

2007

Common Pitfalls in Statistical Thinking

Mark C. Paulk
Carnegie Mellon University

Elaine B. Hyder
Carnegie Mellon University

Follow this and additional works at: <http://repository.cmu.edu/isr>

Published In

.

This Article is brought to you for free and open access by the School of Computer Science at Research Showcase @ CMU. It has been accepted for inclusion in Institute for Software Research by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Common Pitfalls in Statistical Thinking

MARK C. PAULK AND ELAINE B. HYDER
Carnegie Mellon University

According to many best practices frameworks for quality management and process improvement, statistical thinking is an intrinsic part of building organizational capability. In all of these frameworks there is a degree of flexibility in how the statistical thinking concepts are implemented, because these frameworks apply to many different contexts. Flexibility, however, also leads to ambiguity and inconsistency. The purpose of this article is to identify what has been observed in some organizations implementing statistical thinking. These observations can be lessons learned to help accelerate the learning curve for others implementing these frameworks.

Keywords: control chart, evidence-based management, process behavior chart, SPC, statistical process control, statistical thinking

SQP References

A Practical Process to Establish
Software Metrics
Linda Westfall
vol. 8, issue 2

Steer Clear of Hazards on the Road
to Software Measurement Success
Patricia McQuaid & Carol Dekkers
vol. 6, issue 2

The Uses and Abuses of Software Metrics
Pat Cross
vol. 6, issue 2

INTRODUCTION

Statistical thinking is an integral concept to many best practices frameworks, such as the Capability Maturity Model[®] for Software (Software CMM[®]) (Paulk et al. 1995), CMM IntegrationSM (CMMI[®]) (Chrissis, Konrad, and Shrum 2006), the eSourcing Capability Model for Service Providers (eSCM-SP) (Hyder, Heston, and Paulk 2004a; Hyder 2004b), ISO[®]/IEC 15504 (Process Assessment) (ISO/IEC 2003), Six Sigma[®] (Breyfogle 2003), and ISO 9001 (Quality Management Systems) (ISO 2000). Statistical thinking basically means the decision makers incorporate a rigorous understanding of variation into the decision-making process. Speaking as authors and contributors to a variety of improvement frameworks, the authors of this article have observed on many occasions that the implementation of statistical thinking by the organizations implementing these frameworks often fall short of their expectations (Paulk et al. 2005; Paulk, Hyder, and Heston 2006).

Statistical thinking is based on three fundamental axioms (Hare et al. 1995):

- All work is a series of interconnected processes.
- All processes are variable.
- Understanding variation is the basis for evidence-based management and systematic improvement.

Statistical thinking with a view toward management and control acknowledges that variation must be taken into account in the decision-making process; performance is constrained by the variation built into the system. Statistical thinking with a view toward improvement implies minimizing variability, which

in turn implies that the sources of variation must be identified and eliminated or reduced. Variation can only be eliminated if the distinction between common and special causes is understood, since the actions to correct special causes of variation differ significantly from those needed to systematically change the common cause system.

This article presents a discussion of what has been observed in some organizations when implementing quality and process frameworks (Paulk, Goldenson, and White 2000; Paulk and Chrissis 2000; Paulk and Chrissis 2001; Wheeler 2003). Where appropriate, the discussion of mistakes that organizations make with respect to statistical thinking is tied to specific frameworks, but the observations are generally applicable to other improvement frameworks. Many of the organizations where these observations were made are software developers using multiple frameworks, including Six Sigma, ISO 9001, the eSCM-SP, Software CMM, and/or CMMI. The authors have been contributors to a variety of models and standards, including the eSCM-SP, the Software CMM, and ISO/IEC 15504, and reviewers of other frameworks, including ISO 9001 and CMMI.

BEST PRACTICE FRAMEWORKS

There are two major strategies for improving performance: framework based and principle based. A framework-based strategy uses models and standards as best practice frameworks to identify what processes and systems should be implemented in a successful organization. Certification in some framework-based strategies, such as ISO 9001, is binary; an organization is either compliant with the standard or not. Models such as CMMI and the eSCM-SP measure organizations or processes using a form of ordinal scale; five-level scales, such as capability levels or maturity levels, are common. Best practice frameworks identify what to do, but do not usually describe how to do it or performance levels for specific tasks.

The second strategy is principle based. The organization's processes and systems are measured and compared to business and improvement objectives to identify needed improvements. Measurement trends are used to confirm and quantify improvements. Framework-based strategies naturally evolve toward principle-based strategies tailored to the business needs of the organization as the foundational capabili-

ties described by the framework are successfully put in place. By focusing on its business objectives, the organization can leverage a variety of strategies, allowing it to develop an integrated improvement strategy. For example, many organizations initiate improvement efforts with ISO 9001, then adopt more focused frameworks such as CMMI for product development, and then use principle-based strategies such as Six Sigma to drive their improvement after achieving the higher levels in CMMI.

All of the multilevel frameworks explicitly rely on establishing measurement and analysis capabilities at the lower levels and applying statistical thinking in decision making at the higher levels. Measurement and statistical thinking are intrinsic to the principle-based strategies.

The eSCM-SP exemplifies this roadmap. It has three purposes: 1) to give service providers guidance that will help them improve their capability across the sourcing life cycle; 2) to provide clients with an objective means of evaluating the capability of service providers; and 3) to offer service providers a standard to use when differentiating themselves from competitors. Released in April 2004, eSCM-SP v2 comprises 84 practices, structured in five capability levels, with level 5 being the most advanced.

Measurement and analysis are embedded throughout the practices (Paulk et al. 2005). Statistical thinking is primarily embedded in two capability level 4 practices on process capability baselines and benchmarking capability baselines. These identify the expected performance of a process and are typically based on comparing the natural process limits of a control chart to the specification limits for the process. Benchmarking relies on being able to identify when two processes have significantly different performances—both statistical significance and practical significance.

The authors of the eSCM-SP had certain expectations for a complete implementation of both measurement and analysis and statistical thinking (Paulk, Hyder, and Heston 2006), which would be shared by most framework authors. To achieve the level of statistical thinking intended by the authors, an organization should do six fundamental things:

1. Identify the measurable objectives that will be the basis for judging success. These objectives should address both project and organizational

objectives and include business, improvement, and client objectives.

2. Create processes that are well defined and can be measured for their impact on those objectives.
3. Create a measurement program that uses a small set of balanced measures to help management determine the actions to take to meet the objectives.
4. Create a measurement infrastructure, including the tools and repositories, to support decision making.
5. Use appropriate analytical techniques, including statistical process control (SPC), to manage the critical processes.
6. Use measurement data and analysis in making decisions, both to achieve business objectives and to improve performance (note that improvement should be both statistically and practically significant).

The authors of the eSCM-SP believe the model should not overly constrain an organization's implementation of either measurement and analysis or statistical thinking. While an organization should incorporate data into its decision making where it will add value and should demonstrate the understanding of variation intrinsic to statistical thinking at higher capability levels, the model must allow a variety of implementations to fit the environments where the model may be adopted. The requirements in the model are therefore written at a high level of abstraction, which allows flexibility, but which can also lead to inappropriate and ineffective implementations when an organization implements the form of the practices without an understanding of their heart. This philosophy of not over constraining the implementation of best practices is broadly held among developers of improvement frameworks, although the implications of "over constrain" may change depending on where the framework is targeted.

As previously stated, the authors of this article are also authors of the eSCM-SP, as well as authors, contributors, and reviewers of a variety of other quality and process frameworks. In discussions with their colleagues, they believe it is fair to say that they share the fundamentals described previously, although they have had many interesting debates on imple-

mentation specifics. For example, while agreeing on the importance of statistical thinking, some of the authors' colleagues would argue against the appropriateness of control charts for predominantly creative work, such as design, while others would argue that SPC is the optimal strategy to use for improvement for all processes.

MISTAKES WHEN IMPLEMENTING STATISTICAL THINKING

There are a number of common mistakes that organizations make when implementing statistical thinking:

- Using common sense to ignore flawed processes or inappropriate statistical techniques
- Setting natural process limits using statistical techniques that are not economically efficient
- Using incorrect formulae, such as standard deviation, for calculating natural process limits
- Failing to perform the causal analysis that leads to taking informed action based on statistical signals
- Confusing targets and natural process limits as triggers
- Analyzing data from mixtures of (operationally) different processes
- Failing to balance stability and improvement
- Using measurement for motivational purposes

The use of inefficient techniques for statistical analysis and control does not necessarily mean that an organization has not satisfied the level 4 practices in the eSCM-SP, but there are opportunities for improvement. Whether a particular statistical technique is inefficient (and perhaps ineffective) for a particular environment can be debated. Suggesting that there are improvement opportunities does not mean the organization has failed to address the intent of a practice. From a certification perspective, the authors do not judge the goodness of an implementation, simply whether there is an adequate implementation that can act as the foundation for action and continual improvement.

Applying Common Sense

A primary concern in deploying best practices is that the organization may not always follow its own documented processes. Staff may apply common sense to ignore real or perceived flaws in the documented process. The authors do not wish to discourage the use of common sense when following the documented process would not add value, understanding, or insight. It does, however, raise an issue about compliance to the documented processes and the effectiveness of the process implementation. This is a sensitive point, but requiring that staff follow an obviously inappropriate process would damage the credibility of the authors of the framework being used, the assessment or audit team, and those defining the process in question. An assessment/audit team cannot ignore such a problem; it must report both the noncompliance and the issues with the documented process.

For example, some organizations set upper and lower natural process limits at the 5th and 95th percentiles of their data rather than using control charts. Analyzing the bottom 5 percent of the data may be pointless, particularly when there is a natural boundary for the data, such as zero. For a skewed distribution where a relatively high percentage of the data—more than 5 percent—may be at zero, analyzing the points in the bottom 5th percentile appears unreasonable. Observations where the data have a value of zero can be expected to occur occasionally in this process, and analyzing the bottom 5 percent provides little or no insight. Analyzing how often—and whether—zero events should occur may be a valid question, but exploring the “whys” of the common cause system is a process improvement question, not a process control question.

Ignoring data below the 5th percentile limit because it does not make sense to treat those observations as atypical is an example of doing the right thing, even though the guideline, which implements a poor statistical technique, would require doing something else . . . and where implementing a superior statistical technique, for example, an XmR chart, would lead to doing the right thing for the right reason. Blindly doing a causal analysis where common sense indicates that no value would be obtained would be counterproductive.

Sometimes the meaning of the 95th percentile limit is misinterpreted. It does not mean to pick the five largest data values and do a causal analysis. It means pick the largest 5 percent. If one has 20 data points,

it means the largest value; for 1,000 data points, it means the largest 50 values.

The prescription for simple arithmetic errors is training and education, with perhaps better tool support. The prescription for failing to follow a nonsensical process is more challenging, since the solution involves defining processes—and analytic techniques—that are appropriate for the kind of work being done. Defining “good” measures and analytic techniques is an extension of the implied need for “good processes” at the lower levels of models such as the eSCM-SP and CMMI. While how-to “goodness” is deliberately out of scope for what-to-do frameworks, a process is what one does, not what he or she documents. Processes that are not used fail to build on the lower-level capabilities that are a prerequisite for the higher-level statistical thinking in these models.

The Economic Benefit of Natural Process Limits

There are several ways to determine the natural process limits for a capability baseline. The expectation is that the limits will be defined based on a statistical technique, such as control charts or prediction intervals, although simple graphical techniques may be sufficient to address the analytical needs (Wood, Capon, and Kaye 1998).

Classical SPC uses 3σ limits for reasons of economic benefit. There is an economic cost associated with analyzing process observations. Any data points outside the 3σ limits are unlikely to be part of the normal operation of the process, so there will be few “false alarms” (Wheeler’s Empirical Rule for a set of homogenous data, that is, the common cause system, is that more than 99 percent of the observations will be within the 3σ limits with no requirement of normality (Wheeler and Chambers 1992; Wheeler 2000)).

Using 5 percent and 95 percent limits is not statistically “wrong,” but identifying 10 percent of the data as signals of possible special causes is inefficient since 90 percent or more of the signals are likely to be false alarms. Unnecessarily analyzing 9 percent of the events being controlled (90 percent false alarms of the 10 percent of the observations that are signals) is inefficient when more effective tools are available. Control chart limits are designed to be economically efficient and simple to calculate (Wheeler and Chambers 1992;

Wheeler 2003). The 3σ limits in control charts have been demonstrated to balance signals and false alarms at an economically useful point. While other criteria may be used to trigger action, they should be used by people who understand the tradeoffs they are making.

The prescription for failing to use effective and efficient tools builds on applying both relatively sophisticated statistical expertise and a profound knowledge of the engineering methodologies and application domains associated with the processes to be controlled. Inefficient techniques are a cost reduction opportunity, but inefficient techniques are arguably superior to the heuristics and intuition used in measurement and analysis at the lower levels.

Using Standard Deviations in Natural Process Limits

A classic mistake from an SPC perspective is using the sample standard deviation in calculating the upper and lower limits for a control chart. When Walter Shewhart described the control chart, he pointed out the importance of using the average (or median) of dispersion statistics rather than using a global dispersion statistic (Shewhart 1931). This mistake continues to be made despite numerous articles demonstrating that incorrectly calculated natural process limits using this formula can be dramatically wider than those using the correct formulae (Paulk 2000b; Pyzdek 1998; Wheeler and Chambers 1992; Wheeler 1994; 2003).

The sample standard deviation (s) is an estimate of the population standard deviation (σ) in classical statistics, but one of the assumptions (sometimes not articulated) in using it as an estimator is that the sample comes from a homogenous population. The standard way of writing the formula for the sample standard deviation is:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Two alternative, equivalent formulas for calculating the sample standard deviation are:

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n - 1}}$$

or

$$s = \sqrt{\frac{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}{n(n - 1)}}$$

While it may be simple to use the “s” key on a calculator, or the STDEV function in a spreadsheet, to calculate this estimator for σ , when there are special causes of variation in the data, the $\pm 3s$ limits can be five to six times as large as the correct $\pm 3\sigma$ limits. One common response to these wide limits is to use $\pm s$ or $\pm 2s$ as the limits in an attempt to establish more credible bounds. Anytime the limits are “adjusted” because they are too wide, the natural suspicion is that there are fundamental problems in collecting valid data, aggregating dissimilar data (perhaps from different underlying processes), or using inappropriate statistical techniques.

The correct formulae are also simple. The upper and lower natural process limits for the X part of an XmR chart, for example, are calculated by

$$UNPL_X = \bar{X} + 2.66\bar{mR}$$

$$LNPL_X = \bar{X} - 2.66\bar{mR}$$

$$UCL_X = 3.268\bar{mR}$$

where the “bar” above the variable denotes the average of the individual X values and the moving ranges, respectively. A moving range is the difference between successive data points. The lower natural process limit for the moving range chart is always zero. There are many different kinds of control charts that one may select, depending on sample size (the XmR chart uses individual observations), statistical distributions assumed, whether variables or attributes data are being used, and so forth.

The $\pm 3s$ limits are not inflated if there are no special causes of variation in the data, since the global measure of dispersion s is a reasonable estimator of σ in the common cause system under those circumstances. But that implies identifying the signals using a different technique and just using the natural process limits for calculating the capability baselines—a consequence of cascading inappropriate and ineffective statistical techniques. Instances where “natural process limits” are based on the 5th and 95th percentiles or $\pm 2s$ are usually indicators of incorrect calculations that have fatally expanded the limits from a usability perspective.

The prescription for addressing $\pm 3s$ limits differs from that for percentile-based limits for a simple reason: the mistake is based on a misunderstanding of

the correct statistical techniques. Train the analysts in SPC! There are times when the sample standard deviation is correct and times when it is not. Useful statistical debates depend on understanding the difference. One can claim to have considered the cost intrinsic to analyzing 10 percent of the data, however, it is much more difficult to defend using limits that may be five times as large as correctly calculated limits.

Ignoring Causal Analysis

Special causes of variation should not be removed without a causal analysis of why the special event occurred. Special causes are a form of outlier, and while outliers may skew the results of a statistical analysis, outliers that are not clearly erroneous should neither be completely discarded nor blindly included in a statistical analysis (Neter, Kutner, and Nachtsheim 1996). Discarding outliers without root cause analysis can adversely affect the validity of conclusions. Sometimes the outliers contain useful information about the common cause system, which can be extracted during the causal analysis.

The prescription for ignoring causal analysis depends on an assumption that decisions will be made as the result of identifying unusual events. A lack of causal analysis is usually associated with a lack of action and suggests that a deeper problem needs to be addressed. Folding together measurement, analysis, and action is built into evidence-based management and SPC.

Confusing Targets and Capability

The authors have observed benchmarking guidelines that state that if actual performance is within an arbitrary threshold, say 5 percent of the benchmark, then performance is acceptable; if the benchmark is more than 5 percent better, improvement should be done to the process. The choice of 5 percent as the trigger for action is essentially arbitrary. Such arbitrary thresholds do not indicate an understanding of variation.

If a capability baseline has been determined for the process, then the natural process limits for performance can be used as the trigger for considering whether action should be considered (other techniques include analysis of variance or test of hypotheses, but the capability baselines provide an existing criterion). Given the large degree of variation in many software processes, it is likely that a 5

percent difference from the average performance will be within the bounds of the typical variability of the process, so it is unclear that the benchmark process is truly superior to the process.

Using a 5 percent criterion also ignores the pragmatic consideration that a difference in performance should not only be real but also practically significant. If the software process is improved to the level of the benchmark—or even better—will that improvement be of value to the client or the provider? That is a decision that should be made in the context of other possible improvement actions that may have greater impact from a practical perspective.

An even worse criterion sometimes used is a simple greater-than comparison. Variation happens. Capability level 4 organizations should have an understanding of variation: Even if the values are different, is the difference statistically significant? Or is the difference simply part of the normal variation intrinsic to the process, which may be up in June and down in December? Identifying these patterns is a potential use of the capability baselines. If the performance is within the natural process limits of the capability baseline, then performance is considered unchanged (assuming that the process is stable, that is, that any special causes of variation have been analyzed and addressed appropriately). Only when the performance is outside the bounds of the capability baseline and in the “positive” direction should the trend be characterized as improving.

The prescription is simply to set both objectives based on needs and desires and limits based on the voice of the process. From a goal-driven measurement perspective, one needs both. The understanding of variation intrinsic to statistical thinking implies more-informed decisions, while goal-driven measurement establishes the context for making decisions in the first place.

Mixing and Stratifying

It is common for the process data to suggest that a mixture of different kinds of entities, that is, observations from different processes, have been grouped together on the control chart. When the special causes of variation are clusters of “complex” versus “simple” entities, for example, one inference may be that the “simple,” “typical,” and “complex” entities are enough different that they should be separately analyzed.

Different strata of data may be readily visible on a run chart when examined with the perspective that different processes may be overlaid on the chart. The decision on how to disaggregate the data into homogeneous pieces is best made by someone with a profound understanding of the processes and measures in question. The question of data homogeneity is a common one in statistics, and it should be thoughtfully considered.

The prescription for mixing and stratification is both fundamental and challenging, since expertise in measurement, statistical thinking, the processes, the methodologies, and the application domain is needed to understand what the data mean. Understanding the data must precede sophisticated analyses.

Stabilizing the Process

In addition to the informal stabilization of the process previously discussed, there is a tension between having a stable process and continual process improvement. Continual improvement means the process data for the previous process may not be valid for the new process, and new natural process limits may need to be recalculated on an ongoing basis. Even when most changes are incremental, compounded small changes can lead to dramatic improvements. Wheeler observes that continual improvement is a journey consisting of frequent, intermittent improvements, interspersed with alternating periods of predictable and unpredictable performance (Wheeler and Poling 1998; Wheeler 2003).

The prescription is to use statistical thinking in both process control and process improvement. While this may seem a trite recommendation, it has appeared to be a useful insight for more than one organization. All too often, the users of best practice frameworks focus on the practices in the framework and lose sight of why they are being implemented in the first place.

Using Measurement for Motivational Purposes

It is always tempting to use the data collected for SPC as the basis for rewarding (or punishing) the individuals collecting the data. If measures are used to evaluate individuals, however, counterproductive behaviors are likely to arise.

W. Edwards Deming was a strong advocate of statistical techniques, yet he was strongly averse

to performance evaluations, declaring performance measurement the most powerful inhibitor to quality and productivity in the Western world. Although arguably the most controversial of Deming's points, Austin (1996) has shown that the potential for dysfunction arises when the measures used do not comprehensively cover all of the important factors affecting success. Unless the latitude to subvert measures can be eliminated, or a means established for preventing the motivational use of data, dysfunction is destined to accompany organizational measurement.

The simplest prescription for preventing this kind of dysfunctional behavior is to divorce the measures used for process control and product insight from the reward and recognition system—but this simple prescription differs from normal practice in most organizations. Identifying a comprehensive set of measures, using methods such as the balanced scorecard (Kaplan and Norton 1996), would also address this concern, but it seems likely to remain a challenge for the foreseeable future.

CONCLUSIONS

The purpose of statistical thinking is to obtain measurable value rather than to satisfy the requirements of a model or standard. The tools of SPC (including control charts) should be part of the toolkit for a capable organization. Appropriate tools should be used to answer questions that are important in achieving business objectives and to support continual, measurable improvement. Without this anchoring in business reality, statistical thinking cannot add value.

The requirements in frameworks such as the eSCM-SP are intentionally general and do not overly constrain implementations, but they should not be viewed as a license for “teaching to the test.” Statistical thinking should address a business need more effectively and efficiently than nonstatistical techniques if it is appropriately applied. Many different statistical techniques can be used to address these business needs; the examples in this article illustrate a few of the more commonly used ones. The mistakes in statistical thinking described in this article are neither new nor unique. The fact that a mistake is common does not mean that it will be forgiven in today's competitive environment, but it does mean organizations implementing statistical thinking correctly will have a competitive advantage.

Common Pitfalls in Statistical Thinking

REFERENCES

- Austin, R. D. 1996. *Measuring and managing performance in organizations*. New York: Dorset House Publishing.
- Chrissis, M. B., M. D. Konrad, and S. Shrum. 2006. *CMMI: Guidelines for process integration and product improvement, second edition*. Boston: Addison-Wesley.
- Hare, L. B., R. W. Hoerl, J. D. Hromi, and R. D. Snee. 1995. The role of statistical thinking in management. *Quality Progress* 28, no. 2 (February): 53-60.
- Hyder, E. B., K. M. Heston, and M. C. Paulk. 2004a. *The eSourcing Capability Model for Service Providers v2: Model Overview*. CMU-ISRI-04-113. Pittsburgh: Carnegie Mellon University, Institute for Software Research International.
- Hyder, E. B., K. M. Heston, and M. C. Paulk. 2004b. *The eSourcing Capability Model for Service Providers v2: Practice Details*. CMU-ISRI-04-114. Pittsburgh: Carnegie Mellon University, Institute for Software Research International.
- ISO 9001:2000. 2000. *Quality management systems—Requirements*. Geneva, Switzerland: International Organization for Standardization.
- ISO/IEC 15504-2:2003. 2003. *Information technology—Process assessment—Part 2: Performing an assessment*. Geneva, Switzerland: International Organization for Standardization and International Electrotechnical Commission.
- Kaplan, R. S., and D. P. Norton. 1996. *The balanced scorecard: Translating strategy into action*. Boston: Harvard Business School Publishing.
- Neter, J., M. H. Kutner, and C. J. Nachtsheim. 1996. *Applied linear statistical models, fourth edition*. Chicago: Irwin.
- Paulk, M. C., C. V. Weber, B. Curtis, and M. B. Chrissis. 1995. *The Capability Maturity Model: Guidelines for improving the software process*. Boston: Addison-Wesley.
- Paulk, M. C., D. R. Goldenson, and D. M. White. 2000. *The 1999 survey of high maturity organizations*. CMU/SEI-2000-SR-002. Pittsburgh: Carnegie Mellon University, Software Engineering Institute.
- Paulk, M. C., and M. B. Chrissis. 2000. *The November 1999 High Maturity Workshop*. CMU/SEI-2000-SR-003. Pittsburgh: Carnegie Mellon University, Software Engineering Institute (March): 38-39.
- Paulk, M. C., and M. B. Chrissis. 2001. *The 2001 high maturity workshop*. CMU/SEI-2001-SR-014. Pittsburgh: Carnegie Mellon University, Software Engineering Institute.
- Paulk, M. C., S. L. Dove, S. Guha, E. B. Hyder, M. Iqbal, K. O. Jacoby, D. M. Northcutt, and G. E. Stark. 2005. *Measurement and the eSourcing Capability Model for Service Providers v2*. CMU-ISRI-04-128. Pittsburgh: Carnegie Mellon University, Institute for Software Research International.
- Paulk, M. C., E. B. Hyder, and K. M. Heston. 2006. *Statistical thinking in the eSourcing capability model for service providers*. Working paper, ITSqc.
- Pyzdek, T. 1998. How do I compute sigma? Let me count the ways. *Quality Digest* (May).
- Shewhart, W. A. 1931. *Economic control of quality of manufactured product*. New York: Van Nostrand: 302-304.
- Wheeler, D. J., and D. S. Chambers. 1992. *Understanding statistical process control, second edition*. Knoxville, Tenn.: SPC Press.
- Wheeler, D. J. 1994. Charts done right. *Quality Progress* 27, no. 6 (June): 65-68.
- Wheeler, D. J. and S. R. Poling. 1998. *Building continual improvement: A guide for business*. Knoxville, Tenn.: SPC Press.
- Wheeler, D. J. 2000. *Normality and the process behavior chart*. Knoxville, Tenn.: SPC Press.
- Wheeler, D. J. 2003. *Making sense of data: SPC for the service sector*. Knoxville, Tenn.: SPC Press.
- Wood, M., N. Capon, and M. Kaye. 1998. User-friendly statistical concepts for process monitoring. *Journal of the Operational Research Society* 49: 976-985.

® Carnegie Mellon, Capability Maturity Model, CMMI, and CMM are registered trademarks of Carnegie Mellon University.

SM CMM Integration and SEI are service marks of Carnegie Mellon University.

® ISO is a registered trademark of International Organization for Standardization.

® Six Sigma is a registered trademark of Motorola, Inc.

BIOGRAPHIES

Mark Paulk is a senior systems scientist at the IT Services Qualification Center at Carnegie Mellon University, where he works on best practices for sourcing of IT-enabled services. From 1987 to 2002, Paulk was with the Software Engineering Institute at Carnegie Mellon University, where he led the work on the Capability Maturity Model for Software. Paulk received his doctorate in industrial engineering from the University of Pittsburgh, his master's degree in computer science from Vanderbilt University, and his bachelor's degree in mathematics and computer science from the University of Alabama in Huntsville. He is a Senior member of the IEEE, a Senior member of ASQ, and an ASQ Certified Software Quality Engineer. You can reach him at Mark.Paulk@cmu.edu and his home page is <http://home.comcast.net/~Mark.Paulk/>.

Elaine Hyder is a systems scientist at the IT Services Qualification Center at Carnegie Mellon University, where she works on best practices for sourcing of IT-enabled services. She has been working on the eSCM since its inception and is the main author of eSCM-SP versions 1.0, 1.1, and 2.0. She has also been involved in developing course curriculum related to the eSCM-SP, developing eSCM-SP capability determination methods, and teaching eSCM-SP coursework. Current research interests include conflict resolution, standards development, and the use of technology to improve the quality of life and work. Hyder has a doctorate in information systems from Carnegie Mellon University. She is a member of the Academy of Management, IEEE, and INFORMS.