

Six Sigma Solutions

SCORECARDS

The Improvement Of Scorecard Management

Comparing Deming’s red bead experiment to red-yellow-green scorecards

by **Forrest W. Breyfogle III**

W. Edwards Deming is well known for helping Japan improve product quality in the 1950s. Deming also is acknowledged for the messages conveyed during his many four-day seminars. In these 1980s training sessions, Deming used a red bead experiment to highlight several poor management practices, which underscored aspects of his 14 points of management¹ and seven deadly diseases.²

Deming’s red bead experiment

When conducting the red bead experiment, Deming had audience recruits use a paddle to draw 50 beads from a container of red and white beads, as shown in the illustration. White beads were considered satisfactory while red beads were not.

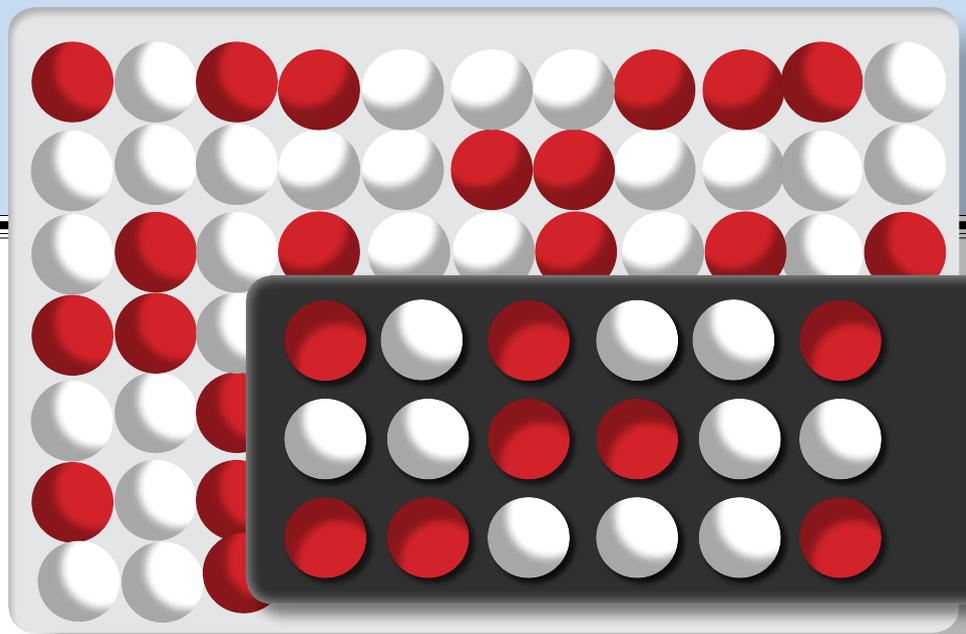
In this exercise, the number of red beads resulting from a paddle draw was reported. A count of red beads in a paddle drawn was to be instrumental in monitoring a fictitious organization’s overall product quality. In this tracking, between-worker quality of workmanship was highlighted in this exercise’s reporting, in which a lower number of red beads in a paddle draw was better.

In Deming’s workshop experiment, he announced the management goal of no more than three red beads for each paddle draw. The poorest-performing operators (those drawing beads from the container) relative to not meeting this minimum red bead number objective were “let go” from Deming’s organization company. The

TABLE 1

Number of red beads in 60 paddle draws

Paddle-draw number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of red beads	7	7	9	12	13	13	13	10	8	9	7	9	13	11	10	14	5	6	7	10	10	14	13	13	15	11	6	9	8	7
Paddle-draw number	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Number of red beads	10	5	7	11	9	8	7	10	12	10	5	8	8	9	14	7	5	7	10	10	10	5	15	13	8	7	4	9	11	8



organization was eventually shut down because none of the operators met the minimum number of allowable red beads for a paddle draw.

Red-yellow-green scorecards

Organizations often use red-yellow-green (RYG) scorecards to track individual performance metrics relative to goals. At regular times, an RYG scorecard colored response assesses how well a metric response performs relative to its goal:

- + Green—satisfactory.
- + Yellow—warning.
- + Red—not satisfactory.

If a metric's color is red, action must be taken to resolve why performance at this point is unsatisfactory. With RYG scorecards, when a green color follows a red-colored metric, management typically feels good because it believes that the red-issue problem was resolved.

Online Figure 1, found on this column's webpage at www.qualityprogress.com, illustrates an actual organizational RYG scorecard report-out.

From a management's scorecard point of view, a question that this column addresses is whether there are any basic differences from a management point of view between Deming's red bead experiment and commonly used organizational RYG scorecards.

If there are no fundamental differences between Deming's workshop exercise and RYG scorecards, the same lessons learned from Deming's experiment also would apply to the use of RYG scorecards in businesses.

Comparing methods using a data set

In the red bead experiment, 20% of the container's beads are red. Hence, on

average, 10 red beads are expected from a 50-hole paddle draw. Rather than having a goal of three or fewer for any draw (as Deming did in his workshops), let's set the goal to be 12 or fewer red beads for paddle draws.

Table 1 shows the number of red beads in 60 example paddle draws. This data will be used to compare the management techniques used in Deming's red bead experiment and RYG scorecard reporting.

For this set of red bead draws, a Deming red bead experiment management style would be telling operators to do better for the 12 paddle draws circled in Table 2—that is, whenever the number of red beads was 13 or more.

Colors for possible metric draws for a RYG scorecard of Table 2 data are:

- + Green: 11 or less.
- + Yellow: 12.
- + Red: 13 or greater.

TABLE 2

Paddle draws with Deming's red bead management style operations feedback

Paddle-draw number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of red beads	7	7	9	12	13	13	13	10	8	9	7	9	13	11	10	14	5	6	7	10	10	14	13	13	15	11	6	9	8	7
Paddle-draw number	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Number of red beads	10	5	7	11	9	8	7	10	12	10	5	8	8	9	14	7	5	7	10	10	10	5	15	13	8	7	4	9	11	8

TABLE 3

Paddle draws with a red-yellow-green scorecard management style operations feedback

Paddle-draw number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Number of red beads	7	7	9	12	13	13	13	10	8	9	7	9	13	11	10	14	5	6	7	10	10	14	13	13	15	11	6	9	8	7
Paddle-draw number	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Number of red beads	10	5	7	11	9	8	7	10	12	10	5	8	8	9	14	7	5	7	10	10	10	5	15	13	8	7	4	9	11	8

Table 3 shows the results of an RYG scorecard management metric reporting style.

With this RYG scorecard reporting approach, management again would be reacting to the 12 paddle draws containing 13 or more red bead paddle draws. In addition, management would probably think that the red occurrence problem was resolved whenever the metric color changed to green on the next draw.

A process output response can either

be the result of natural variation from the process or a special cause event. The container of red and white beads had 20% red beads. For this simplistic situation, we know that the variability from a paddle draw in this container should be common cause. In the red bead experiment and RYG situations, the management style was to react to any paddle draw that contained more than 12 red beads as though the situation were special cause (that is, the operator did

not perform satisfactorily).

If a process has an undesirable common cause response, the process must be improved. Talking about individual out-of-specification occurrences in isolation as though they were a special cause is an ineffective management practice.

Metric management with control charts

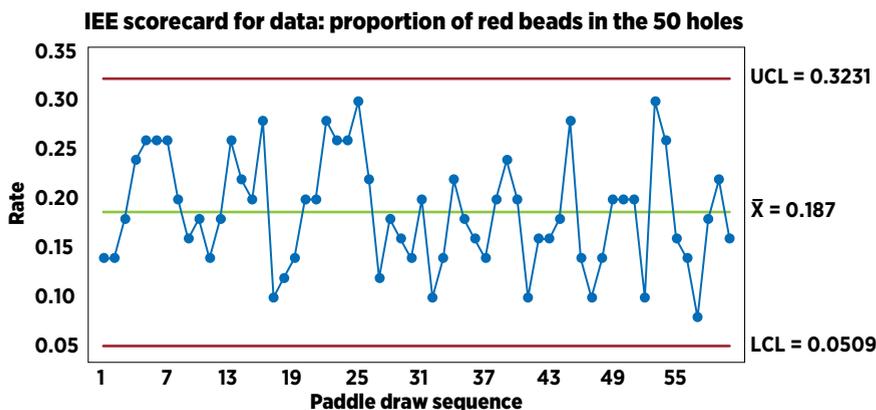
A traditional control charting approach separates common cause from special cause events with the focus on identifying special cause events so the issue can be resolved in a timely manner. If we considered the number of red beads on a paddle draw as the response, we might select a c-chart for this common-special cause event separation. Online Figure 2 shows a c-chart for the data in Table 1.

But is the c-chart technically the most appropriate chart for this number of red beads in a paddle situation? Because we know from this process that either a white or red bead can be present in any of the 50 holes, a p-chart that tracks over time the proportion of red beads relative to the number of 50-hole paddle opportunities would technically be more appropriate. A p-chart of the data from Table 1 is shown in Online Figure 3.

Online Figures 2 and 3 indicate that the process is in-control or stable. No statements about how a process is performing overall relative to specifications or goals should be made from a control chart. However, it does appear that the process is

FIGURE 1

IEE 30,000-foot-level proportion nonconforming report-out



The current process is predictable. The estimated process nonconformance rate is 0.187.

IEE = Integrated Enterprise Excellence
 LCL = lower control limit
 UCL = upper control limit
 \bar{X} = mean value

incapable of achieving the desired results.

From this control chart analysis, we also can conclude that the process management styles of the red bead experiment and RYG scorecard approaches were ineffective in the achievement of the goal of 12 red beads or fewer in a paddle draw.

From this assessment, we also could conclude that there is no fundamental difference between the red bead experiment and RYG scorecard approaches in that both management styles use an approach to management described as managing of individual occurrences that are not satisfactory relative to a goal.

An alternative approach

The data in Table 1 will now be reported out using an Integrated Enterprise Excellence (IEE) 30,000-foot-level metric reporting method.³⁻¹² This form of metric reporting consists of two steps:

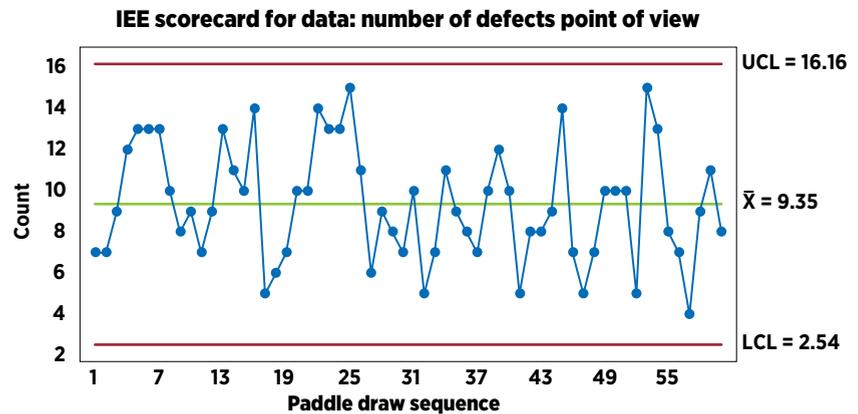
1. The process is assessed for stability from a high-level (30,000-foot-level) vantage point.
2. If the process is stable, a predictive process capability statement is made and reported at the bottom of the chart with easily understood wording.

To illustrate the versatility of this one-page process stability or capability reporting method, Table 1's data will be reported using the following approaches:

1. **Percentage nonconformance in a paddle draw**—With this method, you understand that each hole in the 50-hole paddle can contain either a red or white bead after a paddle draw. This reporting approach will mimic what would be expected as an IEE 30,000-foot-level report-out when the data are analyzed as percentage defective rate.
2. **Count of the number of red beads in a paddle draw**—Examine the data purely from a number of red beads in the paddle point of view, in which we will consider in this analysis that, technically, there could be a number

FIGURE 2

IEE 30,000-foot-level count of red beads report-out



The current process is predictable. The estimated average number of nonconformances is 9.35 per paddle-draw.

IEE = Integrated Enterprise Excellence
 LCL = lower control limit
 UCL = upper control limit
 \bar{X} = mean value

greater than 50 red beads for a given draw. This approach will mimic what would be expected as an IEE 30,000-foot-level report-out if the data were counts of defects from a process.

3. **The number of red beads represents a continuous response in which the goal is to have 12 or fewer red beads for every paddle draw**—A response is unsatisfactory if a draw response is greater than 12. This approach will mimic what would be expected as a 30,000-foot-level report-out if the data were examined as a continuous response—that is, from another process in which the data were truly a continuous response.

Percentage nonconformance report-out: Figure 1 shows an IEE 30,000-foot-level report-out when the data are considered a nonconformance rate. From this individuals chart of the mean, you would consider this process stable with an estimated rate of

nonconformance of 18.7%. If this common cause frequency of nonconformance estimate for paddle draws is unsatisfactory, the process must be improved.

Counting defects report-out: Figure 2 shows an IEE 30,000-foot-level report-out when the data are considered counts of defects—that is, the number of red beads in a paddle. From this individuals chart, you would consider this process stable with an estimated average of counts of defects to be 9.35 for individual paddle draws from the container. If this common cause frequency of nonconformance estimate for paddle draws is unsatisfactory, the process must be improved.

Continuous response report-out: Online Figure 4 shows an IEE 30,000-foot-level report-out if the data from Table 1 are considered continuous. From the individuals chart on the left side of the figure, you would consider this process stable. The probability plot on the right is used to determine process capability/performance from a predictability point of view. For an

upper specification of 12, you would expect that about 17.276% of the time the paddle draw values would be above 12. If this common cause frequency of nonconformance estimate for paddle draws is unsatisfactory, the process must be improved.

IEE 30,000-foot-level reporting

Critiquing individual responses—like that done in the red bead experiment and with RYG scorecards—is ineffective management. From this practice, treating common cause variability as though it were special cause can lead to much firefighting, which can be costly, frustrating and detrimental to an organization.

IEE 30,000-foot-level charting not only assesses process stability from a high process vantage point, it also provides a predictive statement when the process is stable. If this predictive statement is unsatisfactory, this process measurement response pulls for the creation and timely completion of a process improvement project to enhance the performance metric's response.

Control charts are for controlling processes

Deming's red bead experiment illustrates how attempting to manage operators and process outputs through individual reported values is ineffective, which can extend to the use of RYG scorecards.

Control charts are used to control processes. The identification of special cause events is to trigger process resolution activities. Control charts are not to make any statement as to how a process is performing.

IEE 30,000-foot-level report-outs provide, from a high-level point of view, a statement about whether a process is stable. In addition, if a process is stable and its 30,000-foot-level process capability/performance statement is unsatisfactory, the process needs improvement.

Deming's red bead experiment illustrates how attempting to manage operators and process outputs through individual reported values is ineffective.

A 30,000-foot-level report uses an individuals chart when making a process stability assessment. No timely process control efforts are attempted with 30,000-foot-level reporting. This form of performance reporting provides, from a high-level viewpoint, an assessment

on how well a process is performing in which common and special cause events are separated.

This process behavioral characterization separation is important because the actions for special cause can involve the assessment of individual or group events, while an unsatisfactory common cause response suggests that process enhancements are needed.

With 30,000-foot-level reporting, x-bar and R, p-charts, and c-charts are not used to make a process stability assessment. The referenced 30,000-foot-level articles show mathematically why the individuals chart is appropriate for 30,000-foot-level reporting and other control charting techniques are not for this high-level form of performance reporting. [QP](#)

REFERENCES

1. Forrest W. Breyfogle III, *Integrated Enterprise Excellence Volume II—Business Deployment: A Leaders' Guide for Going Beyond Lean Six Sigma and the Balanced Scorecard*, Citius Publishing, 2008.
2. Ibid.
3. Forrest W. Breyfogle III, "Control Charting at the 30,000-Foot-Level," *Quality Progress*, November 2003, pp. 67-70.
4. Forrest W. Breyfogle III, "Control Charting at the 30,000-Foot-Level, Part 2," *Quality Progress*, November 2004, pp. 85-87.
5. Forrest W. Breyfogle III, "Control Charting at the 30,000-Foot-Level, Part 3," *Quality Progress*, November 2005, pp. 66-70.
6. Forrest W. Breyfogle III, "Control Charting at the 30,000-Foot-Level," *Quality Progress*, November 2006, pp. 59-62.
7. Forrest W. Breyfogle III, "No Specification? No Problem," *Quality Progress*, November 2012, pp. 58-61.
8. Forrest W. Breyfogle, "30,000-Foot-Level Performance Metric Reporting," *Six Sigma Forum Magazine*, February 2014, pp. 18-32.
9. Forrest W. Breyfogle III, "Integrating Inputs: A System to Capture and React to VOC Data Can Pay Dividends," *Quality Progress*, January 2011, pp. 64-66.
10. Forrest W. Breyfogle III, "High Vantage Point Report-outs to Reduce Risks of Organizational Problems," *Quality Progress*, December 2015, pp. 58-60.
11. Forrest W. Breyfogle III, "Monitor and Manage: Diabetes Measurement Tracking at the 30,000-Foot-Level," *Quality Progress*, January 2017, pp. 50-53.
12. Forrest W. Breyfogle III, *Integrated Enterprise Excellence Volume III—Improvement Project Execution: A Management and Black Belt Guide for Going Beyond Lean Six Sigma and the Balanced Scorecard*, Citius Publishing, 2008.

NOTE

A no-charge Minitab add-in is available for the easy creation of IEE 30,000-foot-level charts at www.smartersolutions.com/30000-foot-level-add. Organizations can benefit when 30,000-foot-level reporting is integrated with the processes that created them using an IEE value chain.



Forrest Breyfogle III is CEO of Smarter Solutions Inc. in Austin, TX, and holds a master of science degree in mechanical engineering from the University of Texas in Austin. An ASQ fellow, Breyfogle has authored or co-authored more than a dozen business management and process improvement books.