# Unlocking the Hidden Logic of Process Improvement: Peer Reviews

Marilyn Bush
*Marilyn Bush Associates*

*The success of many of the most useful procedures recommended by the higher maturity levels of the Capability Maturity Model® (CMM®) and the CMM Integration℠ (CMMI®) depends on understanding the silent links connecting them to practices associated with Level 2. Certain activities that can seem peripheral to Level 2 goals during an assessment, in fact, prepare organizational foundations for what will become fundamental improvements. In trying to get through Level 2 as quickly as possible and neglecting the global logic of the CMM and the CMMI, companies sidestep the momentum-builders that can catapult them to world-class status. Perhaps the most significant illustration of this phenomenon concerns the function of peer reviews and the procedures associated with detecting defects early.*

The success of using peer reviews can vary widely. Many companies start to use peer reviews and then abandon them. These companies introduced formal training, but it did not help – within a year or two, very few if any projects were still actively doing peer reviews. Other companies swear by peer reviews, and boast that "they have saved us millions of dollars." So why do peer reviews work in some companies and not in others? And what difference does it make? The answers have to do with organizational culture.

All of this could be gleaned from reading the Software Engineering Institute's Capability Maturity Model® (CMM®)/CMM Integration℠ (CMMI®). But you need to read between the lines, since the CMM and CMMI enumerate necessary tasks but do not always describe the logic that sequences them. Part of this hidden logic involves the fact that moving toward and achieving Level 2 involves a general organizational discipline that recognizes and deals with problems early, accepts independent quality reviews, and promotes discomfort when quality procedures are missing.

This cultural change affects far more than the tasks specifically identified in Level 2, and it is essential for getting the most out of, for example, peer reviews. Although some companies seem to be able to assemble this kind of discipline without having all the pieces of Level 2 in place (for example, if they have previously employed other kinds of quality programs such as Crosby), this kind of discipline amounts to having all the practices of Level 2 in place. Level 2 has to do with management discipline. Most Level 1 managers cannot protect their workers because their process is too chaotic. Without such protection, workers do not have the freedom to perform peer reviews effectively.

## The Inner Logic of the CMM/CMMI: Technical Progress Conditions Cultural Discipline

The CMM and CMMI enumerate necessary tasks but do not always describe the logic that sequences them. Both models assume the importance of detecting defects early and then preventing them, yet these goals are not spelled out in the early stages of the maturity scale. At Level 2, all the process areas/key process areas ostensibly concern project management activities – risk assessment, creating consistent processes, etc.

## Technical Culture and Organizational Culture

Technical culture in this article is defined as the sum of an organization's technical practices and methodologies. Changing the technical culture involves instituting better practices, or organizing them in a different sequence. It involves *what* employees do, not why they do it.

Organizational culture deals with the underlying values that motivate individuals as they relate to an organization and its present and future goals. In the words of Harrison Trice and Janice Beyer, cultures consist of:

> … shared sets of beliefs, values, and norms that both impel people to action and justify their actions to themselves and others. With the passage of time, [these sets of beliefs] tend to move away from the forefront of people's attention and become implicit and taken for granted. [1]

## Why Maturity Levels Cannot Be Skipped

Process capability grows in stages. Key processes are only effective after prerequisite processes are stabilized. Engineering processes usually do not improve, for example, before management stabilizes the way it makes decisions. If management changes the work conditions day to day or week to week, the best processes in the world do not have a chance to succeed. Therefore, Level 2 is largely concerned with management decision processes. Further, as management discipline solidifies, so does a more general quality discipline [2].

In technical terms, at Level 2, managers learn to prepare estimates with their team rather than by themselves and methodically track actual estimates against original estimates. These actions constitute an important increase in technical rigor. They also, however, involve changing perspectives as well as changing practices. Managers are empowered by enhanced information to take corrective action early rather than late, and get into the habit of doing so when they need to. When costly problems are found in reviews, and fixed early and easily, teams start to see the benefit of independent and methodical reviewing and begin to feel discomfort when consistent processes are missing.

In other words, changes at Level 2 not only alter what people do but also how they think about it, that is, they substantively alter organizational culture. The big picture looks like this: When all members of the project are involved in planning activities, the whole team has to come to a consensus about goals and necessary quality standards in the product and the process. This can be a slow and bruising process, but it can happen in a timely manner with the proper culture and team participation. Reaching a consensus means the entire team buys in. People are less afraid to raise problems because they have survived an open discussion, and because their first prior-

ity is to fix things before it is too late. Quality comes before blame. (This is sometimes called *decriminalization*.) A culture in which problems are handled early – without the fear of blame – also accepts independent review without defensiveness.

How does all this happen? Where does it lead?

## Changing the Technical Culture for Defect Detection[1]

One of the most powerful payoffs of the CMM and the CMMI has to do with the savings made possible by finding and fixing defects before testing (the technique that is used is called *peer reviews* in both the CMM and CMMI). Peer reviews offer *huge* savings when done correctly. For example, to find and fix a defect before testing costs only *one percent* of what it would cost to fix the same defect in operations. That is, to find and fix one defect found early in development might cost $100. To find and fix the same defect during integration and test can cost $1,000 to $2,000. And to find and fix the same defect post-delivery costs an organization anywhere from $10,000 to $20,000. (Fixing critical defects found post-delivery can cost much, much more [3]).

Today, software developers expect to find at least three to six post-delivery defects per thousand lines of code[2] [4]. That means a loss in the range of $50,000 to $100,00 in unnecessary costs per thousand lines of code. For a typical system that contains millions of lines of code, the unnecessary costs that can be saved are astronomical.

An operational defect, defined as a defect encountered by the product user as a failure, causes the product to malfunction relative to a product specification. (If it does not do what the user wants, it does not work.)

Since many projects now consist of anywhere between 250,000 and millions of lines of code (with programs getting more complex every day), the real costs of quality are enormous, and the real benefits of process improvement are equally high – the potential saving of hundreds of millions of dollars of preventable costs.

Levels 4 and 5 build naturally on the technical and organizational cultures conditioned by Levels 2 and 3. But peer reviews require not only a management discipline but also a cultural discipline. This kind of developing cultural discipline forms an essential element of the CMM /CMMI program of maturity.

## Why Defect Detection Does Not Work Without Level 3 Maturity: Doing Peer Reviews in an Immature Organization

Even when they seem to be doing so, teams without a culture of cooperation do not really look for operational defects.

Managers who have not internalized the culture of quality scare staff away from discussing problems by quickly assigning blame. Lacking the technical discipline of planning and monitoring and control procedures, they too often rush to drop quality procedures when schedules are slipping.

Their subordinates pick up these signals and act accordingly. Finding too many defects might mean that they will be blamed for not doing their job properly. Rifkin states,

> If we fear for our jobs then we are less likely to take the chances that are inherent in performing some new action, making the inevitable mistakes. We would fear that such mistakes would count against us, and we may form a basis for poor performance and then we could lose our jobs. [5]

An operational defect, encountered by the product user as a failure, causes the product to malfunction relative to the product specification. Of course, not all defects are equally disabling. *Critical* defects render a product unusable and require immediate attention. In the case of *serious* defects, the customer's use is severely restricted. Defects of medium severity involve limitations that are not critical to overall operations.

Low severity defects permit users to circumvent the problem and use the product with slight inconvenience. For peer reviews, a determination must be made between major defects and minor defects. Critical and serious defects would be considered major defects, medium severity could be either major or minor depending on the nature of the defect, and low severity defects would fall under the category of minor defects [6].

In a Level 1 organization, finding a few defects (no matter how many more there were or whether the critical ones were caught) is *good enough*. Finding them will *impress the boss*. If too many are found, the boss will think that "we are not very good at our jobs."

The same attitudes shape the way teams in a Level 1 organization conduct peer reviews.

Managers who do not understand the culture of quality scare staff away from discussing problems. They then do not raise them because they are afraid they will be blamed for them – or for slowing down the schedule by raising them.

Without real changes in organizational discipline, peer reviews typically result in only one operational defect detected per review[3].

Consider this example. On one project, the project manager of organization X (which was Level 1) mandated that every development team use formal inspections. One team manager was sure he really did not need to do this since his team did not make mistakes (and even if they did, no one would know about them). But since he had to do it anyway, he decided to implement the order in the following way. He told the project manager his team did not have time to be trained. They would have to read the material on their own and would be directed to perform the formal inspections on Saturdays.

Needless to say his team inspection defect rate was very low. But since there was no one on the project regularly reviewing the inspection data as it was produced, this team manager was allowed to continue with his practice. When the system went into testing, over 60 percent of the defects found were from this one team's modules. It slowed the entire project schedule down by over three months, and cost the project $200,000. (Remember, the difference between one defect found in inspections versus testing is approximately $100 versus $1,000.)

Worse yet, after peer reviews produce disappointing results, companies may get discouraged and adopt a resistance to improvement that places them farther away from improved productivity than they were before.

## The Cost of Doing Peer Reviews Without a Mature Organizational Culture

Organizations quickly understand how doing peer reviews can save vast amounts of money. They less frequently see the point of the intermediate activities that really make peer reviews work. Getting to Level 2 can seem to take forever, and instituting Level 2 activities can often seem pointless. In the meantime people wonder why they cannot implement more radical techniques right away. "Why wait for Level 3 to do peer reviews?" they ask. "What's the use of all the *tinkering* the CMM/CMMI requires beforehand?" "Why do you need to progress through

| Name of Level 5 Assessed Company | Quality Improvement | Productivity/ Profit Improvement | Predictability Improvement | Customer Satisfaction |
|---|---|---|---|---|
| **Telcordia Technologies** Assessed Level 5 May 1999 [10] | **1992/93:** 48 Faults/Thousand Function Points (KFP). **1997:** One Fault/KFP. | **1992** Cost to customer 35-40 percent higher than in 1997 and profit margin substantially higher. Cost of testing a line of code is less than 1/3. | **1992:** Projects took two years. **1997:** Projects take six to nine months. | **1992:** 60 percent **1997:** 95 percent Link to satisfaction is that Severity 1 and 2 defects halved over two years. |
| **Onboard Space Shuttle Software (IBM Houston)** Assessed Level 5 November 1989 [11] | Two orders magnitude reduction in defects delivered/kloc. | 300 percent improvement since early 1980's. | Consistently predicts costs within 10 percent of actual expenditure. (Missed one deadline in 15 years.) | No information available. |
| **Motorola India** Assessed Level 5 November 1993 [12] | 50 percent of software delivered had no known defects (defect levels running at 30 defects/million lines of code). | Increased 3.5 times going from Level 3 to Level 5. | No information available. | No information available. |
| **BAESYSTEMS, CNI Division** Assessed Level 5 March 2002 [13] | Post-delivery defects at Level 5 amount to less than 0.26 defects per thousand lines of code. | CNI's process improvement costs averaged 3-5 percent of the software engineering directorate staff. Productivity has improved by 16 percent moving from Level 3 to Level 5. | On-time schedule commitments have risen to over 90 percent from Level 2 to Level 5. Cost performance has remained at or above 1.0 moving from Level 3 to Level 5 (a 26 percent cost improvement). | External Customer Satisfaction improved 9 percent in the past year. |

Table 1: *Results From Companies That Have Changed Their Culture*

Maturity Levels?" "Why not just do those things that provide a real payoff?"

Alternatively, organizations often try to get away with attaining Level 2 with only a cursory independent quality review in place. Consequently, their organizations do not develop a culture that is aware of the importance of finding problems early. Organizations also sometimes delay putting in place substantive measures to aid tracking critical elements at critical times. (For example, they estimate lines of code before they have done requirements and then collect no actual data.) Hence, their numbers give no early warning during the requirements and design phases, and the organizations delay developing a culture in which identifying issues early is positively reinforced. The result: management remains in crisis management mode rather than in a proactive mode.

But teams without a culture of cooperation cannot see operation-critical defects because they do not really know how to look for them. More precisely, they are not equipped as a culture to look for them. Without the management and cultural disciple conditioned by Level 2 activities, peer reviews produce derisory results. Hewlett Packard, we know, took 10 years to reach a 25 percent adoption level [7] because they did not have a Level 2 cultural discipline in place. The resulting discouragement usually leads to companies bowing to ever-present resistance to continuing process improvement. *Once staff turns against peer reviews, they will not attempt them at all, and they hesitate to adopt further improvement measures.* Nor do most managers understand enough to explain what has happened. All that anyone sees is wasted effort and unrealized savings.

## Implications for Senior Managers: Assessments
Senior managers as well as mid-level managers need to be aware of the technical and the cultural implications of the CMM/CMMI. They need to understand the value of a process and measurements that gives them real visibility early in the development process and allows them to be proactive rather than reactive. Being proactive is the key to quality.

Trice and Beyer state that,

In order to manage cultures of work organizations successfully, managers must (1) be culturally aware – that is, they must understand and take into account what culture is and how it works; (2) know the cultures they are managing; (3) recognize and use the levers they have available to influence their organizations' cultures; (4) resolve the ethical dilemmas involved in managing cultures; and (5) be clear about whether they seek to maintain existing cultures, change existing cultures, or establish new ones. If managers understand the nature of culture, they will be better able to recognize the opportunities and constraints it poses for managerial action. [8]

In other words, executives are key to the success of implementing change (which is never merely technical change). *Without executive vision, positive change is unlikely to occur.*

Assessments are an effective method for management and practitioners to get expert insight into the organization's

maturity and culture. Senior managers should not put too much stress on the numerical *grade* of an assessment and should stress instead an assessment's salutary stimulation. Organizations sometimes try to *game* an assessment, which sidesteps the self-reflection that leads to real change and hinders the growth of cultural discipline that will generate the major benefits to come.

Pretending things are better than they are does not improve things; only laying firm foundations helps. An organization's emphasis should not be "How badly did we do?", "Are we still at Level 1?", or "Who is to blame?" Rather it should ask: "What aren't we doing right and how can we fix it?" Asking these questions is already a big step toward higher maturity. Always, the greatest payoff is in heightened self-consciousness and the self-discipline that goes with it. With self-discipline the big payoffs later on are easy. Without it, they are nearly impossible [9].

Going from Level 1 to Level 2 the right way forces everyone in the organization to be more self-conscious. Middle and project managers begin to understand what to look for, what to ask for, and what the answers mean. Managers have real information and people become less frustrated. Once managers can see the road ahead, their expectations become more realistic, and so developers stop feeling they are being asked to do the impossible.

At that point, and not before, an organization is ready to do peer reviews properly.

## Implications for Senior Executives: The Bottom Line
Without the hands-on and technically informed input of executives, the big decisions (not to game an assessment, to take the recommendations that come out of an assessment seriously) do not get implemented. And unless they are implemented, nothing changes.

To make this kind of decision with confidence, senior managers need to understand the huge payoff in profit margin when defects are caught or prevented before testing, and how these payoffs are tied to a changed organizational culture. Unless they personally understand the way a culture of cooperation and discipline evolves through the levels of process improvement, their organization will inevitably take all the easy ways out.

Being world class brings enormous

rewards. The charts in Table 1 illustrate four Level 5 organizations and their reported benefits. It should be noted that these organizations are in different businesses, have different product lines, and in general have different business concerns. Yet the numbers show the same order of results.

## Conclusion

Peer reviews as prescribed at Level 3 by both the CMM and the CMMI are immensely profitable, but only if they are done right. When attempted too soon in a process improvement program, they not only can disappoint but also prove counterproductive. (The disappointment is so great that it interferes with further process improvement.)

Knowing when to perform peer reviews depends on understanding the silent links between practices associated with Level 2 and Level 3 of the CMM/CMMI improvement programs. Certain activities that seem peripheral to Level 2 goals, in fact, prepare the cultural foundations for more sophisticated activities. In trying to get through Level 2 as quickly as possible and neglecting the global logic of the CMM/CMMI, companies sidestep the momentum-builders that can catapult them to world-class status. This paper has discussed what some of these submerged links are.◆

## References

1. Trice, H.M., and J.M. Beyer. The Cultures of Work Organizations. Englewood Cliffs, NJ: Prentice Hall, 1993: 75, 76.
2. Humphrey, Watts S. Managing the Software Process. Addison-Wesley, 1989.
3. Bush, M. Improving Software Quality: The Use of Formal Inspections at the Jet Propulsion Laboratory. Proc. of the IEEE 12th International Conference on Software Engineering, May 1990: 96-199.
4. Bush, M. Getting Started on Metrics: Jet Propulsion Laboratory Productivity and Quality. Proc. of the IEEE 12th International Conference on Software Engineering, May 1990: 133-42.
5. Rifkin, S. "Why New Software Processes Are Not Adopted." Ed. Marvin Zelkowitz. Advances in Computers Vol. 59 (2003): 22.
6. Grady, R., and Caswell, D. Software Metrics: Establishing a Company-Wide Program, Englewood Cliffs, NJ: Prentice-Hall, 1987.
7. Rifkin 12.
8. Trice 355-56.
9. Bush, M. Do CMM-Based Assessments for Internal Process Improvement Help Companies Stay More Competitive? Proc. of the European SEPG Conference, London, 1998.
10. Ahuja, Sanjiv. Process Improvement in a Rapidly Changing Business and Technical Environment. Proceedings of the European SEPG Conference, Amsterdam, 1999.
11. Billings, C., et al. "Journey to a Mature Software Process." IBM Systems Journal 33.1(1994): 46-61.
12. Pellegrino, J. "Birds of a Feather Session." SEPG Conference, 1995.
13. Howard, Peter. "Operating at Level 5." Internal BAE SYSTEMS Conference, Nov. 2002.

## Notes

1. Peer reviews in the CMM and CMMI concern the defect detection, removal, correction, and verification process carried out by small groups during the *pre-test* phases of the development life cycle. The primary objective of peer reviews is to remove defects *early* in the development process. Peer reviews supplement, not substitute for major milestone reviews. A trained moderator and a group of developers (limited to about four to six people) draw from the area of the life cycle being completed to carry out peer reviews. Everyone participating should have a vested interest in the work product. Peer reviews should never be used as a tool to evaluate workers or assign blame for defects. Team members, after undertaking special training, are assigned specific roles (for example, author, reader, recorder, moderator, etc.) Checklists of questions derived from previous experience are used to fine-tune defect finding. The checklist is regularly updated. Afterwards, statistics on the number and types of defects found and the time expended by engineers on peer reviews are kept as a historical database for later trend analysis. Peer reviews enhance the development life cycle by creating shorter feedback loops. They are not tied to any specific methodology or tool. They are usually done at the end of the following phases of the development life cycle: system requirements, functional design, software requirements, architectural design, detailed design, source code, test plan, and test specification. Peer reviews in the CMM constitute a Level 3 Key Process Area. By the time an organization has achieved Level 3 maturity, one would expect it to perform peer reviews on every project. In the CMMI, peer reviews are included not in a separate process area but as a Level 3 activity found in the process area called Verification, Goal 2 [5, 6].

2. Although these figures come from a 1990 study, and no more recent study is available, their accuracy has been informally confirmed in work with dozens of companies through 2003.

3. Presentation by A. Warman at GEC Marconi SPIRE 99.

## About the Author

**Marilyn Bush** is an independent management consultant who specializes in working with executive managers. She is one of the authors of "The Capability Maturity Model: Guidelines for Improving the Software Process" and the Software Engineering Institute's (SEI) Capability Maturity Model® (CMM®) Introductory Course, and is currently co-authoring a book on assessments. As a visiting scientist at the SEI, she helped revamp its CMM-Based Appraisal for Internal Process Improvement Assessment Method and Lead Assessor Course. She is a transition partner for the CMM Integration℠ (CMMI®) and is authorized to lead CMMI assessments (SCAMPI) and to teach the SEI Introduction to the CMMI. Bush, who divides her time between the United States and Europe, is an expert on how European and American software practices overlap and diverge.

**Marilyn Bush Associates**
**8037 Seminole AVE**
**Philadelphia, PA 19118**
**Phone and Fax: (215) 248-5391**

**United Kingdom:**
**39 ST Giles**
**Oxford OX1 3LW**
**Phone and Fax: 44 1865 512289**
**E-mail: m.w.bush@ieee.org**