methods.

In this paper, we present a two-dimensional assessment tool for evaluating student projects based on simultaneous in-person, direct-assessments by faculty and industry. From this assessment tool, we are able to learn the degree to which faculty evaluations, industry expectations, accreditation criteria and student outcomes are in alignment.

Two rubrics were created to evaluate how well the students satisfied each of the senior design requirements mentioned previously: one for use by faculty members and the other for use by industry judges, through in-person evaluations at the end-of-semester project showcase. These rubrics were aligned using ABET student learning outcomes in order to map all participant responses and ensure that each project was evaluated across all criteria by both faculty and industry representatives. In our assessment tool, industry and faculty scores for each project were plotted in two dimensions to allow the user to visualize the degree of correspondence. The strength of correlation is indicated by the slope of the mean line and the sample variance from this line. Additionally, inter-rater reliability is evaluated by means of the Cohen's kappa statistic.

II. Methods

Beginning with the rubric established by the ECE department to evaluate senior design projects across ABET criterion, a streamlined rubric was created for industry judges and distilled into 5 main categories: Resolves a Design Problem, Approach, Innovativeness, Presentation and Communication, and, Engineering and Technical Performance. The motivation for creating a new rubric was to have an instrument that was simpler to digest (for participants not familiar with ABET), to engage their industry experience and intuition in assessments and, to enable industry judges to evaluate multiple teams easily. The mapping and alignment of the evaluation criteria are delineated in table I below.

Simultaneous evaluation of the projects by both faculty and industry took place at the Senior Design Exposition for both the Fall and Spring semesters. This event is an open-forum project showcase in which evaluators were able to personally visit and engage individual senior design teams. This format was preferred because it allowed faculty and industry judges direct and unrestricted access to gathering information for their assessments. Faculty were recruited ad hoc for the evaluation of the projects, while personnel from industry were randomly assigned digital ballots using an online tool.

We present results from ten different metrics of evaluation for thirty different student projects. Analysis of the resulting plots enable an in-depth evaluation of our senior design course, student progress, and assessment rubrics from multiple perspectives. Do faculty members and industry representatives value the same qualities in design projects? Are our engineering programs preparing students to meet the needs of industry? In this paper, we use our assessment tool to investigate these questions.

TABLE I. Mapping and Alignment of Evaluation Criteria

INDUSTRY RUBRIC	**FACULTY CRITERIA MAPPED TO INDUSTRY RUBRIC**	**ABET related Criterion**	**ABET Criterion (student outcomes)**
Resolves a Design Problem	3ci. Identify project objectives based on general project and client requirements	3c.	
Approach/Innovativeness	3cii. Gather and use relevant information for design recommendations 3ciii. Generate and analyze alternatives by synthesizing and applying appropriate engineering knowledge 3civ. Consider all relevant constraints if applicable (i.e. Economic, Environmental Sustainability, Manufacturability, Ethical/health and safety, Social/political) 3cv. Choose the best solution based on technical and economic criteria and considering relevant constraints	3c.	**3b.** an ability to design and conduct experiments, as well as to analyze and interpret data
Presentation/Communication	3gii. Oral presentation	3g.	
Engineering/Technical Performance	3bi. Experimentation designed effectively 3bii. Appropriate and correct data analysis and interpretation	3b.	**3c**. an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability

In our analyses, we calculate standard statistics on the collected data (e.g. mean and variance), as well as the Cohen's kappa (κ) [12]. Cohen's kappa is a statistical measure of inter-rater reliability and can be used to provide additional insight on the level of agreement between two assessors evaluating a population of subjects. The kappa value is calculated as:

$$\kappa = \frac{p_a - p_e}{1 - p_e}$$

where p_a is the relative number of observations in agreement, and p_e is the probability of an agreement occurring by chance. Since Cohen's kappa takes chance agreements into consideration, it is a robust measurement to use when analyzing evaluation results reported by multiple raters. In order to evaluate the kappa values, which assumes there are only two raters, across all faculty and industry, the average rating reported by each group for each category was used. Determination of the kappa value requires evaluators place subjects into distinct categories. Therefore, after the average ratings were computed, the values were partitioned into bins ranging from 0 to 5 in increments of 0.5, allowing the computed average ratings to be discretized and tallied.

There is no standard for interpreting the value that results from computing the Cohen's kappa statistic. However, the most widely cited interpretation is provided by Landis and Koch [13]. In their interpretation the best possible strength of agreement between raters is when $\kappa = 1.0$, values in the range of $0 < \kappa < 1.0$ represent varying levels of agreement from slight to substantial and $\kappa < 0$ indicates poor agreement. The preceding interpretation of κ is summarized in table II.

TABLE II. INTERPRETATION OF COHEN'S KAPPA

κ	Interpretation
< 0	Poor agreement
0.01 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

III. RESULTS AND DISCUSSION

A. *Identifying Project Objectives and Requirements (Resolves a Design Problem vs 3c.i)*

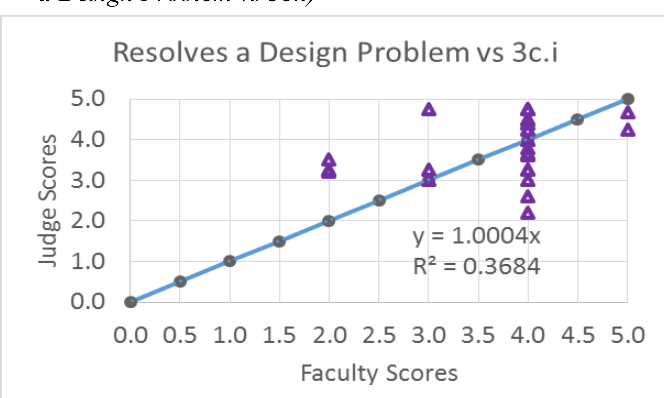

Fig. 1. Resolves a Design Problem vs 3c.i

TABLE III. RESOLVES A DESIGN PROBLEM VS 3C.I

	Judges	Faculty (3c.i)
Mean	3.9	3.7
Standard Deviation	0.7	0.8
Minimum	2.2	2.0
Maximum	5.0	5.0
Range	2.8	3.0
κ	0.0405	

In Figure 1, the slope of the linear fit (1.0004) suggests a strong alignment between faculty and industry with respect to resolving a design problem. On the other hand, the Cohen's kappa coefficient of 0.0405 suggests slight agreement for their ratings. Examining the spread of the scores one can see that while there are mismatches on several projects where industry judges score them higher than faculty or vice versa, these are less common and both industry and faculty scores fall within the range of 2.0-5.0.

B. *Gathering Relevant Information (Design Approach / Innovation vs 3c.ii)*

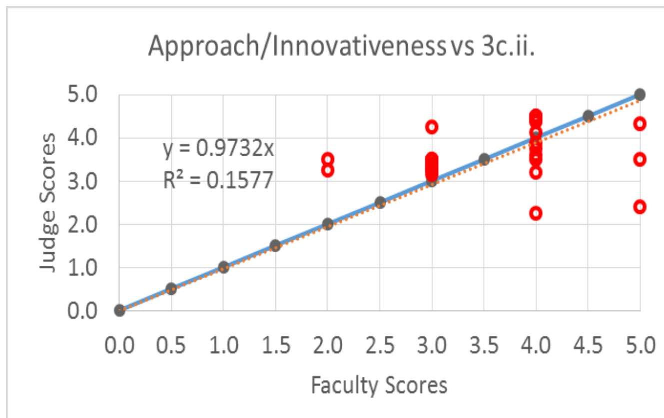

Fig. 2. Design Approach/Innovation vs 3c.ii

TABLE IV. DESIGN APPROACH/INNOVATION VS 3C.II

	Judges	Faculty (3cii)
Mean	3.9	3.6
Standard Deviation	0.7	0.8
Minimum	2.2	2.0
Maximum	5.0	5.0
Range	2.8	3.0
κ	0.1041	

In Figure 2, the slope of the linear fit (0.9732) suggests that faculty have a minor tendency to rate the projects higher than industry judges. Upon closer inspection, the key feature that moved this trend is the fact that several faculty afforded the highest ratings possible to the teams, but no judges from industry did so. The Cohen's kappa coefficient also suggests a slight agreement between industry and faculty ratings with a statistic of (0.1041). A qualitative follow up to these rubrics would be to evaluate the respective faculty and industry expectations of senior design teams. This would provide a better qualitative understanding of their perspectives.

C. Generating and Synthesizing Alternatives (Design Approach/Innovation vs 3c.iii

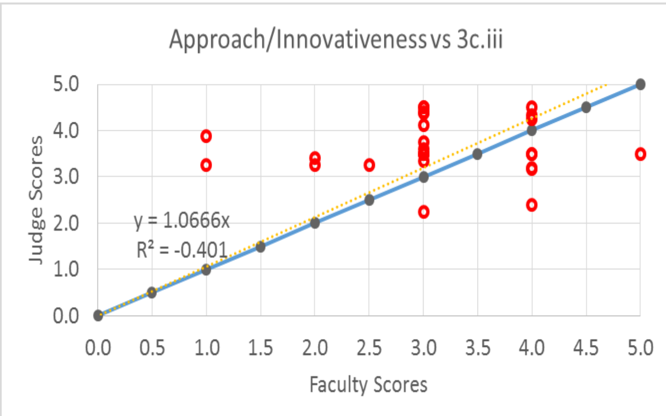

Fig. 3. Design Approach/Innovation vs 3c.iii

TABLE V. DESIGN APPROACH/INNOVATION VS 3CIII

	Judges	Faculty (3c.iii)
Mean	3.9	3.1
Standard Deviation	0.7	0.9
Minimum	2.2	1.0
Maximum	5.0	5.0
Range	2.8	4.0
κ	-0.1634	

As shown in Table V, the faculty evaluators provided overall lower ratings to senior design teams on their ability to generate and analyze alternatives. Here for the first time the faculty scores traverse the complete range of the scoring rubric. The difference demonstrates that the ECE department intrinsically utilizes a more rigorous standard for this criterion thus affirming current departmental practice. The misalignment in average scores with industry judges corresponds with the Cohen's kappa inter-rater reliability coefficient of (-0.1634) which suggests a poor agreement between faculty and industry.

D. Considering Design Constraints: (Design Approach/Innovation vs 3c.iv)

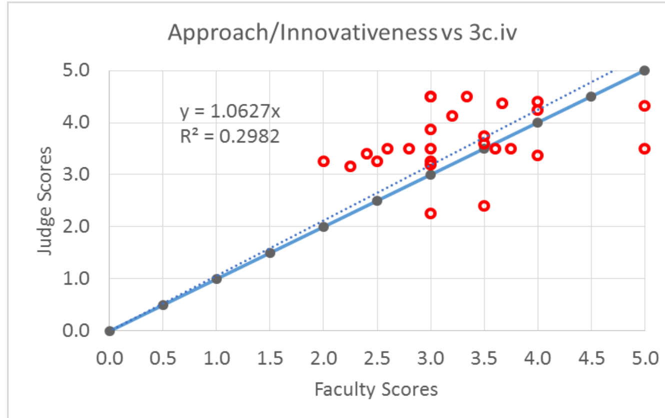

Fig. 4. Design Approach/Innovation vs 3c.iv

TABLE VI. DESIGN APPROACH/INNOVATION VS 3C.IV

	Judges	Faculty (3c.iv)
Mean	3.9	3.3
Standard Deviation	0.7	0.7
Minimum	2.2	2.0
Maximum	5.0	5.0
Range	2.8	3.0
κ	0.1337	

The relevant design constraints considered by faculty when evaluating senior design teams included economics, environmental sustainability, ethics, manufacturability and social/political contexts. In Figure 4, the scores provided by industry judges for innovation and approach category did not correlate well to faculty evaluations as shown by R^2 value (0.2982). However the Cohen's kappa coefficient (0.1337) does suggest that there is a slight agreement between industry and faculty raters. Average industry scores were much higher for this criterion with the standard deviation being equal. As in previous figures there is not a major cluster of scores at the 3.0 and 4.0 mark.

E. Selecting the Best Design Solution (Design Approach/Innovation vs 3c.v)

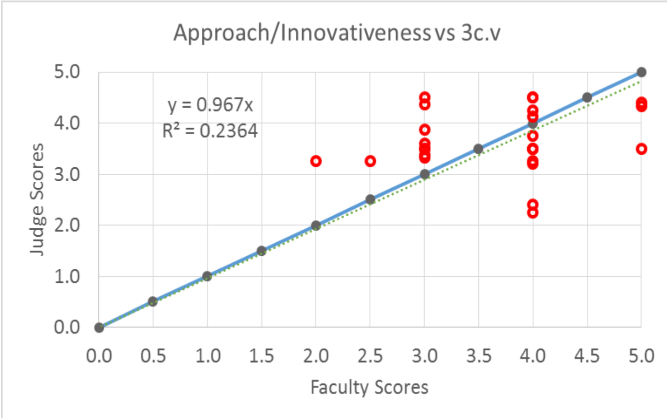

Fig. 5. Design Approach/Innovation vs 3cv

TABLE VII. DESIGN APPROACH/INNOVATION VS 3C.V

	Judges	Faculty (3c.v)
Mean	3.9	3.7
Standard Deviation	0.7	0.8
Minimum	2.2	2.0
Maximum	5.0	5.0
Range	2.8	3.0
κ	-0.1089	

There was positive correlation and alignment for the best design solution. With a smaller standard deviation and smaller range, the data shows that the industry judges generally score the senior design projects higher in this category. The Cohen's kappa coefficient (-0.1089) also suggests a poor agreement between faculty and judge ratings.

F. Oral Presentation (Presentation/Communication vs 3g.ii)

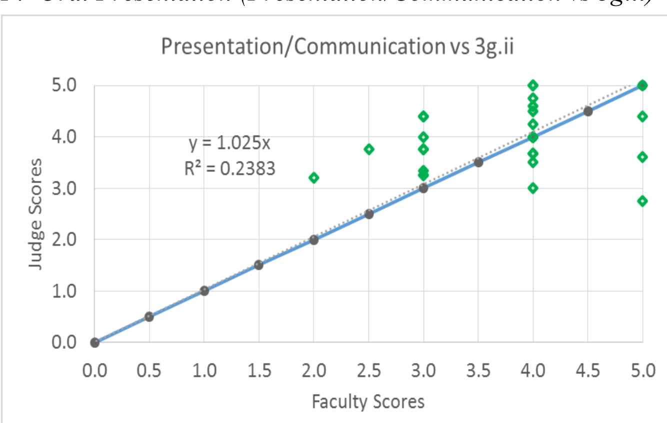

Fig. 6. *Presentation/Communication vs 3g.ii*

TABLE VIII. *PRESENTATION/COMMUNICATION VS 3G.II*

	Judges	Faculty (3g.i)
Mean	3.9	3.7
Standard Deviation	0.7	0.8
Minimum	2.5	2.0
Maximum	5.0	5.0
Range	2.5	3.0
κ	0.0465	

For the last 3 categories of Presentation and Communication, Technical Performance and Overall impact, the industry judges consistently assessed the senior design teams more positively than faculty. With higher averages, smaller standard deviations, and a smaller range of scores, the evaluations from industry judges were clustered toward the upper bound. Moreover the Cohen's kappa of 0.0465 for 3g.ii, -0.105 for 3b.i and, 0.0601 for 3b.ii suggest that there is not a strong agreement between faculty and industry for these categories. Contrasting these scores with the breadth of faculty scores it demonstrates that for these three criteria the faculty standards for senior design projects are adequate for industry expectations.

G. Technical Performance- Experimental Design (Technical Performance vs 3b.i)

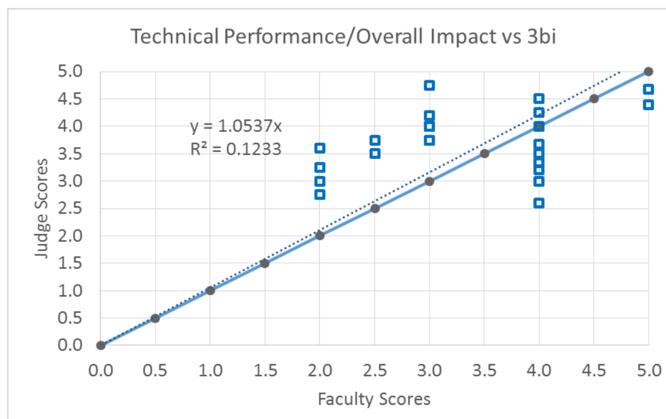

Fig. 7. *Technical Performance vs 3bi*

TABLE IX. *TECHNICAL PERFORMANCE VS 3B.I*

	Judges	Faculty (3b.i)
Mean	3.7	3.3
Standard Deviation	0.5	0.9
Minimum	2.6	2.0
Maximum	4.8	5.0
Range	2.2	3.0
κ	-0.105	

H. Technical Performance- Experimental Design (Technical Performance vs 3b.ii)

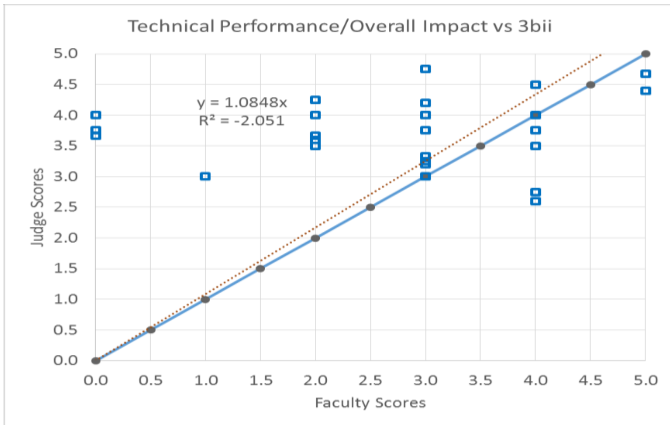

Fig. 8. *Technical Performance vs 3bii*

TABLE X. *Technical Performance vs 3bii*

	Judges	Faculty (3b.ii)
Mean	3.7	2.8
Standard Deviation	0.5	1.4
Minimum	2.6	0.0
Maximum	4.8	5.0
Range	2.2	5.0
κ	-0.0601	

Figure 9 displays a summary of the Cohen's kappa statistics determined across faculty and industry evaluations on student senior design projects.

IV. Conclusions

In this paper we presented a new framework and assessment tool for integrating faculty and industry in-person evaluations of senior design projects. Over 200 combined evaluations of 30 senior design projects were conducted by both faculty and industry judges at an open-forum end of the semester project showcase. A two-dimensional plot and descriptive statistics were used to analyze alignment between both evaluator types and ABET student outcomes. Additionally, inter-rater reliability was evaluated using the Cohen's kappa statistic. Results showed both similarities and discrepancies between faculty and judges using the average, range and variance for their project scores. These results can be used to both affirm positive practices and outcomes of the senior design capstone program as well as areas in need of improvement. The joint evaluation tool will continue to inform the future direction of the senior design program.

References

[1] www.abet.org

[2] L. J. McKenzie, Michael S. Trevisan, Denny C. Davis, S. W. Beyerlein. Capstone design courses and assessment: A national study. American Society for Engineering Education Annual Conference and Exposition. Salt Lake City, UT2004.

[3] N. Hotaling, B. B. Fasse, L. F. Bost, C. D. Hermann, C. R. Forest. A Quantitative Analysis of the Effects of a Multidisciplinary Engineering Capstone Design Course. Journal of Engineering Education 2012;101:630-56.

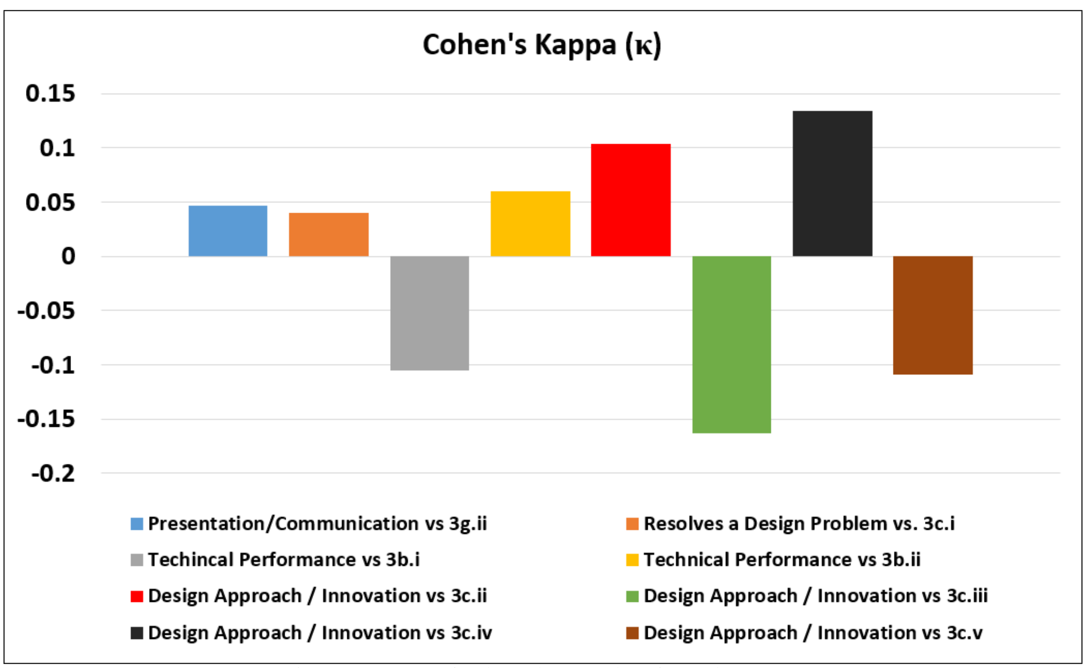

Fig. 9. *Summary of Cohen's kappa values for each category*

[4] R. Bob, A. W. Clifford, T. A. Jacob, R. Peter. Examining the Structure of a Multidisciplinary Engineering Capstone Design Program. Indianapolis, Indiana: ASEE Conferences.

[5] H. Cheng, K. Junichi, G. Lester, S. Mark, A. Richard. A Holistic Approach For Student Assessment In Project Based Multidisciplinary Engineering Capstone Design. Louisville, Kentucky: ASEE Conferences.

[6] M. Faust, A. Greenberg, B. Pejcinovic. Redesign of senior capstone program in electrical and computer engineering and its assessment. 2012 Frontiers in Education Conference Proceedings2012. p. 1-6.

[7] H. Susannah, W. Jessica. 2005 National Survey Of Engineering Capstone Design Courses. Chicago, Illinois: ASEE Conferences.

[8] D. K. S. I. V. K. Jain. Two Instruments for Assessing Design Outcomes of Capstone Projects. American Society of Engineering Education Annual Conference. Salt Lake City, UT2004.

[9] E.-S. Mohamed, S. Matthew, K. Lucy, L. Michael, T. Mark. Assessing Interdisciplinary Engineering Capstone Project. Chicago, Illinois: ASEE Conferences.

[10] T. Phillip, H. Olakunle, D. Denny, B. Steven, T. Michael. Assessment Framework For Capstone Design Courses. Chicago, Illinois: ASEE Conferences.

[11] D. Knight, D. Kotys-Schwartz. Once again around the double triangle: A multi-rater assessment of capstone design skills. 2013 IEEE Frontiers in Education Conference (FIE)2013. p. 1018-20.

[12] Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960; 20, 37-46.

[13] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33: 159-74.