Formative vs Summative ABET Assessment: A Comprehensive Graphic Representation for A New BME Program

Dr. Davide Piovesan, Gannon University

Davide Piovesan was born in Venice, Italy on October 10 , 1978. He is currently Assistant Professor in the Mechanical Engineering department at Gannon University and the director of the Biomedical Engineering Program. He received his M.S.M.E in 2003 and D.Eng in Mechanical Measurement in 2007 at the University of Padova, Italy. His dissertation presented a set of experimental and analytical validation techniques for human upper limb models. From 2004 to 2008 he was a visiting scholar and post-doctoral fellow at the Ashton Graybiel Spatial Orientation Lab at Brandeis University. There, he worked on the mechanics of movement adaptation in non inertial environments as part of a NASA extramural funding program. He joined Northwestern University in 2008, working as a post-doc fellow at the Rehabilitation Institute of Chicago until 2013 in the field of rehabilitation robotics. Davide's main research interest is to gain insights on the role of biomechanics in the neural control of movements, with applications to rehabilitation engineering.

Dr. Karinna M Vernaza, Gannon University

Dr. Karinna Vernaza joined Gannon University in 2003, and she is currently a Professor in the Mechanical Engineering Department and Associate Dean of the College of Engineering and Business. She earned her Ph.D. and M.S. in mechanical engineering from the University of Notre Dame. Her B.S. is in Marine Systems Engineering from the U.S. Merchant Marine Academy. Her primary teaching responsibilities are in the solid mechanics and materials areas, including biomaterials. She was awarded the 2012 ASEE NCS Outstanding Teacher Award, 2013 Gannon University Distinguished Faculty Award and 2013-2014 Gannon University Faculty Award for Excellence in Service-Learning. Vernaza does research in the area of alternative fuels (biodiesel), engineering education (active learning techniques), and high-strain deformation of materials. She is currently the PI of an NSF S-STEM and ADVANCE-PAID grants.

Formative vs Summative ABET Assessment: A Comprehensive Graphic Representation for A New BME Program

Abstract

This paper presents an innovative methodology for the assessment of a new Biomedical Engineering (BME) program. Biomedical engineering programs are quite new in the engineering educational system. To date, in the state of Pennsylvania there are only six programs that have been accredited by ABET out of the 91 programs accredited nationwide[1]. While the guidelines of ABET are quite general and applicable to a variety of programs with different focuses, the scarcity of published data on BME specific programs poses a significant challenge on the preparation and assessment of program specific requirements. Another rather significant challenge is the limited number of students graduating from a brand new program that does not give sufficient statistical power to confirm the reliability of the assessment process.

To obviate these limitations, an analysis for the assessment process has been designed that can help understand if any improvement occurs from the freshman and sophomore year (formative assessment), where concepts are introduced, to junior and senior years (summative assessment), where the concepts are reinforced/mastered and assessed again. By having a formative and summative assessment it is possible to evaluate if improvement occurs within the cohort, allowing for the redaction of a continuous improvement plan.

The goal of the program has been set to have all students at a satisfactory level at the time of graduation. Having a rubric calibrated at "senior level" (i.e. master level), we found that most of the outcomes at the formative level reach a marginal outcome. This was expected as the idea is to observe if there exist learning trends between formative and summative levels where concepts are introduced, internalized and reinforced. This article presents a model that has successfully assessed student learning outcomes at *one institution* and that can be adapted by other programs as they prepare for accreditation.

1. Introduction

The Gannon University BME program has instituted a common assessment rubric for each ABET, Inc. outcome a-k adding one additional program specific outcome. It is important to notice that ABET does not mandate a specific assessment methodology. What is presented in this article is what has been effective for the institution in the last accreditation cycle. The rubrics were employed to present student learning outcomes during the Fall 2014 evaluation visit. *The same rubric is used by each professor to assess the corresponding outcome that is pertinent to his/her course.* The rubrics have a different number of performance indicators (or dimensions) to allow for a comprehensive tool that describes multiple facets of the outcome to be assessed. The performance indicators of each rubric were built in view of the performance indicators of each engineering course in the program. Each outcome specific rubric was agreed upon the faculty and calibrated on a “senior level” of intellectual maturity since ABET’s evaluation is based on attributes achieved by students upon graduation. The assignments were designed specifically to satisfy each dimension of the rubric and consisted in questions or problems presented to the students in midterms and final exams/projects. The four levels of the rubrics are: Unsatisfactory, Marginal, Satisfactory, and Outstanding.

The BME program is an integral part of the Mechanical Engineering (ME) Department which has been accredited since 1965. Hence, many courses are offered as part of the mechanical engineering curriculum. Thus, it was necessary to separate the assessment of the BME student learning outcomes from the assessment of the class. By using the program-specific rubrics, only the students enrolled in the BME program were evaluated within each class, even though the class contained a larger number of students. It is important to note that the ME program has employed several different methods of assessment throughout the years. During the last visit, rubrics were employed.

2. Formative versus Summative Assessment

The separation of Student Learning Outcomes (SLO) assessment in formative and summative poses the problem of intellectual maturity of the students. Intellectual maturity has been approximated using Perry’s levels[2] (refer to Table 1), which helps analyze in different stages of development how knowledge is perceived by the student, as well as how problem solving and the teaching/authority figures are identified. The same set of rubrics was employed for both formative and summative evaluation, assuming that such rubrics would best apply to senior students with a level of maturity between “Relativism” and “Commitment” (see Table 1). However, the level of intellectual maturity between freshmen/sophomores and juniors/seniors is different. While the latter ought to be within the level of intellectual maturity targeted by our rubrics, freshmen/sophomores are most likely to be in the “Dualism/ Multiplicity” level. Hence, it is expected that courses assessed in the formative evaluation might not meet the set criteria threshold. Nevertheless, the formative assessment is important to monitor the improvement trend of the students.

3. Assessment Tools for Student Learning Outcomes

Direct evidence is obtained from the rubrics and other tools which are utilized at specific intervals in order to assess the performance of the program with regard to student learning outcomes. The primary tools for assessment of performance on SLO are evaluated on a semester-by-semester basis, or a modified semester-by-semester basis. The results of these primary tools are used to establish an action plan for each course, (for example, small modifications to be presented in the next course iteration), and thus provide a path for continuous improvement. Secondary tools are implemented on different time scales, and provide corroborative information only. The **program assessment rubrics** are the primary tool used to provide a measure of satisfaction of student learning outcomes.

Table 1: Intellectual maturity according to Perry's levels

Stage of Intellectual Maturity (Approximate Perry Levels)	**Perceptions of Knowledge**	**Abilities to Make Commitments**	**How Solutions to Problems are Perceived**	**Perceptions of the Responsibility of Learners**	**Perceptions of the Responsibility of Experts as Teachers**
Dualism- Individuals at this level are concrete thinkers who believe things are right/wrong, we/they, good/bad.	Knowledge is a set of truths.	I have faith in, and a commitment to, truth and knowledge as it is stated by genuine authorities.	There is a single correct solution to every problem.	I receive explanations of knowledge and become uneasy when asked to think independently, draw conclusions, or give my points of view.	Experts are authorities with an ability to explain and give me correct answers.
Multiplicity- At this level they recognize that diversity in thinking exists. Uncertainty prevails because all opinions are valid.	Knowledge is a matter of educated opinion.	I feel no need to commit to any specific belief or mode of thinking.	There is no one right solution to a problem, because all are equally valid.	I listen to experts, but have a right to my own opinions.	Experts explain course material to me and express their opinions.
Relativism- When they reach this level they perceive that all knowledge is relative, and that they need to orient themselves based on evidence.	Knowledge is not universal, but a matter of context and situation. What is true in one situation may be false in another.	I feel there is a need for some form of personal commitment.	Ambiguity is part of life, so I must defend my own position on problem solutions based on evidence.	I make comparisons to distinguish between weak and strong evidence in determining knowledge.	Based on their experience, experts teach procedures and analytic methods to help me reason and compare alternatives.
Commitment- Finally, at this level they develop the need to take positions and commit to them.	Knowledge is constructed from experience, what is learned from others, and from reflective thinking.	I feel the need to make commitments, especially a personal commitment to learning.	There are many solutions to each problem; some are better, and some are worse. I must take a stand on issues based on my personal values and analysis.	I learn and I integrate new knowledge with what I already know.	Experts are mentors that challenge my assumptions to support my learning.

4. Program Assessment Rubrics

The BME program has instituted a common assessment rubric for each ABET outcome. *The same rubric is used by each professor to assess the corresponding outcome that is pertinent to his/her course.*

The rubrics have a different number of performance indicators (or dimensions) to have a comprehensive tool that describes multiple facets of the outcome to be assessed. The performance indicators of each rubric were adapted[3-4] and re-built in view of the performance indicators of each engineering course in the program. The large amount of indicators at the courses level was dramatically reduced by synthesizing a set of outcome-specific common rubrics that all faculties now use. Rubrics are reviewed each semester as part of the evaluation process after assessment. As an illustrative example, the rubric for outcome h is presented in Table 2.

Table 2: Example of program rubric rating for outcome h

numerical value	1	2	3	4
Outcome h	**Unsatisfactory**	**Marginal**	**Satisfactory**	**Outstanding**
Global impact	Narrow perspective, with consideration only of immediate impact on current users	Awareness of current and future impact on users, with some consideration of potential for broader impact on others	Knowledge of tradeoffs and constraints with respect to current and future impact on users, with some consideration of potential for broader impact on others	Comprehensive analysis and synthesis of current and future impact from a global perspective, and integration of this analysis into engineering design and problem solving.
Societal impacts	Little or no consideration of the risks and benefits to society	Awareness of the basic risks and benefits to society, with some consideration of broader societal impact	Knowledge of tradeoffs and constraints with respect to basic risks and benefits to society, and broader societal impact	Comprehensive understanding and analysis of the risks and benefits to society, and integration of this analysis into engineering design and problem solving.

In adopted rubrics, each performance indicator is used to assess specific assignments or assignment groups in multiple courses. For example, a particular exam question in one course might be included to test whether students correctly understand the historic consequences of a specific engineering failure. This can be abstracted to infer the level of student achievement in understanding the impact of engineering effort on society (student outcome h). On the other hand, in a different course, the student might write an essay on the impact of motion capture in healthcare, which addresses a different aspect of outcome h (i.e. the tradeoffs of different technologies on the specific user). Thus, each professor monitors the performance of all students on these specific assignments, and determines the achievement of each performance indicators that are part of the rubric's outcome. A course including a specific SLO can either cover all performance indicators of a specific rubric or just a few. However, within the program, all the indicators are covered for a specific rubric so it is possible to determine the overall achievement of the SLO. The results of the analysis are used to provide instant feedback to make the course better on the next iteration.

One-third of all courses are evaluated each year so that two evaluation cycles may be completed for each course within the expected period of each accreditation cycle. For new programs applying for accreditation for the first time, the number of students assessed is typically small (e.g. 1 to 7 graduates) and does not allow a clear statistical analysis of the successful result. Thus, we performed an analysis on the assessment examining the SLOs in courses that introduced the topic (formative assessment) and courses in which the topic was reinforced (summative assessment).

5. Student Learning Outcomes Assessment

An example of the analysis is presented in this section. The assessment has been obtained with the subset of courses in Table 3. The specific courses included in Table 3 are all required courses, and were selected for inclusion in order to provide a reasonable cross section of courses and to provide at *least two summative assessments for each student learning outcome*. The correlation of courses to student learning outcomes must be agreed upon by all faculty members, in consultation with one another.

Table 3 should be interpreted as follows: the top set of courses is used for formative assessment, the bottom courses are used for summative assessment. The former courses are mostly freshmen and sophomore courses (with the exception of strength of material ME214), while the latter are program specific courses which occur in the junior and senior years. Shaded boxes indicate the outcomes covered by the course. The cells that contain the number "1" are the outcomes that have been actually assessed. For each course, the instructor is tasked with creating assignments that are appropriate for the performance indicator of the chosen student learning outcomes defined in the rubric such as the one reported in Table 2. Hence, the course instructor applies the rubric that was agreed upon for each student outcome and assesses the results. Actions are then identified (by instructors) to be taken in each course, for the next time the course is taught. Improvement can then be made continuously. In order to collect the assessment data, instructors were provided with an Excel file with clear instructions. These instructions appear in Appendix A.

Notice that the summative assessment of several outcomes is performed in the senior design series (lecture and lab) and major specific courses such as biomedical system model analysis, biomechanics, and bioengineering laboratory.

Table 3: Courses used for the present formative and summative assessment

COURSE (FORMATIVE)	CODE	A	B	C	D	E	F	G	H	I	J	K	L
first year seminar	ENG100				1		1			1			
strength of material	ME214	1		1									
material science	ME315								1		1		
digital computer usage	ME205	1				1						1	
digital computer usage lab	ME206											1	
biology lab	BIOL123		1		1								1
molecular biology	BIOL122												1
engineering thermodynamics	ME312	1				1							
instrumentation lab	ME332		1					1					
assessment per outcome		3	2	1	2	2	1	1	1	1	1	2	2
COURSE (SUMMATIVE)	**CODE**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**
biomechanics	BME420	1				1			1				
biomed. system model analysis	BME430						1	1	1	1	1		1
bioengineering lab	BME440		1		1							1	1
fluid mechanics	ME336	1		1		1		1					
biomaterial	BME410							1			1		
strength of material lab	ME215		1					1				1	
senior design	ME350			1			1		1		1		
senior design lab	ME354			1	1				1	1			
assessment per outcome		2	2	3	2	2	2	4	4	2	3	2	2

As an example, the rubric used to assess the first performance indicator (i.e. dimension) of outcome a (An ability to apply knowledge of mathematics, science and engineering), is presented in Table 4. This dimension is named "A1". For each dimension the performance indicator can vary from 1 to 4. The nominal s used in the rubric as assessment were 1= unsatisfactory, 2= marginal, 3= satisfactory and 4= outstanding. Let us assume in our example that the performance indicator "A1" assessed in a specific class (e.g. Engineering Thermodynamics ME312) is about 2.5 as average among all BME students (see Table 5). The instructor assessed most of the students with either the definition "Chooses a mathematical model or scientific principle that applies to an engineering problem, but has trouble in model development," or with "Applies a mathematical model or scientific principle to an engineering problem." It can be observed that the words like "develop" and "apply" are within different levels of Bloom's Taxonomy[5]. Such words can define the level of students' proficiency in a specific topic as well as allow the instructor to systematically identify such level in the students' work.

Table 4: Example of performance indicator A1 for outcome a

numerical value	1	2	3	4
Outcome a	**Unsatisfactory**	**Marginal**	**Satisfactory**	**Outstanding**
Mathematical Modeling	Does not understand the connection between mathematical models and systems in engineering	Chooses a mathematical model or scientific principle that applies to an engineering problem, but has trouble in model development	**Applies** a mathematical model or scientific principle to an engineering problem	Combines mathematical and/or scientific principles to **formulate** system models relevant to engineering

All the data from the assessment files received from the instructors performing the assessment must be averaged within each performance indicator, among all classes that measured it. An example of this analysis for all the performance indicators for outcome a is illustrated in Table 5. The top row of Table 5 indicates the semester (F= Fall, S= Spring) and year the data was collected. Each column represents the assessment of the students in one course. Data are averaged among courses within the specific categories (i.e., either formative or summative).

Table 5: Average of performance indicators among courses

	F12	F13	F13	S14	**Formative**	F12	F13	F13	S14	**Summative**
	ME312	**ME205**	**ME214**		**Average**	**ME336**			**BME420**	**average**
A1	2.50	2.00	2.75		2.42	4.00			3.00	3.50
A2	2.50		3.00		2.75	4.00			4.00	4.00
A3	2.50		3.00		2.75	3.00			4.00	3.50
A4	2.50	2.00	4.00		2.83	4.00			4.00	4.00

6. Program Assessment Graphic Representation

Figure 1 provides a **visual representation** of the assessment of student learning outcomes for each performance indicator. Each radius of the radar plot, defined as a letter and a number, represents the average among all BME students of the performance indicator used to assess a specific student outcome. A legend for all the performance indicators is reported in Table 6. *Our goal is to have all dimensions at or above level 3= satisfactory, which is represented by the red circles in the graph.*

As mentioned above, both a formative and a summative assessment for the program was performed. The green line represents the condition of the program in courses that are mostly at the freshmen and sophomore level (formative assessment) where the concepts are introduced. The blue line represents the assessment in junior and senior level classes (summative assessment) where concepts are reinforced or mastered. *It is immediate to understand that: i) if the assessment line is outside of the threshold line (red) the performance indicator of said outcome is satisfied; ii) it is expected that the formative assessment line would be encircled in the summative assessment line.* Indeed, when the concept is introduced, some students might find it difficult to grasp, and their level of intellectual maturity might be at an early stage. However, as time passes and concepts are reinforced, a general improvement ought to be observed. Instances that require analysis are when the performance indicator is not satisfied in the summative assessment, or there exist a performance decrease from the formative to the summative assessment.

The latter instance is important to analyze for cases when classes offered within the BME program are shared with other programs, whose students might have different background.

From Figure 1 we can notice that the threshold we impose is not fully satisfied for the following performance indicators:

B5. Data Analysis and Interpretation

C3. Application of Design Constrains

C4. Design Originality

G5. Use of reference in written communications

K2. Knowledge of professional tools.

However, the performance level is between marginal and satisfactory, and never unsatisfactory. All the other 42 indicators are fully satisfied. It should be also noticed that if one performance indicator is not satisfied, the whole outcome on average might be satisfied. Indeed, we can observe that only outcome c is not satisfied most of the time in this assessment cycle.

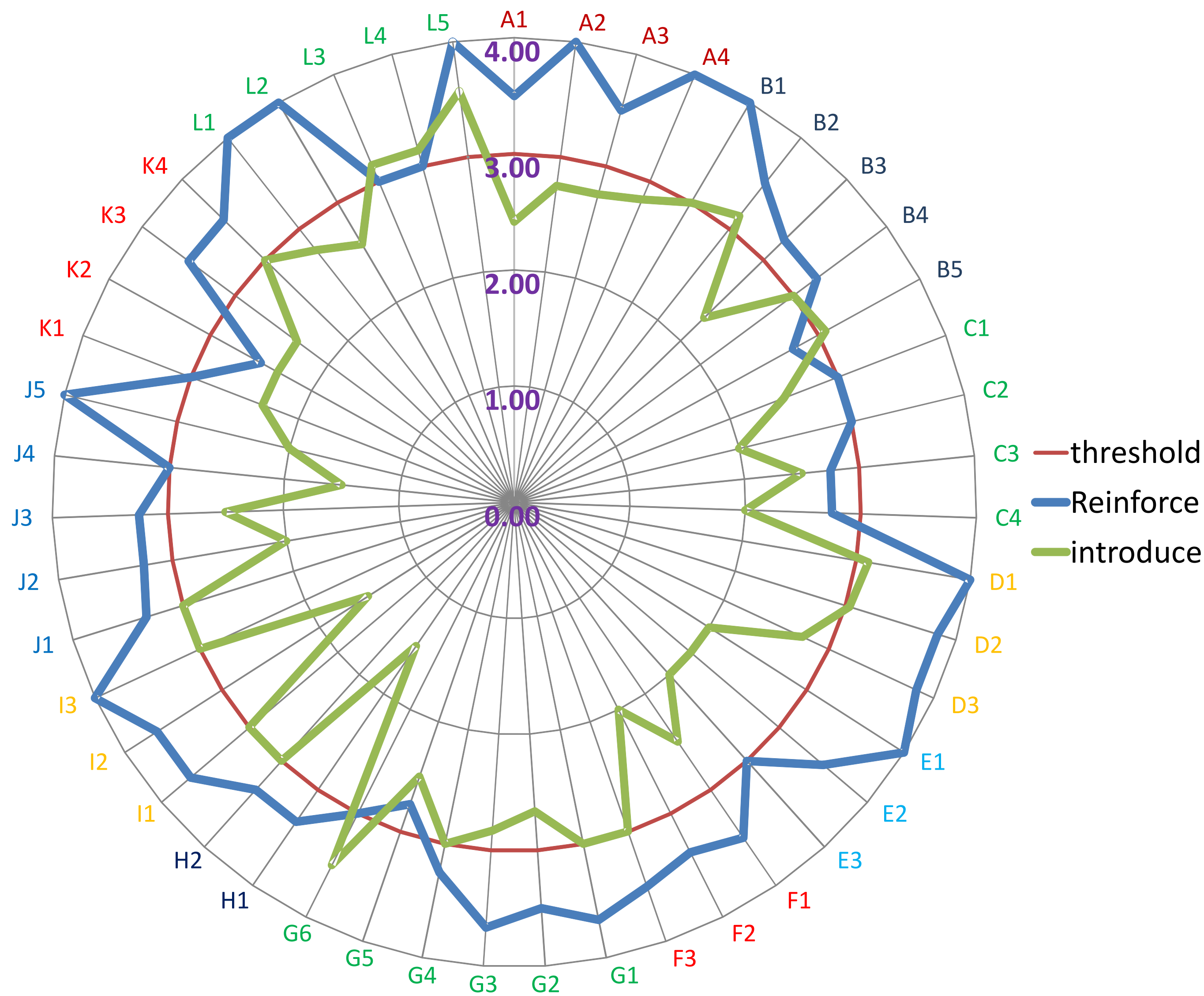

Figure 1: Assessment of Student Learning Outcomes using Rubrics

Table 6: Legend for the dimensions of the assessment graph

	A) An ability to apply knowledge of mathematics, science and engineering		**G) An ability to communicate effectively (WRITTEN)**
A1	Mathematical modeling	G1	Style
A2	Application	G2	Organization
A3	Theory	G3	Relationship of graphics/figures/tables to written material
A4	Calculation	G4	Formatting of graphics/tables/figures
	B) An ability to design and conduct experiments as well as to analyze and interpret data	G5	References
B1	Lab safety	G6	Computer code comments
B2	Experimental procedure		**H) The broad education necessary to understand impact of engineering solutions in a global, economic, environmental and societal context**
B3	Data gathering and documentation	H1	Global impact
B4	Tool selection and operation	H2	Societal impacts
B5	Data analysis and interpretation		**I) A recognition of need for, and ability to engage in life-long learning**
	C) 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	I1	Ability to find information outside of normal class sources (pursue of knowledge)
C1	Design techniques	I2	Ability to recognize need for improvement of current engineering solutions (self-motivation)
C2	Documentation and support	I3	Technical society affiliation
C3	Application of constraints		**J) A knowledge of contemporary issues**
C4	Originality	J1	Identify specific, real and current problem in need of engineering design solution
	D) An ability to function on multidisciplinary teams	J2	Awareness of how economic/social/political/technical issues affect feasibility of engineering solutions
D1	Fulfills team roles & duties	J3	Ability to communicate thoughtfully on the likely effect of new engineered solutions on current economic/social/political/technical climate
D2	Shares equally	J4	Effect of policy on engineering
D3	Listens to other teammates	J5	Ability to analyze an out of field issues using engineering tools
	E) An ability to identify, formulate, and solve engineering problems		**K) An ability to use techniques, skills, and modern engineering tools necessary for engineering practice**

E1	Identify	K1	Selection of tools
E2	Formulate	K2	Knowledge of professional tools (equipment, instruments, software)
E3	Solve	K3	Ability to develop new skills and expertise
	F) An understanding of professional and ethical responsibility	K4	Ability to find needed information/ resources
F1	Maintains professional interaction with others		**L) An ability to apply in depth knowledge of biology**
F2	Abides by the ASME Code of Ethics for Engineers	L1	Structure (anatomy)
F3	Demonstrates objectivity	L2	Function (physiology)
	G) An ability to communicate effectively (Oral)	L3	Process
G1	Mechanics	L4	Interaction (between organisms or subsystems)
G2	Organization	L5	Ecology (interaction with the environment)
G3	Delivery		
G4	Relating to audience		

An analysis can help to understand if any improvement occurs from the freshman and sophomore year, where concepts are introduced to junior and senior years where the concepts are reiterated and assessed again. In this assessment example (refer to Figure 1), consistent improvements are observed among all performance indicators except

B5. Data Analysis and Interpretation

G6. Clarity of computer code

L3. Biological processes

L4. Interaction (between organisms or subsystems)

However the variation is negligible and, as previously mentioned, the student sample is very small so it is difficult to make a statistical analysis on the data. Corrective actions should be identified to improve the performance on these performance indicators. Proposed actions for improvement should be based upon the outcome assessment rubrics.

7. Conclusion

For new programs, it is critical to design a process that provides a meaningful summative assessment. The model presented here has proven its effectiveness at assessing each defined dimension within each student learning outcome. As the program grows, this methodology will provide the basis to implement data driven changes coming from summative evaluations. In the meantime, improvements can be made in response to ongoing response to measured and student-stated achievement at the formative evaluation level. The use of the same program rubrics at the course level will identify actions to improve student learning. Overall, this method provides an effective analysis at identifying improvement trends from freshman/sophomore years to junior/senior years. Lastly, this is one program's approach to ABET assessment; it is the authors' goal to provide a model for other to adopt.

Bibliography

1. ABET, Find Accredited Programs, [Online] Available: http://main.abet.org/aps/Accreditedprogramsearch.aspx
2. Johnson, V.R. Gaining Intellectual Maturity: Becoming an Independent Learning Professional, [Online] *IEEE-USA Today's Engineer*, May 20015 Available: http://www.todaysengineer.org/2005/May/intellectual_maturity.asp#2
3. Market University, Department of Biomedical Engineering, Mission, Objectives and Outcomes, accessed August 7, 2013, http://www.marquette.edu/biomedical-engineering/mission-objectives-outcomes.php.
4. Cornell University, Electrical and Computer Engineering, [Online] Available: http://people.ece.cornell.edu/land/courses/ece4760/rubrics
5. Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. New York: David McKay Company.

Appendix A

Each instructor involved in the assessment received an Excel file containing:

- All the rubrics from outcome a-l
- Copy of Table 3, in which it is indicated which outcome(s) need to be assessed for the chosen course
- The following set of instructions:

0). Create a file for each outcome that you have to assess. Faculty, Course, and Outcomes to be assessed are in Course_Map tab.(NOTE: assess only the outcome that have "1" in the gray cell)

1). **Save file as** *Outcome_CourseCode_YourName_Year&Semester.xlsx* (e.g. A_BME420_Smith_13F.xlsx)

2). Apply rubric (described for each outcome in tabs named w/ a letter) and provide for each dimension one of the following rating 1=Unsatisfactory, 2=Marginal, 3=Satisfactory, 4=Outstanding (no decimals).

3). Use a specific question of an assignment, or a full assignment, to determine a numerical value for each dimension of a rubric

4). After choosing an assignment for assessing a dimension of the rubric use that for each student

5). The number of assignments selected will depend on how many are needed to cover all the dimensions of the rubric.

6). Each dimension of the rubric should to be assessed at least once

7). Since the maximum number of dimensions can be 6 and the minimum can be 2, If a dimension IS NOT part of the rubric, leave the cell empty and highlight the column

8). If a dimension (indicator) IS part of the rubric BUT cannot be measured, leave the cell empty.

9). Assess all biomedical students in your class, and an equal number of non-biomedical students (NOTE: For Biology only BME Student)