ABET’s new criteria, “EC 2000,” has brought assessment and evaluation to the forefront of engineering education. Concomitantly, the focus of ABET’s program evaluation has shifted from “what are you [the program] doing?” to “how is what you’re doing achieving the desired outcomes [what are your students doing]?” In short, accreditation will be concerned as much or more with outcomes rather than inputs or processes. More important, accreditors will be looking at how problems are identified and improvements are made in order to affect the program’s outcomes. Therefore, a comprehensive evaluation system should not only accurately measure student outcomes, through proper instrument development and administration, but also possess a well-designed feedback mechanism that allows for the tracking of these outcomes over time. By appropriately tracking outcomes, engineering educators should be able to identify areas for improvement, as well as monitor the effectiveness of programmatic interventions.

The paper will discuss a potential role for process control monitoring of student outcomes for program evaluation and improvement. Using example data about freshman engineering student attitudes, we will present two non-parametric control-charting methods that are appropriate for either survey data or data that are known to be non-normal. We will then illustrate how the charts may be used to identify improvement opportunities and track interventions.

Introduction – Use of Quality Control Charts in Engineering Education

EC 2000 directs engineering faculty to not only demonstrate that their students have achieved eleven specific outcomes upon graduation, but it also encourages them to continuously improve, in innovative ways, the learning experience. To accomplish this, engineering educators will need evaluation protocols and measurement instruments that will facilitate feedback and the resultant improvements.

Statistical process control (SPC) and control charting, in particular, can be used as a feedback mechanism. Industry has commonly used SPC techniques to assure that production remains “in control” according to pre-determined specifications and process capability. Recently, several authors have proposed applying these concepts to engineering education; i.e., to assure that educational processes and outcomes are in control. See for example 4,5,6. Why would an institution be interested in control charting? As engineering educators become more
knowledgeable about assessment and evaluation, we propose that one result will be the need to track changes in outcomes, particularly in response to programatic changes. One useful tool is a monitoring system that can alert engineering educators to shifts in outcome data from one assessment period to the next. Such trends might be indicative of either positive or negative changes occurring within the engineering program. For example, suppose for the past two years that graduating seniors have scored above the national average on the GREs. Does this warrant reviewing the learning process to determine what factors are contributing to this positive outcome? Is this just a natural fluctuation in the data, which does not indicate that a positive significant change in the engineering education system has occurred?

Not only is SPC a valuable tool for identifying trends or changes, but it can be use data obtained from surveys and questionnaires. If surveys are administered periodically (e.g., semi-annually, yearly, or even bi-annually), it is possible to investigate how these measurements change over time by plotting sample statistics about the mean, range, or proportion. Using the appropriate control chart, one can then observe if a sample is exceeding previously defined limits (which may have been obtained through process capability studies \cite{7}). If a sample exceeds a control limit, there is strong suggestion that an assignable cause exists for this variation, such as major differences attributed to the implementation of a new initiative. If a sample does not exceed a limit, then sample-to-sample variation may just be due to common-cause variation. By allowing variations to be examined in a logical manner, control charts can provide engineering educators with the information needed to make systematic changes. However, in order to yield viable results, the underlying assumptions for its use, the process being measured and the type of data that are being used to measure the process, all need to be carefully considered.

Typical Shewhart control charts used in industry to monitor quality-based variables include the $\overline{X}$ and $R$ (or $s$) charts, and the $p$ chart for attributes \cite{8}. The $\overline{X}$ (the average of sample) and $R$ (the range of the sample) control charts are widely used to monitor the mean and variability. If a survey question requires a ‘yes-no’ or ‘satisfied-dissatisfied’ response, the $p$ chart can be used to monitor the proportion of those individuals responding “yes” or “satisfied”; e.g, with their engineering education. To gain finer discrimination, though, ordered responses may be elicited on questionnaires (e.g. 1 – “not very confident” to 5 – “very confident”). In this latter case, one might consider using control charts for variables. However, several issues should be considered and investigated before one implements these charts for questionnaire data. First, proper use of $\overline{X}$ and $R$ assumes that individual responses be continuous and normally distributed. Ordinal scales used on many questionnaires are discrete rather than continuous, and the data often fail tests for normality. Figure 1 displays a histogram of survey questionnaire responses that do not satisfy the normality assumption.

Second, constants used to develop $\overline{X}$ charts are not amenable to the large sample sizes that are customary in questionnaire studies. Third, the response rate for administering questionnaires varies, thus requiring the control limits to change accordingly. Use of traditional charts may yield misleading results and possibly increase Type I (concluding that the process is in-control when it is really out-of-control) and Type II (concluding that the process is out-of-control when it is really in-control) errors \cite{9}. 
The implementation of non-parametric control charts can help to overcome these issues. Because they are distribution-free, they can handle observations that are neither normally distributed nor continuous. Further, the process variance does not need to be known in order to establish a control chart for the mean, which is not the case for the $X$ chart. Also, depending on the type of data being charted, many non-parametric charts can accommodate large sample sizes. Non-parametric control charting is gaining interest in industry, as well.

For this paper, we present two non-parametric control chart alternatives to the traditional $X$ and $R$ charts, the $\chi^2$ control chart and the modified-$p$ chart, that work well for closed-form questionnaire data. Using four years of questionnaire data from our Freshman Engineering Attitudes Survey, we present a worked example of how one might derive control limits and discuss some possible uses for the control charts.

Two Alternative Non-Parametric Control Charts

This section describes two alternative non-parametric control charts, the $\chi^2$ and the modified-$p$ charts, that may be used when assumptions for typical $X$ and $R$ charts are possibly violated.

The $\chi^2$ Chart

The theoretical basis for the $\chi^2$ chart has been established for some time. Recently $\chi^2$ charts have been used to monitor customer satisfaction survey data in the medical field, making the chart a plausible candidate for use with questionnaire data about students’ attitudes. This chart is based on using the $\chi^2$ goodness-of-fit statistic to compare an actual distribution with a theoretical distribution, as shown in the equation below.

$$\chi^2_j = \sum_{s=1}^{k} \frac{(Y_{sj} - E_{sj})^2}{E_{sj}}$$

where: $k$ is the number of categories, which corresponds to the number of ordered responses on a questionnaire.
\( Y_{xj} \) is the number of individuals responding \( x \) on sample \( j \)

\[ E_{xj} = n_j p_x \] and is the expected number of individuals responding \( x \) on sample \( j \)

\( (n_j \) is the total number of individuals from sample \( j \) and \( p_x \) is the expected number of customers giving response \( x \) on sample \( j \).

The \( \chi^2 \) control chart has a single upper control limit (UCL), as compared to the traditional \( \bar{X} \) chart which has both an upper and lower control limit. If a sample exceeds the UCL, then a shift in the underlying distribution has occurred. Duncan \(^{14}\) computed standard upper control limits for each of the possible \( k \) categories, as shown in Table 1. Duncan's values for the UCLs assure that the probability of obtaining a Type I error is similar to that using traditional Shewhart charts using three standard deviations.

**Table 1. UCL Values for the \( \chi^2 \) Chart**

<table>
<thead>
<tr>
<th>( k )</th>
<th>UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>9.00</td>
</tr>
<tr>
<td>3</td>
<td>11.83</td>
</tr>
<tr>
<td>4</td>
<td>14.16</td>
</tr>
<tr>
<td>5</td>
<td>16.25</td>
</tr>
<tr>
<td>6</td>
<td>18.21</td>
</tr>
<tr>
<td>7</td>
<td>20.06</td>
</tr>
<tr>
<td>8</td>
<td>21.85</td>
</tr>
<tr>
<td>9</td>
<td>23.57</td>
</tr>
<tr>
<td>10</td>
<td>25.26</td>
</tr>
</tbody>
</table>

The \( \chi^2 \) chart does not require that the observations be normally distributed. In addition, the chart can easily accommodate different sample sizes. Unfortunately, because it only has one control limit, if a shift in the mean has occurred, the chart cannot indicate the direction of the shift. One would have to review the sample means to determine the shift direction.

**The Modified-\( p \) Chart**

The modified-\( p \) chart \(^{15}\) is an extension of the \( p \)-chart to more than two categories. It is similar to the \( \bar{X} \) chart in that the plotted statistic is the average, but, again, the control limits do not require that the underlying distribution be normal. Rather, the underlying distribution is based on the sum of the responses. As with the \( \chi^2 \) chart, there can be \( k \) possible response groups. Equally important, the sub-group sizes can vary and can be large. Unlike the \( \chi^2 \) chart, though, there are two control limits. This allows the engineering educator to observe whether a trend in the data is present as well as indicate any out-of-control points.

From background data, the percentage of responses for \( x \) is known and is given as \( p_x \). The response \( X \) given on a questionnaire is a random variable with probability \( \Pr\{X = x\} = p_x \). The formulas for the expected value and population variance for the distribution are given below.

\[
\mu_x = \sum_{x=1}^{k} xp_x
\]
\[ \sigma_x^2 = \sum_{k=1}^{k} x^2 p_x - \left( \sum_{k=1}^{k} x p_x \right)^2 \]

For a sample of \( n \) survey responses, the sample mean is equal to the expected value and the sample variance is equal to \( s_x^2 = \frac{\sigma_x^2}{n} \). The control limits for the modified-\( p \) chart are then given as follows.

\[
\sum_{x=1}^{k} x p_x \pm \frac{3}{\sqrt{n}} \sqrt{\sum_{x=1}^{k} x^2 p_x - \left( \sum_{x=1}^{k} x p_x \right)^2}
\]

**Example - Charting of Freshman Engineering Attitudes**

The Pittsburgh Freshman Engineering Attitude Instrument\textsuperscript{©}, originally developed and tested at the University of Pittsburgh\textsuperscript{16}, has been administered every academic year since 1993-94. The questionnaire measures several facets of student attitudes including their opinions about the engineering profession and the reasons for studying engineering. Students also are asked to rate their confidence in the pre-requisite background knowledge and skills, and in their abilities to succeed in engineering. Finally, students rate their study skills and their interest in working in groups. The closed-form questionnaire contains 50 items that students rate on either a Likert scale or an ordinal-based self-assessed confidence scale. These 50 items have been clustered into thirteen student attitude and self-assessment measures, as shown in Table 2. The questionnaire is administered to students twice - “the pre survey” at the beginning of the freshman year during orientation (before beginning classes) and the “post survey” at the end of the freshman year\textsuperscript{ii}.

The survey was originally used to evaluate innovative changes to the freshman engineering curriculum at the University Pittsburgh during academic years 1993-1995\textsuperscript{17}. Over two hundred students took the survey each year. The instrument has also been used to develop empirical models for identifying those students who may potentially leave the freshman engineering program before they begin their freshman studies. These models have been used for proactive advising at the freshman level\textsuperscript{18}. Since then, the questionnaire has been used by a number of other institutions to help evaluate their freshman engineering programs, as well as to participate in a national cross-institutional study of freshman engineering attitudes\textsuperscript{19,20}.

The questionnaire has become a standard evaluation instrument at several schools, including the University of Texas - El Paso, the University of Pittsburgh, and the North Carolina State University, who have all used the questionnaire for several years. With the continued use of the questionnaire, monitoring and tracking freshman engineering attitudes over time becomes of greater importance. We want to know if continued improvements we make to our freshman engineering program are reflected in positive student attitudinal changes. In addition, we would like to know if ‘sudden’ year-to-year attitudinal variations reflect a substantial change in the program or just a random fluctuation for that particular year.

\textsuperscript{ii} Some of the ten schools currently using the instrument administer the “post” survey at the end of the first semester rather than the end of the freshman year.
Table 2. Student Attitude and Self-Assessment Measures and Their Definition

<table>
<thead>
<tr>
<th>Student Attitude and Self-Assessment Measures</th>
<th>Definition</th>
<th>Rating Value</th>
</tr>
</thead>
</table>
| General Impressions of Engineering            | How much a student likes engineering | 1 – does not strongly like engineering  
5 – strongly likes engineering |
| Financial Influences for Studying Engineering | Belief that engineers are paid well and that having an engineering degree helps assure career security | 1 – does not strongly hold this belief  
5 – strongly holds this belief |
| Perception of the Work Engineers Do and the Engineering Profession | Considers engineering a respectable field and the work engineers do has a positive impact in solving the world’s problems | 1 – does not strongly hold this belief  
5 – strongly holds this belief |
| Enjoyment of Math and Science Courses         | Preference for math and science courses over liberal arts courses | 1 – does not strongly hold this preference  
5 – strongly holds this preference |
| Engineering Perceived as Being an “Exact” Science | Belief that engineering is an exact science | 1 – does not strongly hold this belief  
5 – strongly holds this belief |
| Engineering Comparing Positively to Other Fields of Study | Preference for engineering over other fields of study | 1 – does not strongly hold this preference  
5 – strongly holds this preference |
| Family Influences to Studying Engineering     | Belief that parents are influencing student to study engineering | 1 – does not strongly hold this belief  
5 – strongly holds this belief |
| Confidence in Chemistry                       | Self-assessed confidence in chemistry knowledge | 1 - has low confidence  
5 - has high confidence |
| Confidence in Communication Skills            | Self-assessed confidence in writing and speaking skills | 1 - has low confidence  
5 - has high confidence |
| Confidence in Basic Engineering Knowledge and Skills | Self-assessed confidence in knowledge of calculus and physics, and in computer skills | 1 - has low confidence  
5 - has high confidence |
| Adequate Study Habits                         | Beliefs about the adequacy of current study habits | 1 - not comfortable with study habits  
5 – comfortable with study habits |
| Working in Groups                             | Preference for working in groups | 1 – prefer working alone  
5 – prefer working in groups |
| Confidence in Engineering Skills              | Belief that one has the creative thinking, problem solving and design skills required to survive in engineering | 1 - has low confidence  
5 - has high confidence |

Development of Control Charts

For the purposes of this paper, we have focused our non-parametric control charting efforts on one attitude measure from the questionnaire, “Enjoyment of Math and Science Courses.” This measure assesses a student’s preference for math and science courses over liberal arts courses. It consists of two statements: “I enjoy the subjects of science and mathematics the most” and “I enjoy taking liberal arts courses more than math and science courses,” with the response averaged (the second statement is reverse-scored). This attitude measurement was chosen for demonstration because it typically has been found to deteriorate during the freshman year. Not only is the “annual” decline in this measure a concern, but also is the magnitude of that decline. In particular, over time are our students becoming increasingly more disenchanted with their math and science courses or are the negative changes that are observed over the freshman year basically “constant” from one year to the next?

To chart these measures, both the $\chi^2$ and the modified-$p$ charts are utilized. Because the $\chi^2$ chart only has an upper control limit, we focused its use on determining if there were significant differences from the pre-survey to the post-survey and whether or not these were consistent from year-to-year. For the modified-$p$ chart, attention is focused on two objectives: (1) were there
differences from the pre- to the post-survey, as with the \( \chi^2 \) chart?, and (2) are there any trends occurring in the data that may be reflective of continuous improvements in the system?

The \( \chi^2 \) Chart
To determine the upper control limit for the \( \chi^2 \) chart, we use responses from the pre-survey to establish the upper control limit and plot the post-survey responses for the attitude measure “Enjoyment of Math and Science Courses.” For each year in the study, new control limits were established based on the pre-survey data. To determine the number of categories, \( k \), and calculate the expected number of individuals \( E_{xj} \) there should be five or more observations per category. For example, suppose in the first year the questionnaire was administered there were only two people that had an average of ‘1’ for the measure “Enjoyment of Math and Science Courses.” In that case, these two individuals would be combined with those individuals whose average ratings were in the next category, say ‘2.5’ or ‘2’. Figure 2 displays the \( \chi^2 \) chart for the attitude measure “Enjoyment of Math and Science Courses.” For academic years 1993-94 and 1994-95, data for each year grouped into \( k = 4 \) categories. Therefore the upper control limit for these two particular years was 14.16. Whereas, for years 1995-96 and 1996-97, data for each year grouped into \( k = 3 \) categories, resulting in upper control limits equal to 11.83 both years.

As shown in the figure, for academic years 1994-95 through 1996-97 at the University of Pittsburgh, the post-survey attitude measure is out-of-control (above the control limit) when compared to the pre-survey attitude measure. The results here are not surprising. (From the data, we also know that this change represents a decline.) We have found such significant differences between the pre- and post-survey responses consistently during these same years using non-parametric t-tests. The value of control charting is that the year-to-year results can be displayed on one chart.
The Modified-$p$ Chart

To determine control limits for the first modified-$p$ chart, we use responses from the pre-survey to establish the control limits and plot the post-survey responses for the attitude measure “Enjoyment of Math and Science Courses.” For the modified-$p$ chart, we can either establish new control limits for each year, as with the $\chi^2$ chart or we can establish control limits if we have “a priori” information about the process. Because there were no significant differences in the pre-survey data between the four years for the attitude measure “Enjoyment of Math and Science Courses,” we are able to pool the “pre” scores for the year in question. That is, we will consider the pre-survey data from the four years as “a priori” information about the process and estimated the proportions of the possible responses, $\hat{p}_x$, by the following equation,

$$
\hat{p}_x = \frac{\sum_{j=1}^{m} Y_{ij}}{\sum_{j=1}^{m} n_j}
$$

where $Y_{ij}$ is the actual number of individuals that gave response $x$ on sample $j$. $\hat{p}_x$, is the proportion of observed individuals giving response $x$ for all the samples (or academic years) combined. The resulting upper and lower control limits are 3.99 and 4.28, respectively; and the mean pre-survey value for the attitude measure is 4.14. As shown in Figure 3, the modified-$p$ chart produced different results than the $\chi^2$ chart. For the modified-$p$ chart, both the second and fourth years are out-of-control, similar to the $\chi^2$ chart, but the third year remains in control.

![Modified-p Control Chart for 'Enjoyment of Math and Science Courses' Control Limits Established with Pre-Survey Data](image)

**Figure 3. Modified-$p$ Control Chart with Pre-Survey Control Limits**

Because the $\chi^2$ chart uses $k$ categories for the number of ordered responses and requires at least five observations per category, the results can differ if ordered responses have to be combined. For this particular case, the modified-$p$ chart is more accurate because the response categories are not combined. Compared to the $\chi^2$ chart, the modified-$p$ chart provides more insight. For the
modified-\( p \) the control limits, mean average, and average for the academic year reflect the actual range of possible response values on the questionnaire (1 to 5). Hence, an engineering educator can see how improvements affect attitudes. For example, on average students enter their engineering studies with a rating of 4.14 for “Enjoyment of Math and Science Courses;” however, over the course of the freshman year this attitude measure diminished significantly (i.e., became out-of-control) every academic year except for the first year (1993-94). By introducing an educational intervention aimed at improving students’ enjoyment of math and science, we may be able to track future post-survey responses that will determine if, in fact, the intervention worked. That is, subsequent post-survey averages would remain within the control limits and close to the mean average.

The second objective for using the modified-\( p \) chart is to determine if any trends have occurred in the data that may be reflective of continuous improvements in the engineering education system. To determine this, the post-survey responses were used to establish the control limits. In establishing the limits, the same methodology was applied as with previous modified-\( p \) chart. The resulting upper and lower control limits were 3.75 and 4.13, respectively; and the mean post-survey value for the attitude measure was 3.94. Figure 4 shows that all post-survey responses are within control. This is to be expected since the same data were used to establish the limits. Such a chart would be valuable if we continue to use these control limits and monitor future post survey responses.

![Modified-p Chart for 'Enjoyment of Math and Science Courses'
Control Limits Established with Post-Survey Data](image)

Figure 4. Modified-p Chart with Post-Survey Control Limits.

The control chart described above allows us to analyze future samples of freshman engineering students, as well as future trends. If we see a sample point out-of-control on the high side (exceeding the upper control limit), we need to investigate the assignable reasons for this encouraging result; i.e., what is happening in the classroom to enable students’ enjoyment of their math and science courses to increase? In other words, we need to identify what are we doing right during the freshman year so that we can capitalize on it for future freshman.
As a tool for continuous improvement, the control chart becomes a valuable asset for analyzing trends that may be occurring in the data from one sample to the next, but that may not show a significant difference from the control chart mean. For example consider the hypothetical trend for students’ post-survey responses for “Enjoyment of Math and Science Courses,” displayed in Figure 5. The sample points for the academic years 1997-1998 through 2000-2001 are all above the mean, but have not exceeded the upper control limit. Conducting non-parametric statistics comparing each individual year’s sample to the control chart mean may not yield a significant difference. However, by analyzing the chart using ‘rules’ for control charts [22], one would find that the trend is considered ‘out-of-control’ because four out of five consecutive control points fall between two and three standard deviations from the mean. The next step might be to find out why this occurred? Have the mathematics, physics or chemistry faculty introduced any pedagogical changes that would have caused these attitudinal shifts to occur? Further, if there has been a positive shift in the average post-survey response for this attitude measure, then new control limits will need to be established.

Figure 5. Hypothetical Example of the Modified-p Chart Demonstrating Significant Trends and Possible Need to Re-evaluate Control Limits.

Conclusions and Future Work

This paper describes the utility of control charts to monitor attitudinal data. Such a system would not only provide both graphical and statistical documentation for quick review by engineering educators and ABET evaluators, but also should motivate engineering educators to continuously improve their system. Second, just as the engineering education system is inherent with variation, so are the students who experience it. Not only should assessment instruments be sensitive to such variations, but the evaluation tracking mechanisms and analyses should also be sensitive to variation. Although much of the assessment data collected in engineering education today may not be conducive to traditional Shewhart charts, there are other, non-parametric alternatives that may be utilized.
We are still in the preliminary stages of investigating the use of control charting to track freshman engineering attitudes longitudinally. We are currently developing control charts for the other twelve measures in the questionnaire and are investigating different approaches for establishing control limits. In addition, we are researching other types of control charts that may be better suited for charting engineering attitudes and other types of assessment data. For example, multivariate quality control methods, such as the Hotelling $T^2$ Control Chart, and moving average control charts \cite{23} may provide more sensitive ways to identify process shifts in mean values of the sample.

References

7. Ref. 2, pp. 255-278.
9. Ref. 2, pg. 133.
15. Ref. 13.


Ref. 17.

Ref. 2.


### Biographical Information

MARY BESTERFIELD-SACRE is an Assistant Professor in the Mechanical and Industrial Engineering Department at the University of Texas - El Paso. Her research interests include modeling applications for quality improvement in manufacturing and service organizations, and engineering education assessment/evaluation methodologies. She holds a Ph.D. in Industrial Engineering from the University of Pittsburgh.

NANCY Y. AMAYA is a Research Assistant in the Mechanical and Industrial Engineering Department at the University of Texas - El Paso. She holds a B.S. in Industrial Engineering from UTEP. Before returning to UTEP to pursue her M.S. in Manufacturing Engineering, Nancy was an Industrial Engineer with the VF Corporation.

LARRY J. SHUMAN is Associate Dean for Academic Affairs and Professor of Industrial Engineering at the University of Pittsburgh. His primary areas of interest are the application of operations research to health delivery systems, improving the engineering educational experience and the ethical behavior of engineers. Drs. Shuman and Atman were the General Co-Chairs for the 1997 Frontiers in Education Conference held in Pittsburgh, PA.

CYNTHIA J. ATMAN is an Associate Professor of Industrial Engineering at the University of Pittsburgh. Her research interests include engineering education evaluation and documenting cognitive models of engineering students’ learning processes. She is the recipient of a NSF Young Investigator award to pursue her research in engineering education. She received her Ph.D. in Engineering and Public Policy from Carnegie Mellon University.