Measuring Continuous Improvement In Engineering Education Programs: A Graphical Approach*

Graciela de L. Perez, Larry Shuman, Harvey Wolfe and Mary Besterfield-Sacre
University of Pittsburgh

Abstract

This paper presents a method for developing assessment metrics that can be used to efficiently reduce survey data to a format that facilitates quick and accurate faculty feedback as part of an EC 2000 continuous improvement process. Our methodology, the *Pitt-SW Analysis*, is an adaptation of the competitive strategy principle of SWOT (strength, weakness, opportunities and threats). It consists of four steps – data collection, data summarization, display of proportions, and construction of a Strengths and Weakness (SW) table by the application of rules that reflect the desired sensitivity of the methodology. The results of the SW table can be displayed graphically using basic symbols to highlight and track changes in students' perceptions. In this way, student progress towards meeting the program's EC 2000 objectives can be monitored and fed back to faculty. We have tested the method using 1999 and 2000 academic year data to track four student cohorts. The results have been highly consistent and indicate the usefulness of this methodology to efficiently measure student performance.

Introduction

For the past five years ABET, with its EC 2000, has directed undergraduate engineering faculty to implement a continuous improvement process. Following the setting of objectives, an early step in this process has often focused on data collection, typically using surveys to collect outcome information. As a consequence, faculty now find themselves with the task of interpreting a large amount of data while trying not to be overwhelmed with information that, in its present form, may have limited assessment value.

While the concept of continuous improvement may be new to the engineering academic culture [1], the art (and science) of data analysis is not. We possess a number of techniques for organizing data and developing metrics to assess performance, identify areas of weakness and design potential improvements. The challenge is to design data-driven metrics that are stable and robust for small populations and sample sizes, cost effective and easy to use for decision-making. Further, we have found that a well-designed survey can provide very valuable information about students' attitudes and perceptions [2]. Such surveys also can be viewed as a

---

* This paper supported in part by National Science Foundation grant: EEC-9872498, *Engineering Education: Assessment Methodologies and Curricula Innovations* and Engineering Information Foundation grant EiF 98-4: *Perception versus Performance: The Effects of Gender and Ethnicity Across Engineering Schools*.

means of measuring "customer satisfaction," where in this case the customer is the student (or alumni) and satisfaction is determined by measuring their attitudes and perceptions [3]. In addition, we can analyze those surveys relative to the expectations of faculty and administrators (i.e., the service providers). Those expectations are driven by the program objectives that they have established, which may have been influenced by such stakeholders as industry, alumni, and (in some cases) state governments. Our methodology, which we call the *Pitt-SW Analysis* [4, 5], is an adaptation of the competitive strategy principle of SWOT (strength, weakness, opportunities and threats). We have used this four-step methodology to track student cohorts for two academic years. The results have been highly consistent and, consequently, this procedure has been incorporated into our ongoing assessment activities.

Assessment Yes, But . . . What Are We Assessing?

ABET accredits individual engineering programs rather than the entire school. Each program undergoing review prepares a self-study report that serves as a basis for an onsite visit. The EC 2000 criteria force the program's faculty to express educational goals in terms of the graduate's expected characteristics and abilities and to establish a process for assuring that those goals are being met [6]. Hence, both quantitative and qualitative criteria are needed to fully assess a program [7]. Quantitative data is typically obtained by direct observation and testing for the more "technical" outcomes; e.g., measuring the graduate's ability to understand and apply mathematics, science and engineering principles in solving engineering problems (i.e., capstone design). Qualitative criteria may be more appropriate for assessing the "professional" outcomes; e.g., an understanding ethical and professional responsibilities; written and oral communication; or ability to work in multidisciplinary teams. To a large extent, the measurement of whether or not the desired educational outcomes are achieved depends on the graduates' collective perceptions about their acquired abilities and skills. These perceptions are influenced by the culture of the school, the students' prior experiences, out-of-classroom experiences, and interactions with students from other schools as well as the opinions of students and alumni. The more the perceptions reflect reality, the more sound the judgment of the person; and sound judgment means success [8]. However, contrary to a tangible measure, a "perception scale" shows uncertainty in its continuum [9]. This makes applying pure statistical methods to analyze perceptions impractical (and potentially biased). Here resides the core of our approach – an efficient method for providing reliable assessments using qualitative (perception) data about the students' educational achievement.

Completing The Process Improvement Cycle: A Way of Analyzing Data

During the two academic years 1999 and 2000, survey instruments were administered to University of Pittsburgh undergraduate engineering students[†] (freshmen through seniors). These closed form questionnaires were used as measurement tools for obtaining individual perceptions and attitudes about particular topics. The freshmen were surveyed prior to the beginning of the

[†] Because these instruments were administered over two academic years, multiple measurements were obtained for some students.

fall semester (pre-) and again at the end of the spring semester (post-), sophomores and juniors at the end of the spring semester, and seniors just prior to graduation. (An extensive alumni survey was conducted in 1998 and will be repeated in 2001.) The instruments solicited *engineering related attitudes, perceived ability to work in teams, confidence in personal abilities, confidence in having achieved the EC 2000 outcomes, pre-professional experiences* (junior and seniors) and *education and employment information* (senior). Alumni were also asked about the environment and educational experience. Each of the questions was rated using a five-point scale, where 1 represented either "Strongly Disagree," "Not at All," "Poor," or "None" depending on the question, and 5 indicated respectively "Strongly Agree," "A Great Deal" or "Very Good." The choices 2, 3 and 4 represented the intermediate points with 3 being neutral. Data were optically scanned and organized in a spreadsheet format to allow for graphical representation. Frequencies of the categorical data and percentages were then calculated, and the results displayed in a bar-chart format. Responses were further aggregated into favorable/positive and unfavorable/negative categories with respect to each outcome for the SW analysis; i.e., perceptions were re-grouped into positive (4's or 5's), neutral (3's) and negative (1's or 2's). This aggregation is necessary in order to obtain the desired confidence level given the relatively small sample sizes when data were analyzed by program and year. Figure 1 shows an example histogram for one particular survey item.

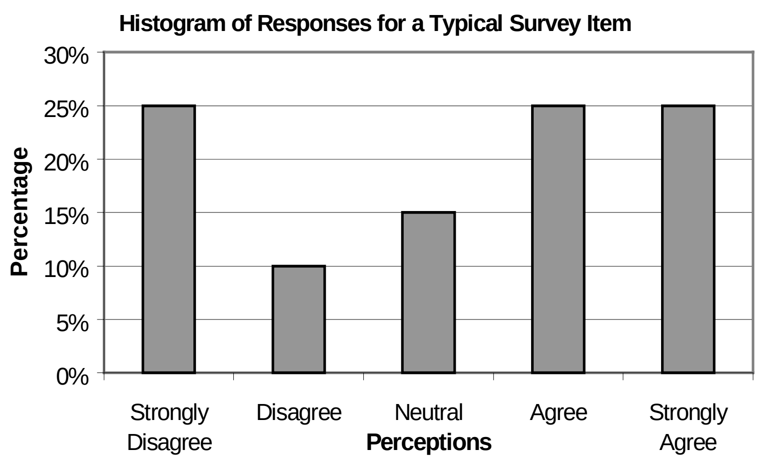

Figure 1: Example Response to Survey Item

The SW table was constructed by applying a set of four classification (decision) rules that represent the collective expectations of faculty and administrators, as shown in Figure 2. Note that the choice of 50% for a *Major* perception (Rule 1) is arbitrary and can be changed as described at the end of this paper. Also note that the 15% difference (Rules 2, 3 and 4) has been statistically validated [10, 11]. These rules have enabled us to classify student responses into five categories - *Major Strength, Possible Strength, Neutral, Possible Weakness* and *Major Weakness*. Two levels of feedback are necessary: one for setting standards for the quality of the schoolwide engineering educational system and a second, for measuring program performances against those standards. Standards are established by determining the level of sensitivity desired and by comparing each program against the schoolwide average or "standard." In order to apply these rules, the school's "standard" must be determined by calculating the average perception in terms of the three aggregated categories - "positive," (4 or 5) "neutral" (3) or "negative" (1 or 2) for each item across all the engineering programs.

The four classification rules (for each item by program or for the entire school) are:

1. Major perception:
   - If one response category (positive or negative) accounts for at least 50% (arbitrary) of the cases, then that item is defined as a *Major Strength* if positive or *Major Weakness* if negative.
2. Possible perception I:
   - If the largest category (positive or negative) exceeds the percentage of neutral responses but accounts for less than 50% of the observations, and it exceeds the other category by 15% or more then the item is classified as a *Possible Strength or Weakness.*
3. Possible perception II:
   - If both positive and negative perceptions exceed the neutral percentage, but each are less than 50% of cases, and the difference between both is greater than 15%, consider the greater value as a *Possible Strength/Weakness* or
   - If one positive/negative perception is equal to the neutral percentage and exceeds the other by 15% then classify as *Possible Strength/Weakness.*
4. Equal Effect:
   - If both positive and negative perceptions are equal or
   - If the difference between the positive and negative perceptions is equal to or less than 15%, or
   - If both perceptions are less than the neutral percentage then classify as *Neutral*

Figure 2. Classification Rules for Pitt-SW

Table 1 shows the aggregated averages for each of Pitt's eight programs (P1 to P8) and the School average for the item "*using mathematical concepts to solve engineering problems*." This display also shows the sample size N for each program.

Table 1: Distribution of Perceptions - Sophomores AY2000

Using mathematical concepts to solve engineering problems									
Program	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**School**
Negative	0.0%	3.0%	5.4%	0.0%	6.1%	3.1%	0.0%	2.4%	2.5%
Neutral	27.8%	30.3%	16.2%	31.6%	28.6%	31.3%	40.0%	29.3%	29.4%
Positive	72.2%	66.7%	78.4%	68.4%	65.3%	65.6%	60.0%	68.3%	68.1%
N=	18	33	37	19	49	32	10	41	239

These four rules are easy to apply as illustrated below. Table 2 shows the EC 2000 criteria assessment for the whole school. Using these rules, a report can be generated for each program for each year (sophomore, junior or senior). In addition to the SW table for a specific program, a summary table of the frequency responses is also provided for backup support. The Dean and Department Chairs can use this to make comparisons among programs or across institutions. We have used these results to not only assess students' progress towards achieving EC 2000 objectives, but also to compare different programs within the school. The rules could also be used to compare similar programs across different schools in order to detect potential trains or areas for improvements that may represent a systemic problem.

Table 2: Schoolwide SW Table - EC 2000 Outcomes for Sophomores (Threshold at 50%)

Confidence in Engineering Outcomes	**Major Strength**	**Possible Strength**	**Neutral**	**Possible Weakness**	**Major Weakness**
Using mathematical concepts to solve engineering problems	X				
Using chemistry concepts to solve engineering problems			X		
Using physics concepts to solve engineering problems	X				
Using engineering concepts to solve engineering problems	X				
Designing an experiment to obtain measurements or gain additional knowledge about the process			X		
Analyzing a set of data to find underlying meaning(s)	X				
Designing a device or process when given a set of specifications		X			
Function as an accountable member of an engineering team.	X				
Formulating unstructured engineering problems			X		
Using appropriate engineering techniques and tools including software and/or lab equipment for problem solving	X				
Understanding the professional and ethical responsibilities of an engineer	X				
Writing effectively		X			
Making professional presentations		X			
Effectively communicating engineering related ideas to others	X				
Listening to and impartially interpreting different viewpoints	X				
Understanding the potential risks and impacts that an engineering solution or design may have	X				
Applying knowledge about current issues to engineering related problems	X				
Recognizing the limitations of my engineering knowledge and abilities when to seek additional information	X				

Testing The Method: Assessing Students' Progress Towards Satisfying EC 2000 Criteria

A major concern has been measuring students' progress towards achieving EC 2000 outcome objectives. The University of Pittsburgh Strengths and Weaknesses analysis (Pitt-SW) enables us to take a "snapshot" of academic progress either by level or across all levels. Further, additional comparisons of students' progress towards satisfying EC 2000 outcomes can be made between post freshman and sophomore levels, sophomore and junior or junior and senior levels. These comparisons can be made for the school as a whole and for each program.

Table 3 illustrates how this information can be displayed (schoolwide). From Table 2, the number of X's in each column serve as a basis for comparison among levels[‡]. The progress toward EC 2000 is shown by the increase in the number of X's in *Major Strengths* and *Possible Strengths* and the decrease in the other categories. The results clearly show the improvement in student

---

[‡] At the University of Pittsburgh, which has a common freshman engineering program, the first year comparison can only be done for the whole School

confidence as they progress through the curriculum. Similar tables can be developed for each program.

Table 3: Schoolwide Student Self-Confidence - Achievement of EC 2000 Outcomes AY 2000

	MAJOR STRENGTH	POSSIBLE STRENGTH	NEUTRAL	POSSIBLE WEAKNESS	MAJOR WEAKNESS
Sophomore	12	3	2	1	0
Junior	15	2	0	1	0
Senior	17	0	1	0	0

Assessing Students' Progress: A Longitudinal Comparison

Another way of measuring EC 2000 progress is by tracking a student cohort as it goes from the sophomore to the senior level and graduation. Table 4 illustrates this comparison for AY 1999 juniors and then as AY 2000 seniors. It shows that six *Possible Strengths* in the junior year became *Major Strengths* by graduation.

Table 4: Student Self-Confidence – Achievement of EC 2000 Outcomes AY 1999 and AY 2000

	MAJOR STRENGTH	POSSIBLE STRENGTH	NEUTRAL	POSSIBLE WEAKNESS	MAJOR WEAKNESS
Junior 1999	11	6	0	1	0
Senior 2000	17	0	1	0	0
Change	6	-6	1	-1	0

Making Comparisons: The One-Dimensional Arrow Graphs

Two different types of graphs were developed to combine "snapshot" and longitudinal results. The first type – one-dimensional – can show either the movement between academic levels or between consecutive years. There are three ways to present comparative information: two consecutive years of a particular level (i.e., Junior AY 1999 and Junior AY 2000), two successive levels at consecutive years (i.e., Sophomore AY 1999 and Junior AY 2000), or simply successive levels at the same year (i.e., Junior AY 2000 and Senior AY 2000). A horizontal arrow is used to indicate the direction of change with the origin (tail) on the cell that corresponds to the initial level (or year) and the tip on the cell representing the subsequent level (or year). No change is indicated on the graph by a small square in one cell. For example, Figure 3 compares Sophomore AY 1999 to Junior AY 2000 for the outcome *using mathematical concepts to solve engineering problems*. The Figure indicates that there were three positive improvements among the eight programs, with two of these being from *Neutral* to *Major Strength* (programs P3 and

P5) and third from *Possible Strength* to *Major Strength* (program P4). For the other five programs this outcome was rated as a *Major Strength* when students were both sophomores and juniors.

This type of figure provides us with a way to see how student confidence changes across the curriculum. It also is useful for comparing different programs, to benchmark across programs, or to detect structural problems within the institution; e.g., when a program's students consistently indicate a particular area of weakness (limited self-confidence). While such graphs can provide valuable information, especially in terms of a specific program or issue, they still present information only on a question-by-question and level-by-level basis, rather than provide a more aggregate picture.

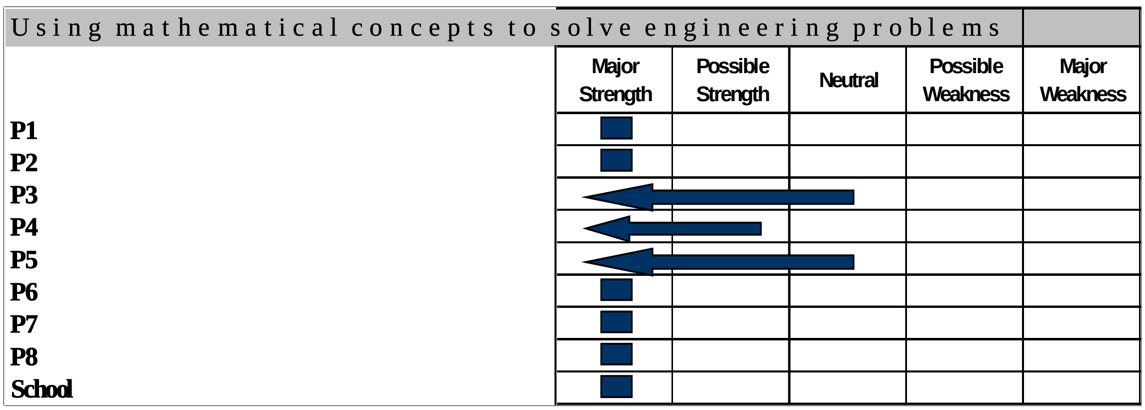

Figure 3: Comparison of Sophomore AY 1999 to Junior AY 2000

Making Comparisons: Two-Dimensional Arrow Graphs

In order to display changes over a three or four-year period, a second type of graph – two-dimensional - was developed. Here, the Y-axis displays the five classifications and the X-axis gives the three academic levels (sophomore, junior and senior). Hence, one graph can be prepared for each EC 2000 outcome and program. By varying the size and color of the dots we can distinguish among years and cohorts. For example, lighter colors (and/or larger sizes) can represent earlier years. These characteristics allow us to present multiple years in the same graph without loosing clarity. An arrow signals a change for a particular cohort between years. The arrow pointing upwards then indicates an improvement and an arrow pointing downward indicates a negative change. Two examples are depicted in Figures 4 and 5.

Figure 4 indicates that the confidence of sophomores, juniors and seniors in their ability *to use chemistry concepts to solve engineering problems* is low (although there was some improvement for AY 2000 seniors over their AY 1999 junior year). If this consistent pattern were unacceptable, finding the root cause would be warranted. Figure 5 shows that AY 2000 students are now more confident *in designing a device or process* than they were the previous year, and across the three levels (sophomore, junior and senior) AY 2000 students are more confident than their AY 1999 classmates for this outcome. These graphs can be presented to the faculty for review and possible action. During the academic year 2001 a third cohort will be obtained; this will enable us to have our first complete cycle of measuring progress and changes.

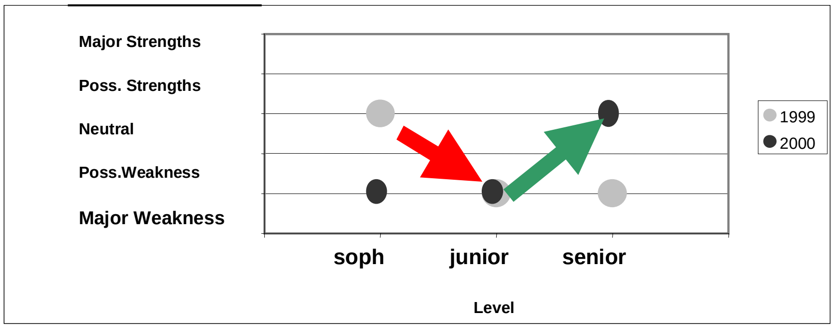

Figure 4: Two-Dimensional Graph:
"Using chemistry concepts to solve engineering problems"

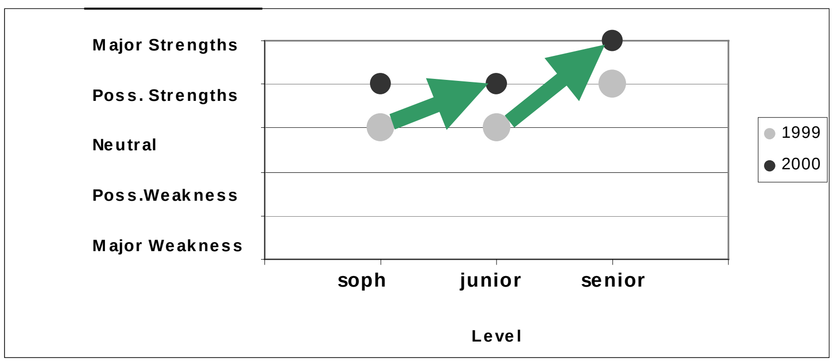

Figure 5: Two-Dimensional Graph:
"Designing a device or process to a set of specifications"

Re-Sensitizing the Graphs: Changing The Threshold

Another concern has been to refine the instruments that we are using in order to make them more sensitive. Having demonstrated the validity of the methodology, it is important to determine its sensitivity, especially in establishing how good is good? That is, what should the threshold be for classifying data as *Major Strength* or *Major Weakness*?

While it initially seemed reasonable to classify an outcome as a *Major Weakness* if 50% or more of the students perceived their ability as a negative, the opposite may not be true when designating an outcome as a *Major Strength*. Hence, we attempted to determine a more appropriate threshold for a *Major Strength*. To look for patterns, we plotted the frequency distribution (percent responding positively) of our original *Major Strength* classification for the sophomore, junior and senior levels. Results are shown in Figures 6 though 8. Note that for each succeeding year, the mean and median percentage is increasing, indicating that students' confidence increases as they move from one academic level to the next.

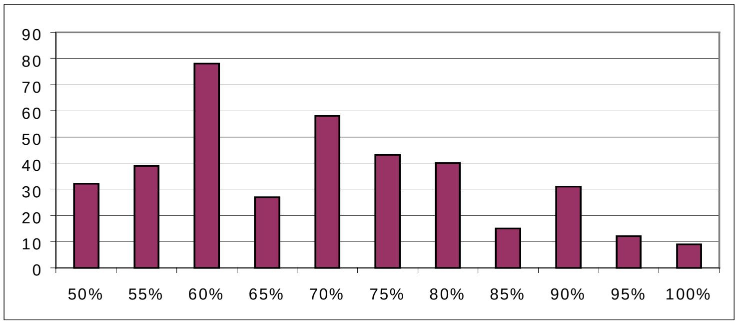

Figure 6: Distribution of Sophomore Major Strength Categories

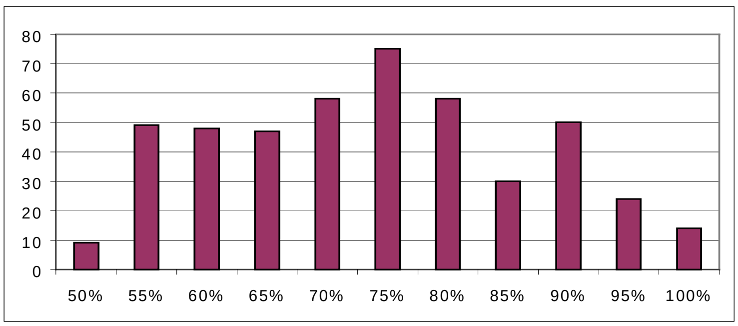

Figure 7: Distribution of Junior Major Strength Categories

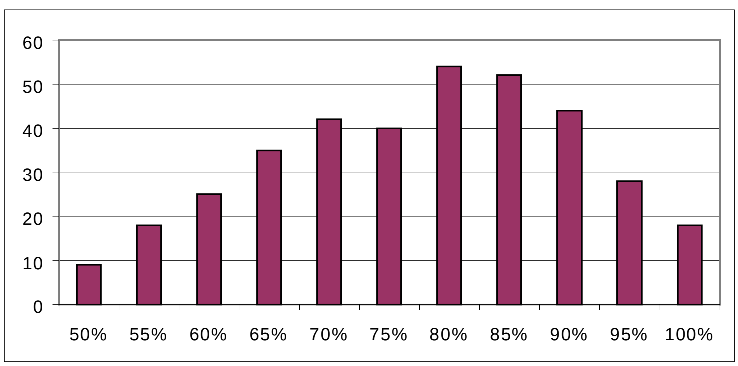

Figure 8: Distribution of Senior Major Strength Categories

We used the Mann-Whitney non-parametric test to confirm the statistical validity of this observation. Where:

H_0: $\mu_{\text{sophomore}} \leq \mu_{\text{junior}}$
H_1: $\mu_{\text{sophomore}} > \mu_{\text{junior}}$

The Mann-Whitney test resulted in a *P-value* of < 0.001; i.e., there is a significant difference between classifications for juniors compared to sophomores. The test showed similar significant differences for the other two comparisons: juniors to seniors and sophomores to seniors.

Recall that a positive response is indicated by a score of either “4” or “5.” The proportion of “5’s in the *Major Strengths* classification by both sophomores and juniors averaged 28%; for seniors the average increased to 46%. This provides additional face validity for our classification procedure and further suggests that we are able to measure progress in student self-confidence.

Table 5: Schoolwide SW Table - EC 2000 Outcomes for Sophomores (Threshold at 80%)

Confidence in Engineering Outcomes	**Major Strength**	**Possible Strength**	**Neutral**	**Possible Weakness**	**Major Weakness**
Using mathematical concepts to solve engineering problems		X			
Using chemistry concepts to solve engineering problems			X		
Using physics concepts to solve engineering problems		X			
Using engineering concepts to solve engineering problems		X			
Designing an experiment to obtain measurements or gain additional knowledge about the process		X			
Analyzing a set of data to find underlying meaning(s)			X		
Designing a device or process when given a set of specifications		X			
Function as an accountable member of an engineering team.		X			
Formulating unstructured engineering problems		X			
Using appropriate engineering techniques and tools including software and/or lab equipment for problem solving			X		
Understanding the professional and ethical responsibilities of an engineer		X			
Writing effectively		X			
Making professional presentations		X			
Effectively communicating engineering related ideas to others		X			
Listening to and impartially interpreting different viewpoints		X			
Understanding the potential risks and impacts that an engineering solution or design may have		X			
Applying knowledge about current issues to engineering related problems		X			
Recognizing the limitations of my engineering knowledge and abilities when to seek additional information		X			

The above analysis has led us to set the threshold for a *Major Strength* at 80%. (The mode of the distribution of senior *Major Strength* categories as shown in Figure 8.) As a consequence of the new limit, the profile of the Pitt-SW has changed with the median classification now being *Possible Strength* rather than *Major Strength*. Further, such outcomes as teamwork, communication skills or professionalism are now considered areas for additional improvement. Table 5 shows these revised results for sophomores. Relative to Table 2, the 12 *Major Strengths* have now been reclassified as *Possible Strengths*. This reclassification now allows sufficient room for improvement as the students progress over their next two years and does not allow us to assume after two years that we have achieved our educational objectives. Using this reclassification scheme for juniors and seniors provided consistent results, with seniors still achieving a relatively high proportion of *Major Strengths*.

Conclusions

In assessing an engineering educational program as part of the EC-2000 process, it is important to consider quantitative as well as qualitative data. We have presented one way of using qualitative information specifically, student self-ratings along different dimensions that influence their education. These ratings are then categorized into a valid, systematic format that is easy for faculty and administrators to use. Simple, but useful reports can be generated that measure the impact of improvements to the educational system. The Pitt-SW analysis provides a picture of the overall state of a program of interest as well as that of the school at certain points in the academic progression. The one-dimensional arrow diagrams allow for comparisons between levels and programs, in both a cross-sectional and longitudinal manner so that changes can be tracked and areas for potential improvement identified for specific outcomes. The two-dimensional arrow diagrams also allow faculty and administrators to track changes in the student perceptions over a period of several years, or to compare different cohorts of students longitudinally. In this way they can obtain relatively quick feedback and identify potential positive or negative shifts that may have occurred during the academic year. Hence, both graphs provide information for needed feedback loops so that interventions can be designed and implemented and resources allocated in a timely manner as part of an effective continuous improvement system.

The methodology developed here can be made more (or less) sensitive by changing the threshold limits. In this way it is possible to screen for different acceptable levels of self-confidence and the perceived acquisition of desired skills. As a result we also can uncover new areas for research and better enable students to improve such as skills, teamwork, communications and professionalism. The methodology is relatively easy to apply and cost effective. As more data is collected and its use becomes routine, it should lead to continued educational improvements in the institution.

References

1. Prados, J.W., “ABET Engineering Criteria 2000: How We Got There and Why?” *Proceedings of the 1997 American Society for Engineering Education Annual Conference and Exposition*, American Society for Engineering Education, Milwaukee, Wiscosin, June 1997.
2. Besterfield-Sacre, M.E., C.J. Atman, and L.J. Shuman, “Engineering Student Attitude Assessment,” *Journal of Engineering Education*, April, 1998, pp. 133-142.
3. Churchil, G.A., *Marketing Research - Methodology Foundations*. Fifth Edition. The Dreyden Press. 1991
4. Perez, G. L., Shuman, L., Wolfe, H. and Besterfield-Sacre M. E., “A Strengths and Weaknesses Data Driven Approach to Assess Engineering Programs,” *Best Assessment Process III*. Rose-Hulman Institute - Terra Haute-Indiana. April 2000
5. Perez, G. L., Shuman, L., Wolfe, H. and Besterfield-Sacre M. E. “A Strengths and Weaknesses Data-Driven Approach for Measuring Performance in Engineering Programs.” Working Paper, Department of Industrial Engineering, University of Pittsburgh, 2001.

6. Prados, op. cit.
7. Rogers, G. M., “EC 2000 and Measurement: How Much Precision is Enough?”, *Journal of Engineering Education.* 89(2) pp. 161-165, 2000.
8. Piket-May M.J.,Chang, J.L. and Avery J.P., “Understanding What Success Means in Assessment,” University of Colorado at Boulder. *Proceedings of the 1999 Frontiers in Engineering Conference*. San Juan de Puerto Rico November 1999.
9. Allport F.H., *Theory of Perception and the Concept of Structure*. Third Edition. John Wiley and Son. 1961.
10. Perez, et. al., op. cit., 2000.
11. Perez, et. al., op. cit. 2001.

GRACIELA DE L. PEREZ
Graciela de L. Perez is currently a doctoral student at the University of Pittsburgh. She received her B.S. in Industrial Engineering from the University of Cuyo in Argentina, a M.I.B. from the Ecole Natieonale des Ponts et Chaussess in France. Before returning to graduate school to obtain her doctoral degree, she worked during 12 years in manufacturing, finance and marketing in her native country. Her research interests are in the decision -making field and modeling applications for quality improvement in service organizations.

LARRY J. SHUMAN
Larry J. Shuman is Associate Dean for Academic Affairs, School of Engineering, University of Pittsburgh and Professor of Industrial. His areas of interest are improving the engineering educational experience, and the study of the ethical behavior of engineers and engineering management. Together with Dr. Cynthia Atman, Dr. Shuman co-chaired the 1997 Frontiers in Education Conference held in Pittsburgh. He is a member of the FIE Steering Committee, and will be Academic dean for the “Semester at the Sea” for the Spring 2002 semester.

HARVEY WOLFE
Harvey Wolfe has been a Professor in the Department of Industrial Engineering at the University of Pittsburgh since 1972 and served as Department Chair from 1985 to 2000. He is a fellow of the Institute of Industrial Engineers and serves as Member at Large of the Professional Enhancement Board of the Institute of Industrial Engineers. He is serving a second six years term as ABET evaluator. After many years working in the area of operations research methods to the health field, he is now active in the development of models for assessing engineering education.

MARY BESTERFIELD-SACRE
Mary Besterfield-Sacre is an Assistant Professor in the Industrial Engineering Department at the University of Pittsburgh. Her principal research interests are in empirical and cost modeling applications for quality improvement in manufacturing and service organizations, and in engineering education evaluation methodologies. She has been assistant professor at the University of Texas -El Paso, and has worked as an Industrial Engineer with ALCOA and with the U.S. Army Human Engineering Laboratory.