AC 2011-775: A DIRECT METHOD FOR TEACHING AND MEASURING ENGINEERING PROFESSIONAL SKILLS: A VALIDITY STUDY

Ashley Ater Kranov, Washington State University

Dr. Ater Kranov is Director of Educational Innovation and Assessment for the College of Engineering and Architecture at Washington State University, USA. She is affiliated assistant professor in the School of Electrical Engineering and Computer Science where she co-teaches the 2-semester senior design capstone sequence.

Dr. Ater Kranov is a leader in university and community internationalization efforts, including developing and assessing global competencies in faculty, staff, and students. The paper describing her collaborative work with faculty in the WSU College of Engineering and Architecture, "A Direct Method for Teaching and Assessing the ABET Professional Skills in Engineering Programs", won the 2008 ASEE Best Conference Paper Award. She has served as evaluator on a number of multi-institutional, interdisciplinary NSF sponsored grants. She is principal investigator on a NSF Research and Evaluation on Education in Science and Engineering project called "A Direct Method for Teaching and Measuring Engineering Professional Skills: A Validity Study."

Mo Zhang, Washington State University

Mo Zhang is a doctoral student major in educational psychology at Washington State University. Her research interests include applied statistics, educational measurement, design of experiments, sampling theories, and item response theory oriented mathematical models. She holds an M.A. in education from Washington State University.

Steven W. Beyerlein, University of Idaho, Moscow

Dr. Beyerlein is a professor of Mechanical Engineering and coordinator of the college-wide inter-disciplinary capstone design program at the University of Idaho where he has been on the faculty since 1987. He is involved in a number of research projects and initiatives related to design pedagogy, professional skills assessment, catalytic combustion, engine testing, and hybrid vehicle realization.

Jay McCormack, University of Idaho

Jay McCormack is an assistant professor in the mechanical engineering department at the University of Idaho where he is an instructor for the college's interdisciplinary capstone design course. Dr. McCormack received his PhD in mechanical engineering from Carnegie Mellon University in 2003.

Patrick D. Pedrow, Washington State University

Patrick D. Pedrow received the B.S. degree in electrical engineering from the University of Idaho, Moscow, in 1975, the M.Eng. degree in electric power engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1976, the M.S. degree in physics from Marquette University, Milwaukee, WI, in 1981, and the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, in 1985. From 1976 to 1981, he was with McGraw-Edison Company, where he conducted research and development on electric power circuit breakers. He is currently an Associate Professor with Washington State University, Pullman, where he is the Associate Director of the School of Electrical Engineering and Computer Science. His research interests are in plasma-assisted materials processing, including the deposition and evaluation of thin plasma-polymerized films. Dr. Pedrow is a member of IEEE, the American Physical Society, Tau Beta Pi and he is a Registered Professional Engineer in the State of Wisconsin. He has served on the Executive Committee of the IEEE Nuclear and Plasma Sciences Society Plasma Science and Applications Committee.

Edwin R. Schmeckpeper, Norwich University

Dr. Schmeckpeper is an Associate Professor at Norwich University's David Crawford School of Engineering, the oldest private engineering school in the nation. Prior to coming to Norwich University he was an Associate Professor at the University of Idaho.

A Direct Method for Teaching and Measuring Engineering Professional Skills: A Validity Study for the National Science Foundation's Research in Evaluation of Engineering and Science Education (REESE)

Introduction

This paper describes an ongoing research project in establishing the validity of a direct method for teaching and measuring undergraduate engineering students' professional skills. Proficiency in engineering professional skills (Table 1) is critical for success in the multidisciplinary, intercultural team interactions that characterize 21st century engineering careers. Yet, programs across the nation have struggled to define, teach and measure professional skills since their introduction as ABET criteria for engineering programs in 2000[1,2,3,4]. In fall 2006, the Washington State University College of Engineering and Architecture partnered with an assessment specialist to create an innovative, direct method to teach and measure the ABET professional skills simultaneously. *No direct method for teaching and measuring these skills simultaneously has existed in the literature prior to the Engineering Professional Skills Assessment* (EPSA)[5].

Table 1.ABET Criterion 3 Professional Skills Student Learning Outcomes

3d	Ability to Function on Multidisciplinary Teams
3f	Understanding of Professional and Ethical Responsibility
3g	Ability to Communicate Effectively
3h	Understanding of the Impact of Engineering Solutions in Global, Economic, Environmental, and Cultural/Societal Contexts
3i	Recognition of and Ability to Engage in Life-Long Learning
3j	Knowledge of Contemporary Issues

The major accomplishments of the four years of on-going research conducted college wide since spring 2007 using the EPSA method at the program level for evaluating the efficacy of the undergraduate engineering curriculum are: (1) an authentic performance task in the form of a scenario and prompts to elicit the ABET professional skills; (2) establishment of initial reliability and validity of the measurement instrument – the Engineering Professional Skills Rubric (EPS Rubric) (Appendix A); and (3) a dedicated community of 40+ engineering faculty using direct assessment to evaluate the efficacy of their own programs, and to plan and implement improvement at both course and program levels.

The EPSA method is a discussion-based performance task designed to elicit students' knowledge and application of the ABET professional skills. In a 45-minute session, small groups of students are presented with a complex, real-world scenario that includes multi-faceted, multidisciplinary engineering issues. They are then asked to determine the most important problem/s and to discuss stakeholders, impacts, unknowns, and possible solutions. Table 2 presents a summary of sample scenarios, and Appendix B provides three full scenarios with instructional prompts. The EPS Rubric, an analytic rubric, was developed to measure the extent to which student performance in response to a given scenario achieved the six learning outcomes associated with the ABET professional skills. This method is flexible, easy to implement, and can be used at the course level for teaching and measuring engineering professional skills and the program level at the end of a curricular sequence for evaluating a program's efficacy.

Table 2.Summary of Sample CD Scenarios

Lithium mining for electrical vehicle batteries	Hanford superfund site clean up
Need for prosthetics in land-mine ridden Iraq	Vehicle retrofitting for wheelchair-bound drivers
Strip mining on Navajo ceremonial lands	RFID tracking device privacy issues
2008 Tennessee Valley coal ash spill impacts	Links between power lines and cancer

The primary research goal of this research project sponsored by the National Science Foundation's Research in Evaluation of Engineering and Science Education (REESE) is to rigorously establish the reliability and validity of the EPSA method and the EPS Rubric through a significant collaboration among three disciplinarily-distinct engineering programs at two large land grant universities in the Pacific Northwest and one private university in the Northeast. *This project will directly contribute to fundamental research in engineering education on a problem of national importance and interest.* Table 3 outlines the project objectives.

Table 3. Project Objectives

Objective Category	Objective Description
1. Performance Task Development	Construct a framework to guide revision of existing scenarios and development of new scenarios (total = 35) with performance task prompts that equally elicit student consideration of the six ABET engineering professional skills.
2. Administration	Develop a manual that provides specifications for: faculty/proctor training & task implementation; scoring & reporting procedures; psychometric properties (e.g., task difficulty, discrimination, interrater reliability).
3. Task & Rubric Validation	Conduct a rigorous validation process by accumulating and examining content, construct, and criterion evidence, as well as by establishing intra &interrater reliability. Externally credible instruments to measure each skill will also be used to establish concurrent criterion validity. An Advisory Board comprised of established professional engineers, psychometric experts, engineering educators, and industry representatives will participate in the validation process by reviewing performance tasks and the rubric validity.
4. Documentation & Dissemination	Document the project, validation process & faculty implementation experiences. Disseminate performance tasks, EPS Rubric, administration manual and suggestions for course and program level use via the project website; presentations at local, national and international engineering conferences; cross-disciplinary journal publications and on-line networks (Center for Advancing Research & Communication/REESE Diffusion & Evaluation Network).

Review of the literature

Professional Skills in Engineering Education

Proficiency in engineering professional skills is critical for success in the multidisciplinary, intercultural team interactions that characterize 21st century engineering careers. Fifteen years ago, in its report *In Restructuring Engineering Education: A Focus on Change*, the National Science Foundation[6] recommended that engineering courses include early and continued exposure to environmental, political and social issues and their international and historical contexts, as well as legal and ethical implications of engineering solutions. This report was one of many that preceded the development of the professional skills and the requirements that engineering programs both teach and assess them beginning this century. To ensure continued competitiveness of American educated and trained engineers in the rapidly changing environment of the world economy and needs, engineering education must help students integrate professional and technical skills for more robust problem solving[7,8,3,9,10]. Therefore, there is a critical need to develop in students a deep understanding of the importance of the professional skills. Colleges and universities must align their curricula and teaching with the 21century workplace demands.

However,engineering programs across the nation have struggled to define, teach and measure professional skills since their introduction by ABET evaluation criteria for engineering programs in 2000[1,2,3,11].Although a variety of methods and instruments have been developed by engineering educators around the nation to teach and assess the ABET professional skills, most of the instruments evaluate one skill at a time[12,13,14,15,16,17,11,18,19]. They are often cumbersome to implement. And more frequently than not, they evaluate given skills indirectly through focus groups, interviews or surveys eliciting student opinions[4].

Engineering Professional Skills Assessment

The EPSA has three components: (1) a performance task including a scenario and prompts; (2) student discussion as a response to the task and; (3) an accompanying analytical rubric called the EPSRubric as a criterion-referenced instrument to measure the quality of the students' performance in demonstrating engineering professional skills. First, in a 45-minute session, groups of five to seven students are presented with a complex, real-world scenario that includes current, multi-faceted, multidisciplinary engineering issues. Second, students are asked to determine the most important problem/s and to discuss stakeholders, impacts, unknowns, and possible solutions. Finally, trained faculty raters use the analytical EPS Rubric to measure the extent to which student performance demonstrate the six learning outcomes associated with the ABET professional skills in response to a given task.

From 2007-2009, program-level EPSA scores were presented in a radar graph format to target audiences (faculty members, evaluation panels, and program administers, among others). See Appendix C for program-level assessment results from 2007-2009. Note that the results for year 2010 are not provided here. This is because the engineering professional skills assessment was delivered online in academic year 2010, instead of in the traditional face-to-face setting. As the validity and reliability of the online administered assessment needs to be further examined, the results were not compared with previous data.

Research goal and questions

The primary goal of this research project is to establish the validity and reliability of the Engineering Professional Skills Performance Assessment in measuring students' professional skills. Participants in a performance assessment "demonstrate their knowledge and skills by engaging in a process [and] or constructing a product"[20].The project's theoretical proposition is that the EPSA effectively elicits and accurately describes the content and constructs that comprise engineering professional skills.

Performance assessment typically has three components: (1) a task that elicits the performance; (2) the performance itself (which is the event or artifact to be assessed); and (3) a criterion-referenced instrument, such as a rubric, to measure the quality of the performance. In our study, the Engineering Professional Skills Performance Assessment also has three components: (1) the CD method (e.g., scenario and prompts) as the performance task; (2) the student team discussion as a response to the performance task; and (3) the EPS Rubric as the criterion-referenced instrument to measure the quality of the student team performance of engineering professional skills.

This research project is driven by the following three research questions:

1) To what extent does the CD method as a performance task equally elicit students' consideration of engineering professional skills when implemented in different course types and at different points in a program's curriculum?

2) Do EPS Rubric scores reliably provide information about students' engineering professional skills proficiency levels?

3) What is the correlation coefficient between the EPS Rubric's scores and scores from other established instruments that measure the same or similar skills?

Research process

The project's leadership team will use *Assessing Performance: Designing, Scoring and Validating Performance Tasks*[14] to guide the validation process framework we undertake. Prior to embarking on a focused validity study, it is crucial to complete a set of performance assessment design, construction and implementation processes and procedures. Therefore, year 1 of this project is devoted to a systematic review of the existing set of 20 performance tasks (e.g., scenarios and prompts), the latest version of the EPS Rubric, materials, processes and procedures in order to determine specifics in each of the following categories, as recommended by Johnson et al[20] (Table 4). An Engineering Professional Skills Assessment Manual will be developed during the first two years of the project.

Table 4. Major Project Validation Procedures

Procedure	Specifics
1. Test Specification	Determine the examinee characteristics Determine the outcomes/skills to be assessed Determine the desired proficiency level Determine the number and type of stimulus materials (e.g., consent forms, scenario, prompts) Determine the equipment needed (e.g., audio recorders)
2. Task Features	Create a framework to guide parallel task development Revise existing performance tasks using framework guidelines Determine the time allotment for each task Determine the number of prompts for each task Develop additional parallel tasks
3. Administration Materials	Develop standard implementation procedure Training sessions for investigators on how to facilitate implementation of the CD method to ensure equal implementation across sites and settings Provide sufficient support
4. Scoring and Reporting	Implement procedures to maintain consistent and accurate scoring Implement training sessions with project scoring groups Define procedures for resolving score differences Ensure scoring is suitably reliable to support the intended interpretation of the scores
5. Psychometric Properties	Determine psychometric properties of the performance tasks (e.g., difficulty, differential functioning levels) Determine psychometric properties of the EPS Rubric (e.g., interrater reliability, score reliability, types of validity evidence)
6. Documentation	Document the entire performance assessment validity process including changes made and rationales, data analyses, results and interpretations, conclusions, recommendations.

When an iteration of the validation process is complete, we will have a solid foundation upon which to carry out a *two-year* validation study (which will, of course, also include refinement of processes and procedures from the first year). The overarching goal is to provide a robust performance assessment that can be implemented and scored in different settings with different people and provide valid evidence of performance quality. Thus, a descriptive case-study methodology is a good fit for conducting the validity study of the engineering professional skills performance assessment.

Methods

A descriptive case-study methodology will be used as the framework for the design and analysis of this collective case study. We define case study as: an empirical inquiry into an event or set of related events within its real life context using multiple sources of evidence[21,22]. Case studies are useful for in-depth study of a particular problem, situation, or area of interest[23]. A collective case study, which we propose, is to study a number of cases to inquire into potential variations of seemingly similar phenomenon[22]. Using a descriptive collective case-study methodology will allow the investigators to understand and examine the contexts in which parallel performance tasks are implemented in three distinct sites and four distinct course-type settings (Table 5).

Design

In order to maximize what can be learned and to provide an adequate number of cases for a collective case study, we will use purposeful sampling at three levels: 1) institution/program, 2) level of course in the curriculum and 3) instructor/course-type. When conducting purposeful multi-site sampling, it's important to select sites that are expected to yield similar results for replication logic purposes (i.e., our theoretical proposition states that we predict similar results across cases).

Table 5. Multi-Site Sample (Participating Institutions, Programs, Course Level, & Course Type)

Institution	Engineering Program	Course Level	Course Type
A	Electrical & Computer	Senior	Capstone Design Sequence
B	Mechanical	Sophomore	Design
C	Mechanical, Bio Ag Electrical & Computer	Senior	Interdisciplinary Capstone Design Sequence
D	Civil	Sophomore Junior/Senior	Statics, Structural Analysis, Reinforced Concrete, Steel Design

A complete randomized design will be used to sample student into control and experiment groups within each course offering, where each course offering is a block of the analysis. The student team will be the primary unit of analysis within a block. Both the control and experiment groups will be presented with a scenario. However, the experiment group will receive two interventions. Experiment group student teams will (1) become familiar with the measurement instrument, the EPS rubric and (2) be provided with discussion prompts.

Instruments

An important strength of the case study is that it involves using multiple sources and techniques in the data gathering process[22]. In order to gather multiple sources of evidence, a number of instruments with established validity will be used to measure student performances and those scores will be compared to the EPS Rubric scores in our efforts to establish the concurrent criterion validity of the EPS Rubric. This provides opportunities for the triangulation of data during the analysis stage to answer our research questions. Table 6 enumerates the additional measures/instruments.

Table 6. Instruments for Determining Criterion Validity of EPS Rubric

Measure/Instrument	Corresponding ABETCriterion 3 Skill	Nature of Measure	Source
Engineering Ethics Rubric	f	D, QN	Shuman et al[4], 2004
Engineering Faculty Survey of Student Engagement	d/g, f, h, i, j	ID, QN	Cady et al[13], 2009
AAC&U Problem Solving Rubric	f, h	D, QN	AACU[24], 2010
AAC&U Lifelong Learning Rubric	i	D, QN	AACU, 2010
ASCE Body of Knowledge Rubric	d/g, j	D, QN	ASCE[12], 2008

Note. AAC & U refers to Association of American College and Universities. ASCE refers to American Society of Civil Engineers. D refers to direct measure. ID refers to indirect measure. QN refers to quantitative data. EPS refers to engineering professional skills.

Participants

We estimate that there will be 70 experimental teams and 66 control teams for a total of 136 teams (796 students) over the two years of the validation study. We estimate that 209 students will participate in the project every fall semester, and 189 students will participate every spring semester over the two years. Students enrolled in each class will be randomly assigned into experimental groups and control groups. The student team is the unit of analysis. Table 7 presents the detailed estimated sample randomization for a given fall semester.

Table 7. Student Team Level Sample in a Given Fall Semester

Institution & Course	std. #	exp.grp	ctl.grp	exp.team	ctl. team
UI-Mechanical	50	25	25	4 teams, ~6 std/team	4 teams, ~6 std/team
UI - Interdisciplinary	50	25	25	4 teams,~ 6 std/team	4 teams, ~6 std/team
NU - Statics	24	12	12	2 teams, 6 std/team	2 teams, 6 std/team
NU - Steel Design	15	10	5	2 teams, 5 std/team	1 team, 5 std/team
WSU - Capstone Design Section 1	35	18	17	3 teams, 6 std/team	3 teams, ~6 std/team
WSU - Capstone Design Section 2	35	18	17	3 teams, 6 std/team	3 teams, ~6 std/team
Total	209	108	101	18 teams	17 teams

Note. std. refers to student. exp.grp refers to experimental group. ctl.grp refers to control group. exp. team refers to experimental team. ctl.team refers to control team. In spring semester, the only anticipated change is the number of students enrolled in the UI – Mechanical course, from 50 to 30. Thus given there are 5 students per team, there will be 3 experimental teams and 3 control teams. The total number will change accordingly as well.

Establishing Reliability and Validity

Reliability is the extent to which a team's scores are "consistent across repetitions of the same assessment"[20]. In other words, would the team's score be the same if the team were to take the performance assessment at a later date, using a parallel form (i.e., a variation) of the performance task and if scored by different raters? Validity refers to the accuracy of the rater's inferences based on interpretation of the assessment scores[20,25]. In other words, will decisions based upon how well the team achieved a given set of outcomes be valid?

Establishing reliability is a prerequisite for establishing validity[25]. The project assessment team made up of the PI/and 4 co-PIs will evaluate the student performances using the EPS Rubric. A consensus estimate approach will be used to estimate interrater percentage agreement, also called consensus estimate[26]. It is based on the assumption that raters should be able to come to exact or near-exact (i.e., within one point, not straddling the cut score) agreement about how to apply a scoring rubric's levels to the observed performances. If two raters come to exact or near-exact agreement, then one can say that they share a common interpretation of a given construct in the rubric[26]. Over the last four years at WSU, we have used a near-exact agreement estimate as defined by Educational Testing Services[27], where raters score within one point of one another after having participated in a norming session, and with the inclusion of anchor performances in the operational rating process. In general, strong inter rater agreement has been observed in previous years. Table 8 presents the adjacent percentage agreement among raters for all the programs from year 2007 to 2009. Results of 2010 are not presented due to the aforementioned reason.

There are three types of evidence that will be examined to support the validity of the EPS Rubric. They are content, construct, and criterion. The questions and concerns that the PI rating team will be asked to consider during each official round of rating are listed in Table 9[25]. The Advisory Board will be asked to focus primarily on those questions addressing content, concurrent and predictive criterion validity. Establishing criterion validity requires that a given performance be assessed using two or more measurement instruments: a) the instrument to be validated and b) other instruments with established validity[28]. Concurrent criterion validity will be established by correlating the EPS Rubric scores with the four established instrument scores. In addition, a program-level team faculty rating group and at least one professional engineer will rate a sampling of student team performances from each institution using the EPS Rubric.

Table 8. Inter Rater Percentage Agreement of EPSA in 2007-2009

	% Adjacent + Exact Agreement		
Program	Year 2007	Year 2008	Year 2009
Bio Engineering	100%	80%	20%
Chemical Engineering	80%	100%	60%
Civil Engineering	100%	100%	100%
Computer Science	100%	100%	100%
Electrical and Computer Engineering (2008)	N/A	80%	80%
Computer Engineering (2007)	100%	N/A	N/A
Electrical Engineering (2007)	100%	N/A	N/A
Materials Science Engineering	100%	100%	100%
Mechanical Engineering	80%	100%	60%

Note. Computer Engineering program and Electrical Engineering program are combined into one program since year 2008. Score differences within one point count as adjacent agreement.

While we will not gather evidence using measurement instruments for predictive criterion validity, we plan to benefit from our Advisory Board's ample professional knowledge to obtain their perceptions on how well the performance assessment elicits and identifies criteria that would indicate successful performance in the 21st century professional engineering environment. This is critical, given the nature of the instrument, as we posit that high scores on the EPS Rubric will suggest high performance in engineering professional skills in the workplace.

Table 9. Questions to Examine Validity Evidence (adapted from Moskal & Leyden[25])

Content	Construct	Criterion
• Do the scoring rubric's criteria address any extraneous content? • Do the scoring rubric's criteria address all aspects of the intended content? • Is there any content addressed in the task that should be evaluated through scoring rubric's criteria, but is not?	• Are all of the important facets of the intended construct evaluated through the scoring rubric's criteria? • Are any of the scoring rubric's criteria irrelevant to the construct of interest?	• How do the scoring rubric's criteria reflect competencies that suggest success on related or future performances? • What are the important components of related or future performances that may be evaluated through the use of the scoring rubric? • How do the scoring rubric's criteria measure the important components of related or future performances? • Which facets of related or future performances are not reflected in the scoring rubric's criteria?

<u>Data Analysis</u>

Our approach to data analysis will show that we relied on all relevant evidence, dealt adequately with all conflicting interpretations, fully addressed the study's research questions, and successfully used investigators' prior expert knowledge as well as the project's Advisory Board's expertise.

We will use pattern matching and statistical analytic techniques to establish the performance assessment's reliability and validity. Pattern matching establishes a detailed set of predictions before the case study is carried out. Because pattern matching is a type of theory testing analysis, the predictions originate from the case study's theoretical proposition. Once a particular pattern has been predicted, investigators can then determine if the case or set of cases match the predicted pattern. The theoretical proposition is confirmed if the case matches the pattern. If the pattern is not matched, then modification of the theory and/or further investigation is required[14]. The complexity of the pattern matching generally varies in relation to the number of independent and dependent variables included in the predicted pattern(s)[22].

Within-case pattern analysis will be conducted prior to cross-case pattern analysis. Tables 10 and 11 illustrates both the within and cross-case pattern analyses. The purpose of within-case analysis is to identify unique patterns within the data for that single case. During the cross-case pattern analysis, investigators compare cases for similarities and differences. Then, they divide the data by type across all cases. This allows investigators to look at the data in different ways so that premature conclusions are not reached. When a pattern from one data type is corroborated by the evidence from another, the findings are stronger. Certain evidence may emerge as being in conflict with the predicted pattern(s). Should that occur, follow-up measures to confirm or correct the initial data should be conducted in order to tie the evidence to the findings and to state relationships in answer to the research questions[30,22].

Table 10.Within-Case Pattern Matching Analysis

Case 1	Example case with 30 students in 6 teams (teams = unit of analysis). All teams rated by same 5 raters.					
	Experimental			Control		
	Team A	Team B	Team C	Team D	Team E	Team F
EPS Rubric	3f	3f	3f	3f	3f	3f
	3dg	3dg	3dg	3dg	3dg	3dg
	3h	3h	3h	3h	3h	3h
	3i	3i	3i	3i	3i	3i
	3j	3j	3j	3j	3j	3j
External	Ethics	Ethics	Ethics	Ethics	Ethics	Ethics
	E-FSSE	E-FSSE	E-FSSE	E-FSSE	E-FSSE	E-FSSE
	AACU	AACU	AACU	AACU	AACU	AACU
	ASCE	ASCE	ASCE	ASCE	ASCE	ASCE

Table 11.Cross-Case Pattern Matching Analysis

Case 1 Predicted Pattern			INDEPENDENT VARIABLES		
			Institution		
			Course Name		
			Level		
			Semester		
			No. of teams		
			Instructor		
DEPENDENT VARIABLES	*EPS Rubric*		EXPRMT	H_0	CNTRL
		3f	**higher**		lower
		3dg	**higher**		lower
		3h	**higher**		lower
		3i	**higher**		lower
		3j	**higher**		lower
	External	Ethics			
		E-FSSE			
		AACU			
		ASCE			

Case 2 Predicted Pattern			INDEPENDENT VARIABLES		
			Institution		
			Course Name		
			Level		
			Semester		
			No. of teams		
			Instructor		
DEPENDENT VARIABLES	*EPS Rubric*		EXPRMT	H_0	CNTRL
		3f	**higher**		lower
		3dg	**higher**		lower
		3h	**higher**		lower
		3i	**higher**		lower
		3j	**higher**		lower
	External	Ethics			
		E-FSSE			
		AACU			
		ASCE			

Although the percentage agreement will remain as an estimation of interrater agreement in this project, such estimation by itself is not adequate for robust estimates of interrater reliability[30], as it does not take the likelihood of chance agreement into account. This drawback is more severe when there is a low degree of freedom of rating categories, which is the case in this study. Thus, in addition to the percentage agreement, quadratic weighted kappa is calculated as another indicator of the level of agreement between raters, which takes into consideration of rating the same category by chance. We will also use a recently developed index: $\alpha_{WG(l)}$by Brown and Hauenstein[5]. This index overcomes limitations that previous indices have, and the advantages also fit the design and purpose of this project. For example, equal $\alpha_{wg(1)}$values will represent the same level of consensus, regardless of the variation of the number of raters and the variation of location of the observed means[31].

Different statistical analyses also will be performed to complement and strengthen the pattern matching analysis: (a) a generalizability study will be conducted during the first year of the implementation, in order to determine the rater and task effects on teams' observed scores from both the EPS rubric and other measurement instruments; (b) Analysis of variance techniques will be conducted in accordance with the randomized complete block design in order to determine the effects of performance task intervention, in other words, whether there are significant differences between experimental and control groups; (c) correlation coefficient will be used to determine the criterion validity of the EPS Rubric; (d) task difficulty will be analyzed using the formula $p = \bar{X}/X_{max}$ on both experimental and control groups, where $\bar{X}$ is the average score for a group, and X_{max} is the maximum score a group can receive. Task difficulty index will be compared among different tasks. Coupled with generalizability study, this analysis will help us to determine the accuracy of team scores from EPS Rubric.

Conclusion

This century's technology-driven economy and change-driven society needs nimble engineers: creative problem solvers who can cross cultural, disciplinary and geopolitical boundaries with confidence. Engineering education in the United States is still short of preparing engineers to address the complex, globally scoped ill-structured problems the world faces.

A robust technical solution to a contemporary engineering problem must include rigorous consideration of human and environmental impacts and interactions. The 2010 BP oil disaster is a prime example of overtly ignoring potential impacts of technical solutions to a narrowly defined problem, prioritization not based on ethical considerations, not having conducted an adequate number of use-case scenarios prior to implementing drilling, and most tragically not being able to use modern engineering tools and skills to fix the subsequent problems in a timely manner. This event highlights how an engineering failure can cross geopolitical, economical, national, and disciplinary boundaries.

Engineering curricula and the corresponding required learning outcomes must be updated to include early and continued exposure to environmental, political and social issues and their international and historical contexts, legal and ethical implications of engineering solutions, as well as how to generate and harness collective innovation using current technology.

Bibliography

1. Duderstadt, J. J. (2008*). Engineering for a changing world: A roadmap to the future of engineering practice, research, and education*. The Millennium Project.The University of Michigan.

2. Felder, R. M. & Brent, R. (2003). Designing and teaching courses to satisfy the ABET engineeringcriteria. **Journal of Engineering Education**, 92(1), p. 7-25.

3. Sheppard, S. D., Macatangay, K., Colby, A., & Sullivan, W. M. (2009). *Educating engineers*. San Francisco: Jossey-Bass.

4. Shuman, L. J., Besterfield-Sacre, M., &McGourty, J. (2005). The ABET "Professional skills"- Can they be taught? Can they be assessed? **Journal of Engineering Education**, 94(1), p. 41-55.

5. AterKranov, A., Hauser, C., Olsen, R. G, & Girardeau, L. (2008, June). *A Direct Method for Teaching and Assessing Professional Skills in Engineering Programs.* **Proceedings from the American Society for Engineering Education Annual Conference and Exposition**, Pittsburgh, PA.

6. National Science Foundation. (1995). *Restructuring engineering education: A focus on change.* Division of Undergraduate Education, Directorate for Education and Human Resources.

7. Almgren, R. (2008). *Perspectives from industry*. **Journal of Engineering Education**, 97(3), p. 241-244.

8. Grasso, D. &Martinelli, D. (2007). Holistic engineering.**The Chronicle of Higher Education**, 53(28), B8.

9. Tryggvason, G. &Apelian, D. (2006). Re-engineering engineering education for the challenges of the 21st century. **Journal of Materials**, 57(10), p. 14-17.

10. Vest, C. M. (2008). Context and challenge for twenty-first century engineering education. **Journal of Engineering Education**, 97(3), p. 235-236.

11. Shuman, L. J., Sindelar, M. F., Besterfield-Sacre, M., Wolfe, H., Pinkus, R. L., Miller, R. L., Olds, B.,&Mitcham, C. (2004). *Can our students recognize and resolve ethical dilemmas?* **Proceedings from the2004 American Society for Engineering Education Annual Conference and Exposition**, Salt Lake City, Utah.

12. American Society of Civil Engineers. (2008). *Civil Engineering Body of Knowledge for the 21st Century: Prepare the Civil Engineer for the Future* (2nd ed.). Reston, VA: American Society of Civil Engineers.

13. Cady, E. T., Fortenberry, N. L., Drewery, M. P., &Bjorklund, S. (2009). *Development and validation ofsurveys measuring student engagement in engineering.* **Proceedings from the Research in Engineering Education Symposium**, Palm Cove, Queensland, Australia.

14 Davis, D., Beyerlein, S., Thompson, P. Harrison, O., & Trevisan, M. (2009). *Assessments for Capstone Engineering Design*. Transferable Integrated Design Education (TIDEE) funded by National Science Foundation. Retrieved February 4, 2009 from http://tidee.org.

15. Loughry, M. L., Ohland, M. W., & Moore, D. D. (2007). Development of a theory-based assessment of team member effectiveness. **Educational and Psychological Measurement**, 67, p. 505-524.

16. Mourtos, N. J. (2003). *Defining, teaching, and assessing lifelong learning skills.* **Proceedings from the 33rd ASEE/IEEE Frontiers in Education Conference**, Boulder, CO.

17. McMartin, F., McKenna, A., &Youssefi, K. (2000). Scenario assignments as assessment tools for undergraduate engineering education. **IEEE Transactions on Education**, 43(2), p. 111-119.

18. Sindelar, M., Shuman, L., Besterfield-Sacre, M., Miller, R. L., Mitcham, C., Olds, B., & Wolfe, H. (2003). *Assessing engineering students' abilities to resolve ethical dilemmas*. **Proceedings from the 33rd ASEE/IEEE Frontiers in Education Conference**, Boulder, CO.

19. Stein, B., Haynes, A., Redding, M., Harris, K., Tylka, M., &Lisic, E. (2009). *Faculty driven assessment of critical thinking: National dissemination of the CAT instrument.* **Proceedings from the 2009 International Joint Conferences on Computer, Information, and Systems Sciences and Engineering,** Bridgeport, CT.

20. Johnson, R. L, Penny, A. J., & Gordon, B. (2009). *Assessing Performance: Designing, Scoring, and Validating Performance Tasks.* New York: The Guilford Press.

21. Noor, K. B. M. (2008). Case study: A strategic research methodology. **American Journal of Applied Sciences**, 5, p. 1602-1604.

22. Yin, R. K. (2009). *Case Study Research: Design and Methods (4th ed.).* Newbury Park, CA: Sage Publication, Inc.

23. Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods.*Thousand Oaks, CA: Sage Publications, Inc.

24. Association of American Colleges and Universities. (2010). *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*. Washington DC: Association of American Colleges and Universities.

25. Moskal, B. M. &Leydens, J. A. (2000). *Scoring rubric development: Validity and reliability*.**Practical Assessment, Research and Evaluation***,* 7 (10).Retrieved March 17, 2010 from http://PAREonline.net/getv=7&n=10.

26. Stemler, S. E. (2004). *A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability*. **Practical Assessment, Research and Evaluation***, 9* (4).Retrieved March
17, 2010 from http://PAREonline.net/getvn.asp?v=9&n=4.

27. Educational Testing Services. (2007). *Guide for Coordinating Scoring Sessions.* Author: ETS.

28. Fortune**,** A. E. &Reid, W. J. (1999). *Research In Social Work(3rd ed.).* New York: Columbia University Press.

29. Soy, S. K. (1997). *The case study as a research method*.Unpublished paper, University of Texas at Austin.http://www.ischool.utexas.edu/~ssoy/usesusers/l391d1b.htm.

30. Lombard, M., Snyder-Duch, J., &Campanella, B. C. (2008). *Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects.*Retrieved May 3, 2010 from http://astro.temple.edu/~lombard/reliability/index_print.html.

31. Brown, R. D. &Hauenstein, N. M.A. (2005). Interrater agreement reconsidered: An alternative to the r_{WG} Indices. **Organizational research methods**, 8, p. 165-184.

Appendix A: Engineering Professional Skills Rubric (Version 2010)

ABET Skill 3f Understanding of professional and ethical responsibility	1 - Absent	2 - Emerging	3 - Developing	4 - Competent	5 - Effective	6 - Mastering
Issue ID/ Resolution	Students do not identify or summarize the issue.		Students begin to frame the issue, although some key details are glossed over. Students discuss one or more approaches to resolve the issue.		Students clearly frame the professional challenge and embedded issues. Students develop appropriate, concrete approaches to resolve the issue.	
Ethics	Students do not identify related ethical considerations.		Students show some recognition of relevant ethical considerations, but don't adequately address them in proposed approaches to resolve the issue.		Students clearly identify relevant ethical considerations and address them in proposed approaches to resolve the issue.	
Stakeholders	Students do not consider stakeholder positions on the issue, focusing only on their own perspectives.		Students may consider perspectives of one or more stakeholders, but do not discuss how they might communicate with these parties.		Students thoughtfully consider perspectives of diverse stakeholders. Students discuss how they might communicate with these parties.	

ABET Skill 3d/g Ability to communicate effectively in group and team settings	1 - Absent	2 - Emerging	3 - Developing	4 - Competent	5 - Effective	6 - Mastering
Group Discussion Skills	Students pose individual opinions and do not build on other student's ideas. Students ignore assumptions or biases underlying the issue. Some students may monopolize or become argumentative.		Students occasionally build on each other's ideas. Students briefly discuss assumptions or biases underlying the problem. Students attempt to share the floor, although this may not always be successful.		Students collaboratively build on other students' ideas to form a team approach. Students deeply examine the biases and assumptions underlying the problem. Students share the floor and encourage participation of all team members.	

ABET Skill 3h Understanding the impact of engineering solutions in global, economic, environmental, and cultural/societal contexts.	1 - Absent	2 - Emerging	3 - Developing	4 - Competent	5 - Effective	6 - Mastering
Context/Impact	Students do not relate approaches to relevant contexts. Students do not consider the impact of contexts on the issue.		Students discuss the impact of their approaches on 1 or 2 relevant contexts. Students briefly consider the impact of contexts on the issue and/or proposed solutions.		Students deeply examine the impact of their approaches on relevant contexts. Students deeply examine the impact of contexts on the issue and/or proposed solutions	

ABET Skill 3i Recognition of the need for and ability to engage in life-long learning.	1 - Absent	2 - Emerging	3 - Developing	4 - Competent	5 - Effective	6 - Mastering
Sources/Knowledge	Students do not consider outside sources of data/evidence, or these sources are seen as irrelevant to the topic. Students do not identify what they still need to know. Students do not recognize inherent biases or assumptions in sources.		Students acknowledge outside sources, and some ability to discern fact from opinion. Students identify what they don't know as well as what they do know. Students briefly address inherent biases or assumptions in one or two sources.		Students seek and evaluate outside sources (possibly including personal experience). Students identify what they still need to know. Students discuss inherent biases or assumptions in several sources.	

ABET Skill 3j Knowledge of contemporary issues.	1 - Absent	2 - Emerging	3 - Developing	4 - Competent	5 - Effective	6 - Mastering
Contemp. Issues	Students do not consider any contemporary issues.		Students show some recognition of contemporary issues and how they might relate to their identified approaches.		Students clearly understand the import of considering contemporary issues and address them in their approaches.	

Appendix B. EPSA Discussion Instruction Prompts and Scenario Examples

Engineering Professional Skills Assessment Discussion Instructions

Imagine that you are a team of engineers working together for a company/organization on the problem/s raised in the scenario below. Discuss what your team would need to take into consideration to address the problems/s. You do not need to suggest specific technical solutions, but try to come to a consensus on what is most important, and discuss stakeholders, constraints, impacts and important unknowns. Address each of the questions below:

- What are the primary problems raised by the scenario?
- Who are the stakeholders and what are their perspectives?
- Consider the following contexts: economic, environmental, cultural/societal, and global. What are the potential impacts of existing or possible engineering solutions on each of these contexts?
- What are some unknowns that seem critical to know to address the identified issues?
- Who [or what] would you need to consult outside your engineering team to best address the problems you've identified? What indicators would tell you that these are valid resources?
- What biases or assumptions do you need to consider in your interpretation of the information provided by the sources given for the scenario (as well as those you consulted on your own)?

Scenario example 1

Lithium mining for lithium-ion electrical vehicle batteries

The US government is investing heavily in sustainable resource research and development in order to decrease national oil consumption, and automotive industries around the world are competing in a global race for "sustainable mobility". There were about 52 million total vehicles produced in the world in 2009, and replacing a significant amount of them with highly electrified vehicles poses a major challenge. The state of California is targeting 1 million electric vehicles (EVs) on its streets by 2020. By that same date, Nissan forecasts that EVs will become 10% of all global sales.

Battery technology is currently the major bottleneck in EV design. In 2009, President Obama announced $2.4 billion in grants to accelerate the manufacturing and deployment of next generation batteries and EVs. Lithium-ion batteries are the first choice for the emerging EV generation, (the Chevy Volt, the Volvo C30, the Nissan Leaf), because they feature high power density, manageable operating temperatures, and are relatively easy to recharge on the grid.
In spite of its potential, lithium may not be the answer to the EV battery challenge. Lithium, which is recovered from lithium carbonate (Li2CO3), is not an unlimited resource. Lithium-based batteries are already used in almost all portable computers, cell phones and small

appliances. Utility-scale lithium-based energy storage devices are in the works for smart grid applications, such as balancing energy supply-demand fluctuations. Lithium is also extensively used in a number of processes we take for granted: the manufacturing of glass, grits, greases and aluminum, among others. This makes accurate estimations of future demand in relation to resource availability almost impossible.

How much lithium is needed to power an electric vehicle?

Energy requirements………………………..16 kilowatt hours (specified for Chevy Volt)
Lithium estimates per kWh……………….0.431 kg (US Department of Transportation estimate)
Total lithium for one Chevy Volt……….6.86 kg
Total Li_2CO_3 for one Chevy Volt .... 36.5 kg
Total Li_2CO_3 one million PHEVs ....36,500 metric tons

According to Meridian International Research, an independent renewable-energy think tank, there is insufficient recoverable lithium in the earth's crust to sustain electric vehicle manufacture based on Li-ion batteries in the volumes required by the mass market. Lithium depletion rates would exceed current oil depletion rates, potentially switching dependency from one diminishing resource to another. The United States Geological Survey reports that the Salar de Uyuni salt pans of Bolivia contain the largest untapped reserve of lithium in the world – an estimated 5.4 million metric tons or almost 50% of the global lithium reserve base. Other estimates put the Bolivian resource as high as 9 million metric tons. Bolivian president, Evo Morales, has consistently rejected bids by Mitsubishi and Toyota to mine lithium in his country and has announced plans to develop a state-controlled lithium mining operation. Prices of lithium carbonate (Li_2CO_3) have more than doubled since 2004. Lithium batteries are costly, too; battery packs for vehicles cost upwards of $20,000 alone, driving up the overall cost. Lithium CAN be recycled, but there is little existing infrastructure. In 2009, a California company, Toxco Waste Management, received $9.5 million in grants from the US Department of Energy to help build the first US-based facility for recycling lithium batteries in anticipation of demand.

<u>Sources</u>

Lithium Dreams: Can Bolivia Become the Saudi Arabia of the Electric-Car Era? (March 22, 2010). *The New Yorker*.
Lithium Largesse?(August 2009). *American Ceramic Society Bulletin*.
US Department of Energy, Press Release. (August 5, 2009)
Bolivia's Lithium Mining Dilemma. (September 8, 2008) *BBC News*.
The Trouble with Lithium: Implications of Future PHEV Production for Lithium Demand. (2007). *Meridian International Research*.

Scenario example 2

<u>BP Deepwater Horizon Oil Spill</u>

According to estimates made in early August 2010 by the US federal Flow Rate Technical Group, the BP Deepwater Horizon oil spill is the largest unintended discharge of oil into marine waters. A key emergency disconnect system failure was noted in a US House of Representatives Committee on Energy and Commerce memo summarizing the early investigation. A working automatic closure system (i.e., the "deadman" switch), would have stopped the oil from flowing. But because of the explosion, signals may not have reached the blowout preventer (BOP). The overall BOP technology maintenance and testing is also under investigation. While the BOP is often referred to as a component, it is actually a system of many complicated systems and components controlled automatically by hardware and software, or controlled remotely by humans.

BP has reported concerns about multiple additional failures related to the BOP: its automated mode function, its shearing functions, and the remote operated vehicle interventions. According to BP, "there were multiple control mechanisms— procedures and equipment—in place that should have prevented this accident or reduced the impact of the spill." The BP response team is now working on a number of efforts to advance work flow, improve coordination, focus efforts and manage risks.

What was the status of the oil in late July 2010?

- approximately 4.9 million barrels of oil had leaked
- about 17 percent (800 million barrels) had been contained
- of the remaining 4.1 million barrels of oil, more than half had been burned, skimmed, evaporated or dispersed
- about 1.3 million barrels of oil had formed tar balls and were covered by sediment and sand, or was floating as a light sheen on the ocean surface

BP has launched two unmanned robotic vehicles to monitor pollutants. The Wave Glider robots operate autonomously with redirecting possible by satellite, receive propulsion power from the waves and use solar power. BP will use the robots to collect data on water quality, including oil that is emulsified, dissolved or dispersed, as well as plankton and oxygen matter. Data collected by the vehicles will be relayed via satellite to BP's control center in Houston.

The BP internal investigative report released September 8, 2010 states: "The [BP investigative] team did not identify any single action or inaction that caused this accident. Rather, a complex and interlinked series of mechanical failures, human judgments, engineering design, operational implementation and team interfaces came together to allow the initiation and escalation of the accident. Multiple companies, work teams and circumstances were involved over time."

Sources
"BP blames oil spill on 8 key failures." (September 9, 2010). *The Financial Post*.
BP Deepwater Horizon Investigative Report. (September 8, 2010). BP Oil Internal Investigation Team.
"BP oil spill robots to report on water pollution."(September 3, 2010). *Computerworld UK*.
"BP: Oil spill taught cloud, workflow management lessons." (September 8, 2010). *Computerworld UK.*
"Memorandum to the Subcommittee on Oversight and Investigations on Key Questions Arising from Inquiry into the Deepwater Horizon Gulf of Mexico Oil Spill."(May 25, 2010). *Congress of the United States, House of Representatives' Committee on Energy and Commerce.*
"Oil giant expected to release details of internal report into accident." (September 3, 2010). *Computerworld UK*.
"Tracking the Oil Spill in the Gulf Deepwater Horizon Oil Spill." (August 10, 2010). *New York Times*.

Scenario example 3

The Need for Prosthetics for Land Mine Victims in Iraq

Each year, thousands of Iraqis and US military personnel lose limbs due to frequent land mines explosions. Landmines are designed to maim, not kill; victims often require extensive medical care including amputations and subsequent prosthetics. The Mosel-based office of the Red Crescent Society, the Islamic affiliate of the Red Cross, estimates that 3,000 additional replacement limbs are needed annually in the northern region alone. For Iraqis, financial constraints limit the rate of advancement in prosthetic rehabilitation, and it is a challenge to find funding for widespread application of affordable prosthetic innovations.Because US soldiers have also suffered from amputations, the US government has invested millions, resulting in great advances in prosthetic technology since the start of the war. However, the primary beneficiaries have been US soldiers.

A high proportion of car bombs and roadside explosions characterize the Iraq war, as do injuries to civilians caused during US air strikes. In 2009, air strikes were five times more frequent than they were in 2006. The Ministry of Health in Iraq estimates that there are "approximately 80,000 amputees of whom some 75 to 85 percent reportedly were caused by mines or unexploded ordinance." Many are women and children. Recently, the leaders of the Basra Iraq Prosthetics Project hypothesized that it would take 20 to 30 years to fully and adequately care medically for the current survivors in Iraq.

Common Materials Used to Make Prosthetics

- Plastic polymer laminates (e.g., acrylic, polyester, epoxy, polypropylene) for prosthetic socket fabrication
- Silicon elastomers used as barriers between the human socket and the prosthetic
- Carbon fiber composites used in artificial limbs
- Titanium implants to the bone
- Aluminum alloys

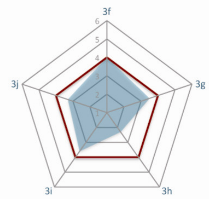
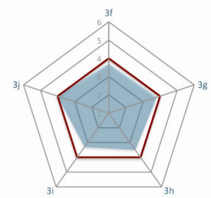
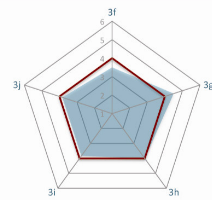
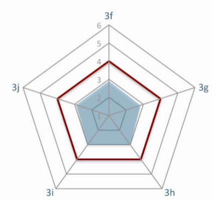
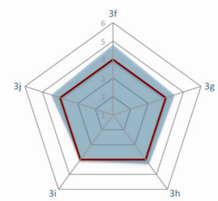
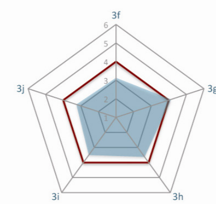
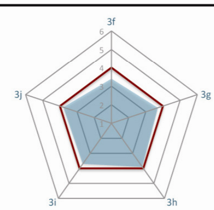
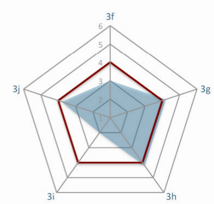
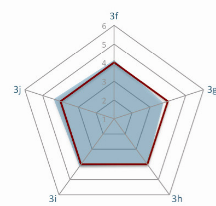
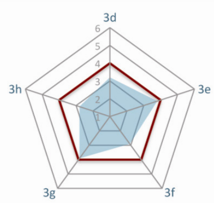
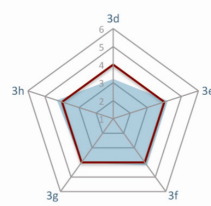
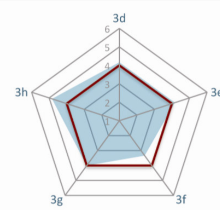
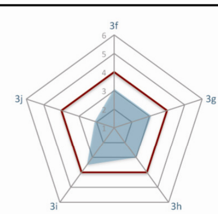
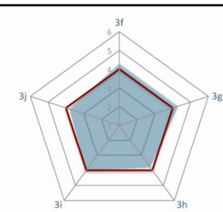
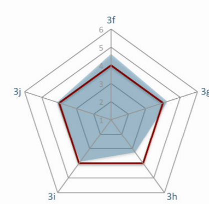
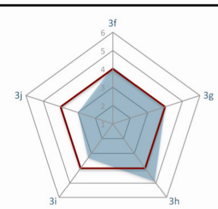
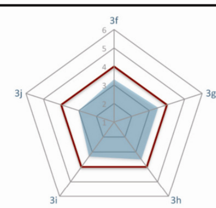
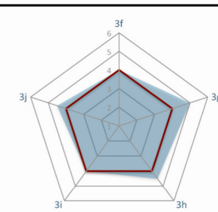
Being an amputee in Iraq has very different implications than it has in the developed world. Because it is not a wheelchair friendly country with wheelchair accessibility infrastructure, such as paved roads, becoming disabled in this nation means a total loss of independence for the survivor. Once you are an amputee in Iraq, “you’ve lost your mobility; you’ve lost your future,” according to Linda Smythe, head of the Prosthetics Project. With no way to travel, survivors cannot support themselves and their families. Also, since high numbers of these amputees are children, they lose access to education in addition to their mobility. A child’s prosthesis must be replaced every 6 months; an adult’s every 3 to 5 years. Girls with amputations are frequently considered a burden for their families, as they become ineligible for marriage.

Iraq lacks enough medical centers to treat its disabled. Many health-care centers have been destroyed and others have been forced to shut down. The remaining centers are undersupplied and do not have sufficient orthopedic doctors or specialists trained in orthotics or prosthetics to help the overwhelming number amputees regain their mobility.

<u>Sources</u>

Land Mine Monitor Fact Sheet. (June 2009). *International Campaign to Ban Landmines*.
Basra, Iraq Prosthetics Project.(Summer 2008).*Journal of Mine Action*.
Iraq’s Prosthetics Crisis. (July 29, 2007). *hc2D.co.uk Virtually Comprehensive Healthcare News*

Appendix C. Radar Graph Representation of Results of EPSA on the Program Level from 2007 to 2009

Program	2007	2008	2009
Bioengineering 2009 Avg = 3.9			
Chemical Engineering 2009 Avg = 3.5			
Civil Engineering 2009 Avg = 4.1			
Computer Science 2009 Avg = 4.1			
Electrical and Computer Engineering 2009 Avg = 4.0			
Materials Science and Engineering 2009 Avg = 4.3			
Mechanical Engineering 2009 Avg = 4.3