

Benchmarking Student Learning Outcomes using Shewhart Control Charts

Steven J. Peterson, MBA, PE
Weber State University
Ogden, Utah

This paper looks at how Shewhart control charts—a statistical tool used in manufacturing quality control—may be used to establish and monitor benchmarks for student learning outcomes. By using Shewhart control charts, faculty can establish meaningful benchmarks based on the current educational process and identify when a statistically significant change (improvement or decline) in student performance has occurred. Included is a case study using data from Weber State University's construction management program, which uses the Associate Constructor Level 1 exam as an assessment tool.

Key Words: Shewhart, Outcomes, Control Charts, AC Exam, Assessment

Introduction

When the implementation of outcomes is discussed, questions often arise regarding the setting of a target threshold (benchmark) for evidence of student learning. Most of the questions center around two issues.

First, where should the benchmark be set so that faculty know that their students are showing adequate evidence of learning? Most faculty agree that there is a problem if the benchmark is set very low so that it requires minimal effort by the students to meet the benchmark. They also agree that it is also a problem if the benchmark is set so high that the majority of the students do not meet it. But where in that broad range, between being too low and too high, should the benchmark be set?

Second, when a decline in student performance occurs, how do faculty determine whether the decline is due to randomness, a change in teaching, or a fundamental change in their students? The teaching process consists of strategies that faculty can control (e.g., teaching methods, assessment methods, curricular content, required prerequisites) and student issues they cannot control (e.g., the students' willingness put forth the effort needed to master the material, the students' level of preparedness for the class). Using their benchmarks as guides, how do faculty determine what to pay attention to and what they can ignore?

Shewhart control charts provide one possible way to address these questions.

Literature Review

In 1956, Western Electric published the *Statistical Quality Control Handbook* as a practical working handbook for their employees and made a few copies available to others who were interested. The handbook was written in non-technical language, and its purpose was to describe the quality control procedures essential to Western Electric's quality control program. Part B of Section I details the use of control charts (often referred to as Shewhart control charts). This book can be downloaded from the internet.

Wheeler (2000) discusses how Shewhart control charts can be used to analyze a wide variety of data, including company financial reports, the US trade deficits, transit times, labor costs, frequency of spills, home runs in baseball, and, of course, product defects. This book provides a good foundation for understanding the use of Shewhart control charts.

Case Study: Shewhart Control Charts with Construction Management at Weber State University 2010-2013

Walter A. Shewhart (1891-1967), while working at Western Electrical Company, was an early pioneer in statistical quality control, which has been used to dramatically improve the quality of manufactured goods for many years. Shewhart recognized that the measures used to assess quality did not remain constant and that these variations could be attributed to randomness within the process—also referred to as noise—and significant changes in the process—also referred to as signal; that noise should be ignored and signals require corrective action. To separate the signal from the noise, Shewhart developed two control charts.

Control Chart using Individual Values

The first control chart monitors the individual values over time and establishes an upper control limit (UCL) and a lower control limit (LCL) for the process. This control chart is created by plotting the individual values with time on the horizontal axis and with the UCL, the average, and the LCL drawn as horizontal lines. The average forms the centerline of the chart midway between the UCL and LCL.

When an individual value falls above the UCL or below the LCL, the variation in the value is greater than that which can be explained by randomness (noise) in the process, and one should look for an assignable cause for the variance. Conversely, if an individual value falls within the range established by the control limits, there is nothing to indicate that the variation is due to anything but randomness in the process; therefore, no action is needed. The control limits are calculated as follows:

$$UCL = \bar{X} + (2.66 \times \overline{mR})$$
$$LCL = \bar{X} - (2.66 \times \overline{mR})$$

where

$\bar{X}$ is the average of the data used to establish the control limits

$\overline{mR}$ is the average moving range—the average of the change between pairs of data points

2.66 is a constant that represents three standard deviations from the average

The average moving range is calculated as follows:

$$\overline{mR} = \frac{\sum_{y=2}^{n} \left| \bar{X}_y - \bar{X}_{y-1} \right|}{n-1}$$

Before graduating, students from Weber State University's Parson Construction Management Technology (CMT) program are required to take the Associate Constructor (AC) Level 1 exam. The AC exam is an entry-level exam for construction management professionals given by the American Institute of Constructors and the Constructor Certification Commission. The exam consists of ten subsections and is given in the fall and the spring (see www.professionalconstructor.org/certification for more information). Students must achieve a score of 210 out of 300 to pass the exam. The exam results for fall of 2010 through fall of 2013 are shown in Table 1.

Table 1

Exam results

Test Date	Fall 10	Spring 11	Fall 11	Spring 12	Fall 12	Spring 13	Fall 13
Program Average	236.11	233.85	228.44	230.83	231.88	223.00	208.50
National Average	214.40	214.47	210.27	210.59	209.64	208.93	204.16
Candidates Tested	44	54	9	23	17	19	12
Candidates Passed	38	48	7	18	14	13	6
Percent of National Average	110.1%	109.0%	108.6%	109.6%	110.6%	106.7%	102.1%
Pass Rate	86.4%	88.9%	77.8%	78.3%	82.4%	68.4%	50.0%

The results for the first five exams (fall 2010 to fall 2012) were used to establish the control limits for the AC exam. The average is calculated as follows:

$$\bar{X} = \frac{236.11 + 233.85 + 228.44 + 230.83 + 231.88}{5} = 232.22$$

The moving range

$$mR = \left|\bar{X}_y - \bar{X}_{y-1}\right|$$

measures the change from one period to the next and is expressed as a positive value. The moving ranges for the first five exams are calculated as follows:

$$mR_2 = |236.11 - 233.85| = 2.26$$
$$mR_3 = |233.85 - 228.44| = 5.41$$
$$mR4 = |228.44 - 230.83| = 2.39$$
$$mR5 = |230.83 - 231.88| = 1.05$$

The average moving range is calculated as follows:

$$\overline{mR} = \frac{2.26 + 5.41 + 2.39 + 1.05}{4} = 2.78$$

Next, the control limits for the program average are calculated as follows:

$$UCL = 232.22 + (2.66 \times 2.78) = 239.61$$
$$LCL = 232.22 - (2.66 \times 2.78) = 224.83$$

A Shewhart control chart for the program average is shown in Figure 1.

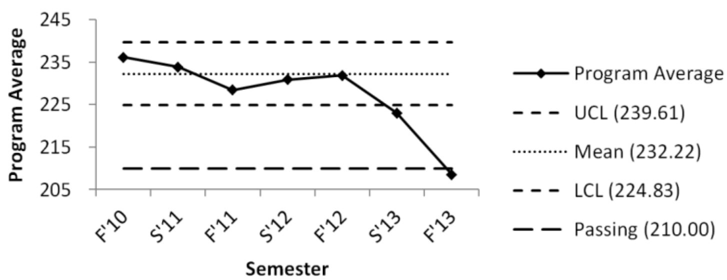

Figure 1: Shewhart Control Chart for Program Average

The LCL allows faculty to establish a benchmark for performance based upon the past performance—also referred to as voice—of the educational process. When implementing outcomes, the CMT program set a goal that the program's average score on the AC exam would be above the pass rate of 210 and above passing on each of the ten subsections. Although these were worthy aspirations, they were arbitrarily set without any consideration to the voice of the educational process. Figure 1 shows that the spring 2013 exam fell below the LCL and above passing; and the fall 2013 exam fell below the LCL and below passing. The calculated LCL indicates that the program should have set the benchmark for the program average at 224.83 rather than 210. Had this been done, the faculty would have recognized that the program average for the spring 2013 exam (223.00) fell below the LCL (224.83); and that they should have begun looking for an assignable cause because the program average was lower than could be attributed to randomness in the educational process. Using the pass rate created a false positive. In spring 2013, faculty failed to recognize a significant decline in student performance.

A Shewhart control chart for the program average for the communication skills section of the AC exam is shown in Figure 2 and is expressed as a percent of the passing score. The percent of passing was used because the number of questions varied from test to test. The program's goal was to be above passing (100%). Figure 2 shows that the fall 2011 and fall 2013 exams fell below the passing rate, but above the LCL of 77.8 percent. This control chart indicates that the educational process was producing the same quality of student learning as it had in the past; in other words, there was no indication that a decline in the quality of the educational process had occurred in this area, despite the fact that the program average fell below the goal (being above the pass rate) twice. Using the passing rate created a false negative; faculty thought there was a decline in student performance when the decline could have been attributed to randomness in the educational process. If faculty want to keep the goal of being above the passing rate on communication skills, a change needs to occur in the educational process so that the LCL moves above the passing rate.

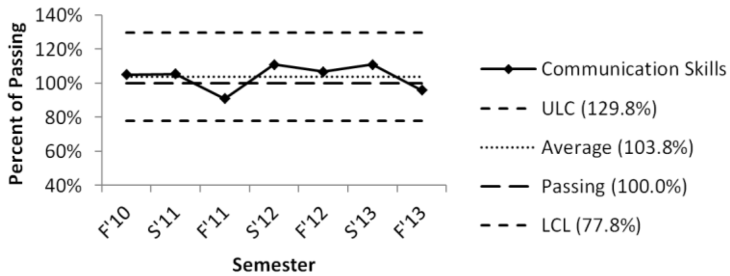

Figure 2: Shewhart Control Chart for Program Average for Communication Skills

The UCL allows faculty to recognize when the educational process produces student learning greater than expected and indicates that the educational process is improving. An increase in measurement that falls within the UCL and LCL may be due to nothing more than random variation, and faculty cannot say that the educational process is improving, even if the scores are improving. However, when a measurement rises above the UCL, there is almost certainly an assignable cause that produced the change, and faculty should try to identify that cause so that it may be repeated in the future.

Besides being above the UCL or below the LCL, there are four other events that are commonly used as signals that indicate a change has occurred that cannot be attributed solely to randomness. For the purpose of Shewhart control charts, a signal is an indication that something other than random variation has likely occurred.

First, a signal occurs when eight or more successive data points are on the same side of the centerline of the chart (Wheeler, 2000, p. 66; Western 1956, p. 27). If the variation in measurement is due to randomness, it is expected that the measurements will move back and forth across the centerline. Having eight successive values on the same side of the centerline is an unlikely event, unless there has been an assignable cause that has changed the process.

Second, a signal occurs when three of four consecutive data points are closer to the control limits than the centerline (e.g., between 1.5 and 3 standard deviations) and on the same side of the centerline. Given that 85 to 90 percent of the data should fall in the middle half of the region established by the control limits, it is rare to have three of four data points fall on the same side of the centerline and be closer to the control limit than the centerline (Wheeler, 2000, p. 57).

Third, a signal occurs when four of five consecutive data points are beyond 1 standard deviation and on the same side of the centerline (Western 1956, p. 27).

Fourth, a signal occurs when two of three consecutive data points are beyond 2 standard deviations and on the same side of the centerline (Western 1956, p. 27).

When any four of these signals occur above the centerline, it indicates an improvement in student performance; any four below would indicate a decline. If an improvement or decline occurs, one would need to look at other data (such as student ACT scores) to determine if the assignable cause is due to a change in teaching, a change in the aptitude of the students, or both.

Figure 3 shows a control chart for the scores on Midterm 1 for a building codes class. The control limits were established using the first five scores. Three of the four remaining scores fall closer to the UCL than the centerline, indicating that faculty should look for an assignable cause. Beginning spring 2012, the instructor had increased the number of quizzes the students had to complete before taking the midterm in hopes of improving student performance on the midterm. Here, the control chart was used to confirm that a positive change in student learning had occurred.

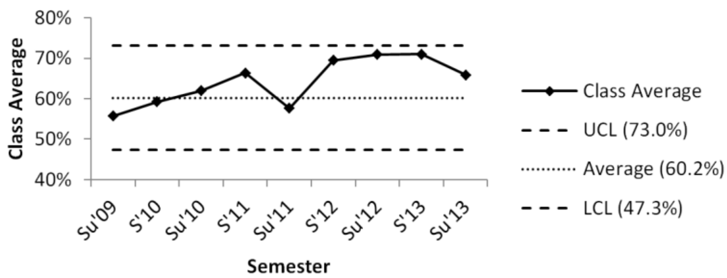

Figure 3: Shewhart Control Chart for Midterm 1 Scores

Control Chart using Moving Range

Shewhart also recognized that a major change from one data point to the next—such as falling from just below the UCL to just above the LCL—could not be attributed to randomness in the process. To address this, an upper range limit (URL) needs to be established, which sets the maximum change that can occur from one data point to the next. If the change from one data point to the next is less than the URL, faculty cannot say that the change is anything more than randomness within the process, unless another type of signal is found (e.g., the individual value is above the UCL). If the change is greater than the URL, there is almost certainly an assignable cause, and faculty should be looking for that cause.

The URL is calculated as follows:

$$URL = 3.27 \times \overline{mR}$$

where 3.27 is a constant.

Because this control chart is based on the moving range—which is the absolute value of the change from one period to the next—all values will be positive or zero. For this chart, there is no lower limit because the smallest possible number—a zero representing no change from the previous period—would not constitute a signal.

The upper range limit for the program average for the AC exam is calculated as follows:

$$URL = 3.27 \times 2.78 = 9.09$$

The second of Shewhart's control charts monitors the moving range—the change from one data point to the next. The control chart is created by plotting the moving range with time on the horizontal axis and with the URL drawn as a horizontal line. The Shewhart control chart for the moving range for the AC exam is shown in Figure 4. From

Figure 4 it can be seen that the change between fall 2012 and spring 2013 is below the upper range limit, and from this chart, faculty cannot say that the amount of change is due to anything more than randomness in the process. It is important to note that a signal may appear on only one of the control charts, so it is important that both control charts be used. For the spring 2013 exam, a signal occurs on the control chart using individual values but not on the control chart using the moving range. The change between spring 2013 and fall 2013 falls above the upper range limit and signals to faculty that they should be looking for an assignable cause.

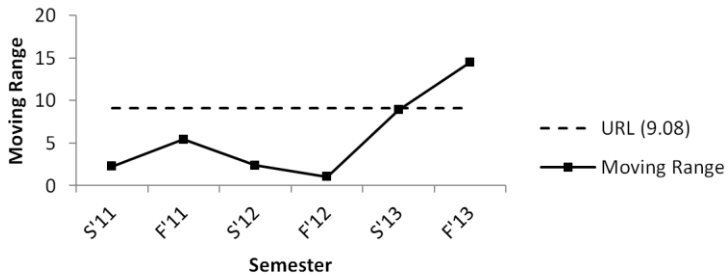

Figure 4: Shewhart Control Chart for Moving Range

There are a number of other ways that Shewhart control charts may be used to analyze the data. Control charts may be created for the program score expressed as a percentage of the national average or the percentage of the students passing the exam.

Establishing Control Limits

One of the most difficult aspects of using Shewhart control chart is establishing control limits. "Useful limits may be constructed with as few as five or six [data points]" (Wheeler, 2000, p. 60). As the number of data points used to establish the limits increases, the uncertainty in the limits decreases. Increasing the number of data points may move the control limits closer together or move them apart.

All of the data points used to establish the limits should fall within those limits. If any of the points fall outside the limits, it indicates that there (1) has been a change in the educational or assessment process and the data points before the change should not be used because they are not representative of the current process or (2) that the educational or assessment process is not repeatable, in which case the faculty should work on improving the consistency of the educational and assessment processes. Shewhart control charts require a consistently repeatable method of measurement (assessment) if they are to be of any use. A repeatable assessment method can be achieved by normalizing a test, using grading rubrics to ensure consistent grading, and so forth.

Another potential problem when establishing control limits is that the limits are too far apart to be of any use (e.g., the UCL is 100% and the LCL is 50%). When this occurs, the faculty need to work on improving the consistency of the educational and assessment processes rather than working on improving student learning, because there is so much randomness in the process. As a result, faculty cannot separate actual improvement in student learning from random changes in the student scores.

When a change (improvement or decline) in the educational process has been maintained over a period of time, the control limits should be recalculated using only the data points from after the change. This is done to ensure that the current control limits are representative of the current educational process. In Figure 3, positive change occurred beginning with the spring 2012 semester. Figure 5 shows new limits that have been calculated, using the data from the spring 2012 to summer 2013 semesters. The data before spring 2012 was not used in the calculation of the limits because it occurred before additional quizzes were included in the course and the resulting improvement in the students' scores on Midterm 1.

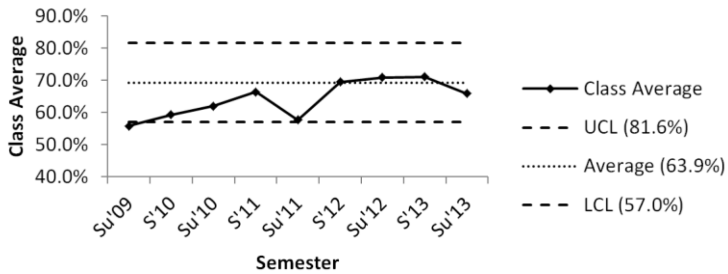

Figure 5: Shewhart Control Chart for Midterm 1 Scores

Limits and Weaknesses

The biggest drawback to using Shewhart control charts in academia is that it requires at least five data points (often equated with semesters or years) to establish meaningful limits, which could take two to five years to collect. To help minimize this weakness, data needs to be collected every semester (if possible), which requires more frequent sampling than is often required for accreditation.

Shewhart control charts require a consistent assessment method and works best with programs that use national certification exams that have been normed and validated. However, control charts can be useful when programs develop and implement a consistent assessment tool for a course or a program.

Like all statistical methods, the results are dependent on the data used in the analysis. In Figure 1, the spring 2013 midterm fell just below the LCL. Had there been more historical data, this may not have been the case.

Implementation

The implementation of any quality control measure in an organization needs to be accompanied by a change in culture. It is natural for people to resist change, especially with change that they see as threatening to their careers. As higher education moves to measuring and analyzing outcomes, it is important that faculty feel that it is safe to report data that indicates there has been a decline in the educational process. When outcomes are used as part of faculty performance, promotion, and tenure reviews, this puts pressure on the faculty to always meet the required standard. This pressure can lead to distortion of the data (e.g., using easier tests/assignments to measure student performance or more lenient grading of student performance) or distortion of the process (e.g., teaching to the test). The focus of measuring and analyzing outcomes should be on using the data to continually improve the educational process and should lead to discussions that ask such important questions as

1. Does the tool (test or assignment) we are using accurately measure student performance? How can the tool be improved?
2. What are the assignable causes of the signals in the data? Should these causes be addressed, and, if so, how should they be addressed?
3. What changes can be made to the curriculum to improve student learning?

With respect to outcomes, faculty performance should be judged based on their engagement in the process of continual improvement (collecting and analyzing the data and looking for ways to improve the educational process), rather than meeting the outcomes. Faculty need to be reassured, by both word and deed, that it is acceptable to not meet an educational outcome as long as they are looking for ways to improve.

Conclusions

As higher education implements outcomes, faculty need a method of accurately measuring the current state of the educational process and determining when changes (both positive and negative) have occurred. Shewhart control charts provide one means of doing that.

Shewhart control charts provide the following advantages over selection of an arbitrary goal:

1. Rather than having a simple pass/fail criterion (faculty meet their goal or they did not), Shewhart control charts have a number of events or signals that indicate when a change in student learning has occurred and may be used to identify both an improvement and a decline.

2. Shewhart control charts separate random variation (noise) within the system from a statistically significant event (a signal that something has changed).

3. The benchmark, or threshold, for student learning is set based on the voice of the current education process.

4. Faculty can determine if they need to work on the repeatability of the process or improving the outcomes of the process. Improvement in the educational process can be shown by (1) improving the repeatability of the process (i.e., tighter control limits) or (2) improved student learning outcomes (i.e., a shift in the average).

5. By allowing faculty to identify the current state of the educational process, Shewhart control charts can help faculty avoid a false negative (reacting to a decline in student performance that is just statistical noise) and a false positive (failing to identify a change in student performance).

Shewhart control charts have a well-established track record of identifying variation in processes and are a useful tool for identifying variation in the educational process. Once variation has been identified, faculty can begin looking for the assignable causes of that variation and how to best address it.

Recommendations for Future Study

This paper details the use of Shewhart control charts in one program at one university and will hopefully spark a discussion of how to use Shewhart control charts to analyze student learning outcomes. The next step would be to implement Shewhart control charts at other universities and in other programs (outside of construction), reporting the findings, including lessons learned, as case studies. As the number of case studies increase, researchers can begin to identify the best practices for using Shewhart control charts in an educational environment.

Bibliography

Western Electric Company, Inc. (1956). *Statistical Quality Control Handbook*. Easton: Mack Printing Company.

Wheeler, Donald J. (2000). *Understanding Variation – The Key to Managing Chaos*. Knoxville: SPC Press.