Using Staged Control Charts for Educational Assessment

Chen Xie[a], Theodore T. Allen[b] and Alah Raqab[c]

[a]The Ohio State University, 1971 Neil Avenue, 210 Baker Systems, Columbus, Ohio 43210
[b]The Ohio State University, 1971 Neil Avenue, 512 Baker Systems, Columbus, Ohio 43210
[c]The Ohio State University, 1971 Neil Avenue, 512 Baker Systems, Columbus, Ohio 43210
Email: [a]xie.72@osu.edu
[b]allen.515@osu.edu
[c]araqab@gmail.com

Abstract

This article proposes the application of "staged control" charting used in industrial quality improvement to educational intervention assessment. We illustrate staged control charting using two education-related case studies. The first case study involved a transition to blended learning instruction from traditional in-person instruction. The second study related to the profitability of industrial projects as influenced by extension instruction. We compare the insights gained from staged control charting with those from nonparametric hypothesis testing and two samples t-testing. We conclude that staged control charting offers enhanced intuition through visual trend analysis and a rigorous process for reducing the possibility that conclusions are inappropriately biased by assignable causes.

1. Introduction

Assessment is widely used in education and the methods for assessment are also a subject of research[1]. Assessment takes place in the engineering education improvement procedure at both the program level and the class level[2]. Often, assessment occurs because educators want to know if their specific educational activities are effective using objective and/or student perception measures. The purpose of this article is to propose a new assessment technique called "staged control" charting and to clarify its possible advantages compared with alternatives.

Several definitions of educational assessment have been proposed, and the key components of assessment definition have been divided into three parts[3]: (1) a statement of educational goals either for the entire class or for individual students, (2) measures of achievement of the goals, and (3) use of the data to improve the education process. Because the reasons to perform assessment are varied, finding an appropriate assessment tool is a challenge in designing a specific assessment approach. In our first case study, we focus on entire class improvement goals at the level of the instructor.

Also, we use the term “assessment” in a broader sense than measuring individual student’s competencies, such as scores on a classroom exam or homework assignments. That type of assessment is important and might be viewed as a type of “quality” assurance for setting specification limits related to specific students. For us, however, assessment also relates to system improvement in relation to evaluating class overall quality levels and major performance improvement initiatives. The possible evolution of assessment from control only to system improvement seems to mirror the transition within the larger “quality” movement which spans multiple industries[4].

In the remainder of this section, we further describe alternative assessment techniques. In Section 2, we describe the staged control charting method and discuss its possible advantages within the educational quality control and improvement contexts. Section 3 includes two case study-evaluations of both staged control charting and alternatives relating to both students’ grades and also the profits of industrial projects influenced by educational interventions. Section 4 summarizes the conclusions and possible topics for future research.

1.2 Types of Assessment Methods

Mcgourty et al. (1998)[5] categorize assessment methods into groups: qualitative methods and quantitative methods and mix of both. The authors have studied a comprehensive framework for assessing engineering education. Leydens et al. (2004)[6] reviewed the previous studies about different qualitative methods in assessment. Those qualitative methods are used in both the data collection and analysis phases. In this article, we focus on quantitative assessments and the analysis phase.

1.3 Assessment Frame Work

Mcgourty et al. (1998)[5] argue that clarification of educational goals is an important first step in developing assessments. Yet, in control chart monitoring, there might not need to be a specific goal other than measuring the system and identifying unusual occurrences (Allen, 2010 p. 88)[7]. Mcgourty et al. (1998) proposed a comprehensive assessment process composed of five steps. They are: (1) Define Objectives, Strategies and Outcomes, (2) Identify Assessment Methods, (3) Develop and Pilot Assessment Processes, (4) Implement/Expand Assessment Processes, (5) and Apply Results.

This five-step process is similar to the “define, measure, analyze, improve, and control (DMAIC)” method in the Lean Six Sigma[7]. In fact, we argue that education shares many similarities with other forms of production systems. Therefore, we propose to transfer the analytical techniques that are applied in Lean Six Sigma process improvement to engineering education. In the following section, we describe the definition of the staged control charting method, and its possible applications in educational assessment.

2. Description of Staged Control Chart Method

In general, statistical process control has been proven in manufacturing to be an effective method for improving a firm's quality and productivity[8]. Perhaps the primary tool of statistical process control is the control chart[8]. The charted quantities in control charts are often called "quality characteristics". In education, these could relate to measures of students' grades or measures of achievement outside of the class room such as profits on real-world projects. The control chart is a graphical display of quality characteristics used for monitoring and system evaluation. Control charts were introduced by Shewhart in the 1920s while working for Western Electric and Bell Labs and, since then, they have been routinely used in manufacturing. For detailed descriptions of these charts and extensive annotated examples, see Juran and Gryna (1999)[4], Montgomery (2008)[9], and Allen (2010)[7].

In general, a control chart contains a center line (CL) that represents the mean value for the in-control process of the charted statistic, which could be the average value of a "subgroup" or an individual value. Subgroups are sets of observations that are generally chosen to be representative of performance during a time period. For example, every 10th student scores on the first and second homework assignment could form a subgroup.

Two other horizontal lines, called the upper control limit (UCL) and the lower control limit (LCL) are also shown on the chart. These control limits are chosen so that, if there is nothing unusual shifting the process, the charted quantities will usually remain between the limits. Conversely, if there is an assignable cause shifting the process, a signal will likely occur because the charted quantity will go outside the control limits. Figure 1 illustrates the related concepts including charted quantities outside the limits representing "out-of-control" signals.

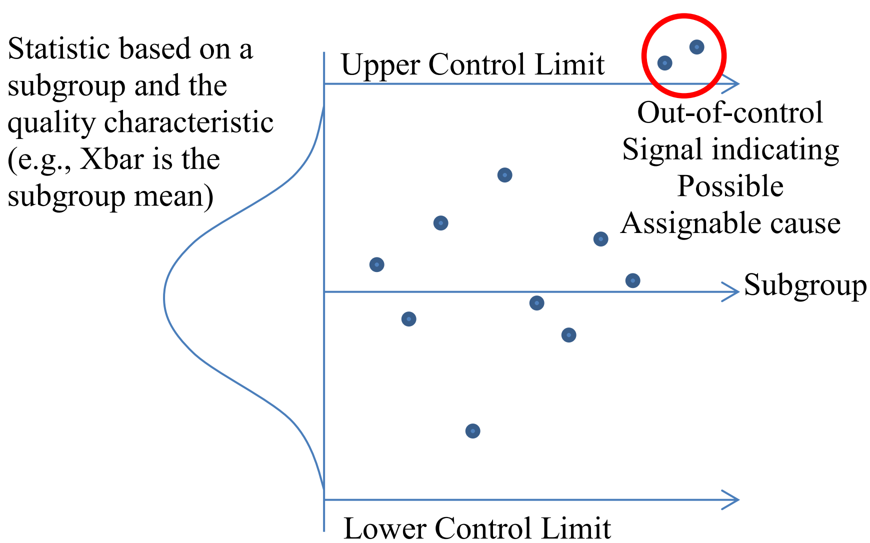

Figure 1. Depiction of a control chart with the reference population (distribution of the charted quantity if there are no assignable causes).[10]

Note that assignable causes sometimes require the attention of local authority or might simply indicate mistakes in the data collection system which could make all assessments misleading.

Also, the overall quality level of the process might not relate in a simple way to monitoring using control charts because the monitoring intent is primarily for the identification of assignable causes that might be fixable by local authority, e.g., the instructor or department chair, without major investment, e.g., through work at the dean's level or through large committees. The chart in Figure 1 could hypothetically indicate that there might be an assignable cause that affected the last two subgroups plotted (e.g., an instructor might need assistance).

Alternative types of control charts have been reviewed in Leydens et al.[6] (2004). The most popularly used one is from Shewhart (1931). Often, in Shewhart charting, the subgroups include only 4 or 5 data points. For example, an academic program might randomly sample five students and pay them to take a test to obtain a subgroup for program evaluation. In general, so-called "three sigma" control limits are used so that the chance that a signal will signal that the process is out-of-control wrongly is approximately 0.0026. Shewhart charting actually involves making two separate charts of the subgroup average value and the range of observations.

A major part of constructing control charts is the inclusion of a "trial phase" during which the limits are provisional. Usually 25 subgroups of data are taken and the initial or "trial" chart is made. Then, any assignable causes are identified such as mistakes in data entry or unusual occurrences that are rare and not part of the system. The associated data are removed if they are deemed not representative of the system and the process is evaluated for stability and assignable causes. Finally, the "revised" limits are determined and the steady state phase begins in which there is monitoring but generally no further limit revisions[7].

From our inspection of teaching related publications, it seems that what might be called "traditional" intervention assessment is based on statistical hypothesis testing such as t-testing and non-parametric testing[Error! Bookmark not defined.]. Apparently, the traditional assessment will treat the education objects original performance as the control group and take the performance after certain education methods been applied as the treatment group. Comparing these two groups' performance gives educators an indication about whether the intervention is effective.

Yet, t-testing and nonparametric evaluation might fail to include some of the benefits of the control chart framework. These benefits could, of course, be obtained if both types of methods were applied together. Yet, t-testing and nonparametric evaluation applied without control charting fail to evaluate:

- The stability of the process before the intervention,
- The stability of the process after the intervention,
- The typical spread of the quality characteristic when the process is stable, and
- A visual indication of trends in the quality characteristic.

Control charting offers all of these benefits. Staged Control Charts (SCCs) are a simple extension of control charting in which the trial phase is performed again after a known intervention. Then, the charting limits are changed and the before and after charts are plotted together. Therefore, the purpose of staged control chart is to compare the system performance after implementing a major change. In the case studies that follow, we use staged control chart to compare between two types of educational systems to show the difference between them and observe the improvement. In the table below we show some comparisons between the traditional statistics tests and the staged control chart methods.

Table 1 Comparison of alternative statistical assessment methods including t-testing, nonparametric hypothesis texting, and staged control charting (SCC).

	t-testing	Nonparametric hypothesis testing	SCC
Formal hypothesis test of subgroup means or medians	X	X	
Includes a trial phase in which unusual or "assignable causes" are investigated to make sure the system quality measurement is fair.			X
Facilitates the identification of assignable causes and process capability.			X
Facilitates the visual investigation of process trends.			X

3 Case studies

In this section, we describe two applications of the alternative assessment techniques relating to real-world case studies. The first relates to a transition to online education of a discrete event simulation laboratory and the second relates to instruction of practicing engineers.

3.1 The first case study

These data are described in Allen, Artis, Afful-Dadzie, and Allam[11] (2013). In that document, those authors explain that the traditional and online cohorts were run simultaneously and the assignment of students to cohorts was performed using a random sampling. Here, for simplicity, we imagine that the transition was traditional during the first time period and online during a second time period. This scenario would be typical of the transitions that are occurring at many universities.

Therefore, assume that there was a transition of course from traditional in-person training on ARENA software to blended learning with almost all instruction on-line. The grading policies and assignments were similar and intended to be at a constant level of difficulty. Also, fewer students were selected for the traditional group in part because students generally preferred being selected into the online group. In addition, we had several students opting out of the study who populated the traditional labs placing a space constraint on the control group.

Table 2. Data showing the de-identified student grades for students taking a simulation laboratory course in either the traditional or online modes.

No.	Stage	Item	Size	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	Trad.	HW 1	11	94	95	100	97	100	97	97	94	94	90	93
2	Trad.	HW 2	11	91.4	99.3	82.9	99.3	100.0	82.9	90.0	97.9	95.0	70.0	83.6
3	Trad.	HW 3	11	90	97	99	97	100	97	89	98	94	51	92
4	Trad.	Quiz 1	11	71.5	78	86.5	85	95	87.5	84.5	89.5	75.5	90.5	75.5
5	Trad.	Quiz 2	11	87.5	84.5	89	90.5	95.5	83	73.5	95	92	91	77
6	Online	HW 1	22	98	96	100	92	97	93	97	85	97	97	94
7	Online	HW 2	22	100.0	93.6	83.6	92.1	95.7	83.6	96.4	82.1	91.4	97.9	97.9
8	Online	HW 3	22	94	97	92	95	98	97	97	93	100	92	99
9	Online	Quiz 1	22	83.5	91	81	78.5	77.5	88	74.5	90	93	93	92
10	Online	Quiz 2	22	92.5	86	85	74.5	78	92	99	87	91.5	87.5	91

No.	Stage	Item	Size	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22
1	Trad.	HW 1	11											
2	Trad.	HW 2	11											
3	Trad.	HW 3	11											
4	Trad.	Quiz 1	11											
5	Trad.	Quiz 2	11											
6	Online	HW 1	22	71	77	100	97	86	97	100	91	93	97	100
7	Online	HW 2	22	55.7	92.9	99.3	96.4	64.3	95.0	90.0	87.9	94.3	96.4	100.0
8	Online	HW 3	22	68	95	97	95	91	94	94	98	90	97	97
9	Online	Quiz 1	22	76.5	82	84	83.5	84.5	81.5	92	89.5	83	94.5	98.5
10	Online	Quiz 2	22	82	83	80.5	90.5	87	88.5	87.5	82.5	84.5	95.5	100

Using the traditional statistical analysis of the data collected for different assignment's grades, we run the t-test at 0.05 significant level to compare two sample mean and see if there is a true difference between them. T-test did not give us a significant difference between the sample grades of online and traditional group. We also conducted three non-parametric tests, Wilcox rank sum test, Wilcox signed rank test, and Kruskal Wallis test. The hypothesis of there is true difference between these two group's means is rejected at significant level of 0.05. Although the sample sizes are not large, usually a sample size as 30 would be a good instance for sample t-test, this result is consistent with previous studies done by Haag and Palais[3] (2002). They also found out that the grades for the web based teaching and non-web based teaching have no significant difference. The results of the statistical tests we conducted in our case study are shown in Table 3.

We also ran a two way ANOVA table to test which factor that really influences on the student grades. As most of the statistical analysis used, we chose 0.05 as the significant level p-value, meaning if a factor has a p-value smaller than 0.05 or close to it, we will conclude it as a significant factor. We can see the assignment factor has a fairly significant P-value which means it has impact on the grades. So this result shows that the difference among the assignments do influence on student's grade. The reason might be the assignments are at different difficult levels. However the online or traditional indicator doesn't show significance as a factor, which is also consistent with the previous statistical tests. Therefore, although the online group shows a slight higher average grade than traditional teaching group does, it doesn't mean that online education will generate a better average student grades. From this two way ANOVA table we can barely tell anything about whether or not online education will help with student's grade. The results of this ANOVA table are shown in Table 4.

Table 3. Traditional assessment measures from the discrete event simulation student grade study evaluating the transition from traditional to online education.

	Mean	Standard Deviation	P-value			
			T-test	Wilcoxon rank sum test	Wilcoxon signed rank test	Kruskal Wallis Test
HW1 Traditional	89	24.57	0.96	0.98	0.84	0.59
HW1 Online	89	21.45				
HW2 Traditional	84	24.60	0.77	0.79	0.50	0.71
HW2 Online	86	21.16				
Hw3 Traditional	85	26.67	0.61	0.64	0.17	0.77
Hw3 Online	90	21.47				
Quiz1 Traditional	78	22.10	0.74	0.47	0.35	0.38
Quiz1 Online	80	19.50				
Quiz2 Traditional	81	22.99	0.88	0.86	0.79	0.83
Quiz2 Online	82	20.05				

Table 4. Two way ANOVA results

	Degree of Freedom	Sum Sq	Mean Sq	F value	Pr(>F)
Assignment	1	1071	1071.2	2.963	0.087
Online or Tradition	1	727	727.2	2.011	0.158
Interaction	1	30	29.7	0.082	0.775
Residuals	171	61821	361.6		

In general, the traditional method does not offer much illumination about how much and whether the intervention helped and whether there are any assignable causes of interest. This follows because the p-values were higher than 0.05 and no significance was found. Hence, we wanted to try out another method-staged control chart methods-to really test out if there is no impact from online instruction to the students' grade.

Figure 2 shows the application of Shewhart X-bar and R charting to create staged control charts. The apparent benefit of using the chart compared with the statistical tables is an improved intuition about the effect of the transition. Immediately the apparent result is, confirmed by the traditional evaluation, that the student grades changed little. We can see that the online limits are narrower which indicates the larger sample size for the online students (22 participating in our study compared with 11 traditional students). Also, we can see that the grade means follow a pattern indicating that the assignments vary in difficulty, which is consistent with the traditional ANOVA table conclusion. This provides feedback to the instructor that the additional efforts to balance the difficulty level might improve the grading stability.

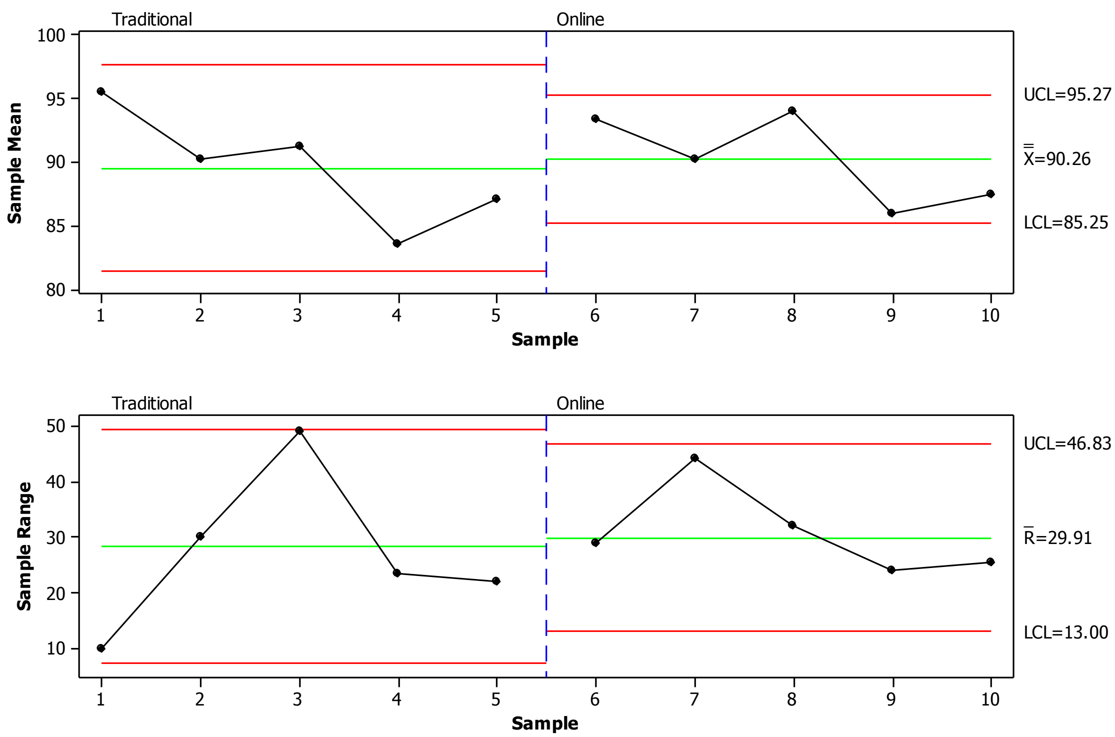

Figure 2. The staged control chart relating to the traditional to online transition and sample mean and sample range grades.

3.2 The second case study

This involves a reanalysis of the project data from Allen, Brady, and Schenk[12] (2011). As described by those authors, the project number 5 differed from the others in that the manufacturer was allowed to change the design given by the customer. Therefore, we removed this project from the analysis. In between projects 31 and 32, many of the engineers were given instruction from a local community college. Using staged control charts, we can estimate that this instruction was worth approximately $50,000 per project. This follows because, pre-instruction the average profit per project was near zero dollars. After the instruction, the average jumped to $55,807. The data is shown in Table 5.

Table 5. The pre and post instruction profits from a collected 39 lean six sigma projects.

Project	Stage	Profit	Project	Stage	Profit
1	Pre	-12700	21	Pre	3025
2	Pre	-7590	22	Pre	3025
3	Pre	-3800	23	Pre	-400
4	Pre	-2900	24	Pre	5400
5	Pre	3874500	25	Pre	-1900
6	Pre	94000	26	Pre	-1900
7	Pre	112775	27	Pre	2860
8	Pre	-220000	28	Pre	4800
9	Pre	66675	29	Pre	-1180
10	Pre	7225	30	Pre	3240
11	Pre	-9300	31	Pre	6650
12	Pre	-2600	32	Post	2975
13	Pre	-2000	33	Post	126540
14	Pre	14240	34	Post	9250
15	Pre	-13600	35	Post	53460
16	Pre	-2000	36	Post	15276
17	Pre	-5000	37	Post	7716
18	Pre	17920	38	Post	220590
19	Pre	6150	39	Post	10652
20	Pre	3025			

We conducted a nonparametric test for two sample data having different sample sizes. Here we conducted a Kruskal-Wallis rank sum test that takes the first parameter as the factor, with levels being pre and post stages. The response is the total project profit. The detailed test result is shown in Figure 3. The conclusion is that there is no significant difference of the project profit between pre and post stages of the project at significant level of 0.05.

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 38, df = 34, p-value = 0.292

Figure 3 Kruskal-Wallis rank sum test result for pre stage and post stage project cost

In our case, the projects were each important and their completions were staggered. Therefore, there was no natural subgrouping and we applied individuals control charting using the so-called individuals and moving range (I-MR) charting procedure.

The resulting trial phase control chart including all the projects is shown in Figure 4. We identified the out-of-control signal associated with project 5 with profit about $3,900,000. By investigating Allen, Brady, and Schenk[12] we found an assignable cause associated with the signal. This project was the only project for which the manufacturing facility was given control of the engineering design specifications. Therefore, we removed the associated data and declared the analysis scope to only include projects without design control. With project 5 removed, the limits were recalculated with the resulting chart shown in Figure 5.

On the Moving Range chart we can see there are two points with extremely high values, which are the difference between profit of project 4 and project 5, and between project 5 and project 6. This is consistent with what we have found in the individual chart that project 5 has an extremely high profit. This also shows project 5’s profit is an interruption of the process stability and we need to take those two values out of the moving range chart too. After eliminating the two difference values from the chart, we see a tightened range with upper limit of $351,757, lower control limit of 0 and mean difference in dollars is 107,660. The staged control chart for the trial phase is shown in Figure 4. The revised chart is shown in Figure 5.

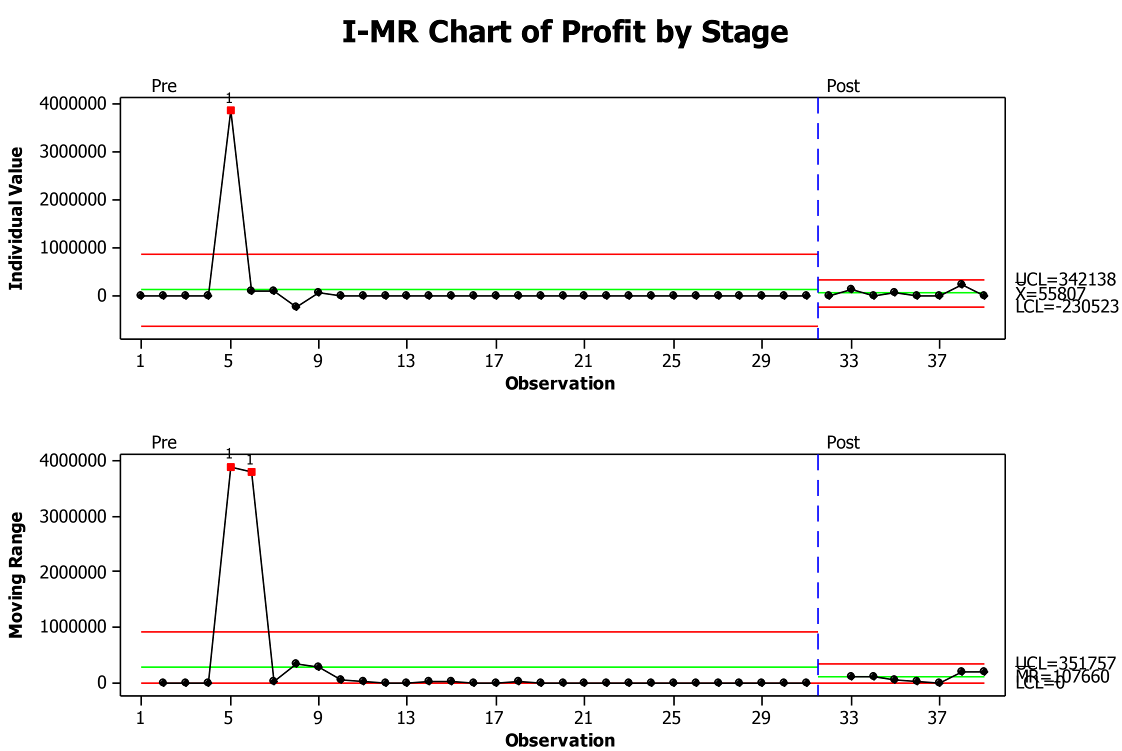

Figure 4. Trial individuals staged control chart of project profitability pre (before) and post (after) the educational intervention relating to extension training in design of experiments of practicing engineers. The out-of-control signal was an assignable relating to the only case in which design control was permitted and should be removed since design control is not considered relevant.

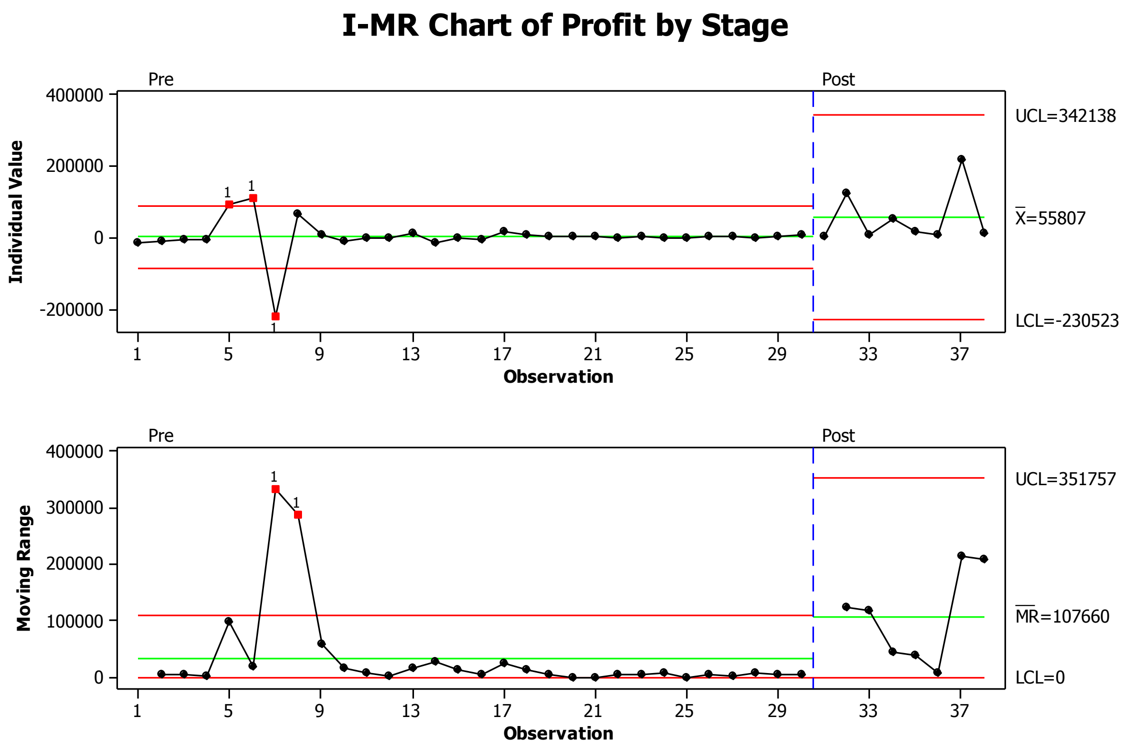

Figure 5. The revised staged control chart for evaluating the effect of the intervention on lean six sigma profitability. The positive effect on profitability is apparent in the shift of the center lines up by approximately $50,000 per project.

4. Discussion

This paper has proposed staged control charting for assessing the value of interventions in education including new methods of instruction (e.g., blended learning) or new programs (such as an extension course for practicing engineers). The staged control chart had been applied as an evaluation tool in the lean six sigma and quality control literature. Also, we demonstrate the additional value of stage control charting supplementing traditional hypothesis testing evaluations.

The results showed that the staged control chart offers additional insights and checks compared to traditional hypothesis testing. Note that the two methods might be combined. In the students' grade example relating to a transition to online laboratory instruction, the staged control chart offered an additional insight about the lack of balancing relating to assignment difficulty. In relation to the lean six sigma project profitability, staged control charting provided more trustworthy results. This is because an outlier relating to an assignable cause was flagged and interpreted. With this data removed, the positive effect on profitability was readily apparent of the extension education. That education improved average project profitability approximately $50,000 per project.

There are a number of topics for future research. First, the issue of subgrouping can be addressed more thoroughly for standard courses. Such investigations can consider whether it makes sense to group grades from different assignments and the impact of using all the grades instead of samples. Second, the objective of intentionally balancing the difficulty of assignments can be considered. One benefit of balancing assignment grades is that it would make the constant mean assumption implied in control charts to be appropriate. The example that we used related to blended learning. The instructor did intend for assignment grades to have the same mean values. Third, special types of control charts to address educational goals can be developed going beyond the Shewhart X-bar and R charts that we used in our examples. Such control charts could permit the efficient identification and elimination of assignable causes by instructors and/or supervisors.

References

[1] Shaeiwitz, Joseph A. Outcomes Assessment in Engineering Education. Journal of Engineering Education (July 1996), 239–246.

[2] Turns, Jennifer, Atman, Cynthia J., and Adams, Robin. Concept Maps for Engineering Education: A Cognitively Motivated Tool Supporting Varied Assessment Functions. IEEE Transactions on Education (May2000), VOL. 43, NO. 2.

[3] Haag, Susan. Palais, Joseph C., Engineering Online: Assessing Innovative Education. Journal of Engineering Education (July 2002), 285–290.

[4] Juran J M, Godfery A B (1999), Juran's Quality Handbook, 5th edition, McGraw-Hill: New York.

[5] Mcgourty, Jack., and Swart, Williams. Developing a Comprehensive Assessment Program for Engineering Education. Journal of Engineering Education (October 1998), 355–361.

[6] Leydens, Jon A., Moskal, Barbara M., and Pavelich, Michael J. Qualitative Methods Used in the Assessment of Engineering Education. Journal of Engineering Education (January 2004), 65–72.

[7] Allen, T. T. (2010), Introduction to Engineering Statistics and Lean Sigma, 2nd edition, Springer: London.

[8] Saniga , Erwin M. Economic Statistical Control Charts Designs with an Application to $\bar{X}$ and R Charts. TechnometricSA, U Cust 1989, VOL. 31, NO. 3.

[9] Montgomery D. (2008), Introduction to statistical quality control, 6th edition, Wiley.

[10] NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/-handbook/date.

[11] Allen, T. T., Artis, S., Afful-Dadzie, A. Allam, Y. (2013), "Case Study Application of Blended Learning for Delivering Discrete Event Simulation Laboratories," Technical Report – The Ohio State University, Integrated Systems Engineering.

[12] Allen, T. T., Brady, J. E., and Schenk, J. (2011), "Analysing Portfolios of LeanSix Sigma Projects, Six Sigma Projects," Prof. Abdurrahman Coskun (Ed.), ISBN: 978-953-307-370-5, InTech, DOI: 10.5772/16732. Available from: http://www.intechopen.com/books/six-sigma-projects-and-personal-experiences/analysing-portfolios-of-leansix-sigma-projects.