28TH ANNUAL
PACIFIC NW
SOFTWARE
QUALITY
CONFERENCE

OCTOBER 18TH – 19TH, 2010

ACHIEVING QUALITY IN A COMPLEX ENVIRONMENT

Conference Paper Excerpt from the

CONFERENCE PROCEEDINGS

Assessing the Health of Your QA Organization

Michael L. Hoffman
capt.navar@gmail.com

Abstract

The aspiration of any software company or organization is the delivery of software products within defined goals of scope, and constraints of schedule and resources. The competitive nature of business requires that organizations live up to these goals, and work within these constraints, more effectively and efficiently over product lifecycle generations.

What constitutes improving over time for a software QA organization varies depending on perspective:

- For a tester it may mean more quickly differentiating between improper and expected behavior.
- For a customer support analyst it may mean less issues found in the field.
- For a test engineer it may mean optimizing test coverage based on risk assessment.
- For a quality engineer it may mean identifying issues in earlier lifecycle phases.
- For a QA manager it may mean making sure a quality/test team has the proper knowledge and tools in time to validate new technologies.
- For a QA director it may mean doing more with less.

The ability to consolidate so many perspectives into a comprehensive evaluation of improvement at the organization level is a complex challenge. A QA organization needs to regularly ask itself these questions:

How do we know we are doing the right thing?
How do we know we are improving over time?

Realizing quantitatively whether your software QA organization lives up to the dynamic needs of your business is not as simple as looking at defect trends over time. An appropriate evaluation involves various aspects of a software QA function, including: operational behaviors, talent, customers, and budgets.

This paper chronicles an approach taken by one software QA organization to evaluate its own health through establishing goals, benchmarking, defining organization-level metrics, and on-going self-assessment.

Biography

Michael L. Hoffman is a 20-year veteran of the software development/QA industry. He has filled a variety of engineering, management, and strategist roles at companies including Jeppesen Sanderson, Hewlett-Packard, Insight Distribution Software, Systematic Designs, Sharp Microelectronics, and GTE Government Systems. Michael's experience includes the areas of communications systems, factory control/integration, embedded systems, PC applications, and web-based solutions. Additionally, Michael teaches various software courses for Oregon Institute of Technology and is Chair of Oregon Tech's software engineering department industry advisory board.

1 Introduction

The success of delivering products and/or services from the perspective of a software QA organization is typically measured by looking at factors such as working within budgets, timely releases, and number/severity of field defects. Metrics around these factors tell how well the organization is doing, but do not indicate whether the organization is effectively aligned with the business.

A good team or organization can perform well, deliver on time, have low field/warranty costs, and still fail to meet goals; particularly if there is not clarity as to how organizational performance is being measured.

It's not matter of accountability. It's a matter of being accountable for the right thing.

Below is the experience that one software quality organization undertook to assess its delivery of value to the company and to ensure that it maintained that value over time.

2 The Challenge

Our Software Quality Assurance (SQA) organization was made up of seven geographically-separated teams that focused on testing and quality improvement of software for consumer electronic products.

Following a high-level company reorganization, SQA reported to a vice president who had not previously been responsible for a QA function. To become familiar with our organization, the vice president requested an overview presentation.

Nearly a week was spent by the SQA leadership team planning, preparing, and practicing our shtick; the resulting presentation was seamlessly delivered. Our goal of showing that we had a well-organized, cohesive SQA organization that consistently delivered on-time, in budget, products, with low field/warranty costs was accomplished.

Our audience, the vice president, accepted our assertions as to our success and did not doubt that there was high value in what we did. Following the presentation he asked us three simple questions:

> *How do you know you are doing the right thing?*
> *How do you know you are improving over time?*
> *How do you compare to similar organizations at our competitors?*

What he was really asking was for us to prove to him that what we were doing was what the business needed from us, that we were getting better and better at our role within the company, and that our approach was the best possible.

We were not prepared for, nor was the vice president expecting, an on-the-spot response to that question. We left the meeting with the task of determining the answers.

3 Determining the Approach

SQA management determined the current approach to metrics was accurate at showing whether we were successful in qualifying the readiness of products for the company, but they also conceded that providing this information was just not sufficient in showing whether we were delivering to the business' goals or that we were constantly improving our value to the company over time.

A cross-organization project team was formed with the goal of defining and driving a plan that would lead to a way we could assess SQA that answered the three key questions posed by our vice president.

The following approach was planned:

- Taking stock
- Benchmarking
- Understanding where we stand
- Defining measures
- A first assessment
- Setting thresholds and actions
- Continuing to focus on the right thing

4 Taking Stock

In today's work environment, virtual teams introduce a variety of factors that impact the effectiveness of an organization. Even people within the same reporting structure can have vastly different views of their organization, influenced by factors such as business unit history, target market segments, culture, and personal experience.

We realized very quickly the need to have a common realization of our own SQA organization before we could effectively benchmark with other companies. We developed a template to document our own self-assessment. In doing so, we also had a tool for sharing information with benchmarking partners and for documenting our assessments of other companies. Our assessments were organized into key categories:

Category	Description	Discussion Points
Organizational Model	How it impacts objectives, priority-setting, and decision-making.	Organization structure Management of outsourcing
Relationship to Customer	How SQA is assessed and held accountable for its responsibilities.	Customers' view of the quality organization Relationships with partner organizations Assessing satisfaction
Roles and Responsibilities	SQA roles, their respective responsibilities, and the partnerships with external teams.	Roles/responsibilities within SQA Assessing/tracking performance On-going individual development
Product Assessment	How product readiness is assessed.	Tracking the health of products Assessing product readiness Historical/predictive analysis
Process Assessment	How the effectiveness of quality/test processes is assessed.	Lifecycle processes Proactive quality processes Assessing process effectiveness Measuring the right things
Organizational Assessment	How the health of the quality organization is assessed and improved-upon.	Responsibility for organization health Assessing/measuring organization health Alignment of organization goals
Organization Performance	How the quality organization is measured by the company.	Measuring organization performance Effective metrics at an organizational level

Although much of the self-assessment was simply a matter of bringing together existing information, the exercise itself was valuable from a few perspectives. It provided us with:

- a realization of levels at which we already were doing self-assessments
- some obvious self-assessment gaps that, until this point, had never been recognized by our leadership team

- a trigger for conversation topics for our upcoming benchmarking efforts
- the basis for a set of information we planned to share with participating benchmarking partners

5 Benchmarking

The goal of benchmarking was to discuss the key categories with SQA organizations outside of our own.

Benchmarking was approached as a mutually beneficial opportunity. As part of the process of soliciting benchmarking participants, we offered an open, two-way conversation (any question we asked of them, we'd be willing to answer ourselves) and we committed to sharing our final benchmarking report.

In order to answer the question "how do we compare to our competitors?", it would have been ideal to benchmark with a competitor. That, however, certainly was not realistic. We did not want to share our business practices with competitors and we suspected they would feel the same way about sharing with us. Alternately, we chose to benchmark with companies that had similar target market segments (high volume consumer electronics), but who produced products that did not compete with our company.

Additionally, we arranged to talk with a couple of companies in vastly different industries so as to include a diverse set of perspectives. Two non-producers of consumer products were chosen: one that focused on the service/logistics industry and the other a major defense contractor. All participants were representatives from quality assurance teams.

Participating companies were provided with a general list of discussion topics (see above) based on the key categories, to set the expectation of discussion scope and to facilitate productive conversations. In a few cases, benchmarking participants reciprocated with additional questions that were incorporated into the benchmarking discussions.

Benchmarking was accomplished with half-day or full-day sessions based on time availability of participants. All sessions were conducted in person, except for one session that was conducted via video conference.

6 Understanding Where We Stand

Each company involved in the benchmarking effort was successful in qualifying their products for market readiness, but no two companies did so in the same way, nor did they use the same measures to assess their success or determine if they were improving over time. This was a result of a number of factors:

- The nature/complexity of the technology in their products
- Commitment of the company toward the QA function
- Maturity of their SQA organization in terms of quality practices
- Resource (people and/or budget) constraints

Consequently, realizing how we compared to the QA teams with which we benchmarked became a subjective activity. Since benchmarking discussions were focused on category-based questions, the categories were used as the comparison points. For each category, the relevant learnings that came out of the benchmarking with each company were discussed and voted upon by our benchmarking team. Using ourselves as a midpoint for a scale, the results provided a picture of our relative comparison to each company.

7 Defining Measures

From our benchmarking results, it was clear that other QA organizations were similarly strong in defining metrics to assess the readiness of their products and the efficiency of processes. Some of the metrics used included:

- Number of defects per lines of code
- Test effort (cost) per defect found
- Mean time to fix / validation
- Number of re-opened defects
- Ratio of automated vs. manual test executions
- Coverage analysis
- Percentage of test plan reviews
- Number of escapes
- Number of quality audits

Despite the varying approaches to assessing product readiness and ensuring minimal defect escapes, there was very limited, if any, focus on assessment of improvement over time. Although this was reassuring in the sense that what we were trying to do was, in fact, beyond the normal focus of self-assessment and effectiveness/efficiency improvement, it showed that there was no simple, easy-to-implement method to accomplish what we wanted.

To move forward from this point, we focused on our company's scorecard provided by senior management. This included major categories of:

- Financial
- Customer
- Operational Efficiency
- Employee

Whatever measures of assessment we used, they needed to convey a story of organization wellness relative to these four focus areas. It was decided that our highest level organization metrics would be broken down by these four areas.

7.1 Financial

This area was straight-forward … measures were dictated by our company's financial organization for the purpose of consistency across the company, and there were tools and processes in place to collect data and provide analytical summaries. All that would be involved for SQA would be to include the already-generated measures as our part of our organization's self-assessment, and develop appropriate action plans as needed.

7.2 Customer

There was consensus that a metric around defect escapes into the field for previously undiscovered issues would give us the ability to assess SQA's effectiveness at determining readiness of a software release. The same is true for issues found during customer beta releases, but this metric would require consideration that software versions typically get released for beta testing before full regression testing has been performed.

From the benchmarking effort, it was discovered that many QA organizations own the responsibility of qualifying product usability in their charter. In fact, it was not uncommon for testers or test engineers to submit defects for areas where usability specifications were not met, were vague, or were undefined. In

the end, however, it was decided not to include this metric in the assessment of SQA since in our case, the evaluation of customer experience was owned outside of SQA, by a customer satisfaction team. The purpose of our metrics effort was to assess our organization's health, and that such a metric would actually be measuring how well our product met our customers' needs and not a reflection our organization's effectiveness.

7.3 Operational Efficiency

This area focused on domain-specific assessment. In our case, the domain was software quality/test engineering. The metrics selected were broken into five categories:

Categories	A measure of our organization's ability to…	Metrics
Readiness	Meet delivery milestones.	Specification readiness Test case readiness Product readiness*
Program Fulfillment	Live up to the intent of the program.	# of escalations # of waivers
Effectiveness	Do the right thing.	Defect removal Spoilage Lifecycle testing
Efficiency	Use less resources over time.	Effort per defect Defect merit
Initiative Status	Make timely progress on improvement efforts.	Varied by initiative

** Although this metric is actually a measurement of the timely delivery by development teams, the ability for SQA to deliver on its commitments on-time to the business are highly dependent on this factor. Additionally, this milestone does reflect effective collaboration between SQA and Development, so it includes aspects of self-assessment.*

7.4 Employee

We learned from the benchmark discussions with other companies that assessing employees is an area where there was no consistency other than tools such as performance reviews. But what we wanted to achieve was the ability to answer these kinds of questions:

1 Do our teams have the right skill sets needed to work effectively?
2 Are we prepared for what the future will demand of our organization (technology, partner-team changes, business model, etc.)?
3 Are employees satisfied with their roles in the organization?

#1 and #2 are really the same question, just focusing on different timeframes. We quickly realized that although there are tools for assessing the readiness of an organization in terms of meeting the needs of the future, we did not have a good handle on what would be expected of us down the road. Our business was changing (i.e. new market segments), technology was evolving, and new working models were being established (i.e. engaging vendors). This required new skill sets, and we did not know what those skill sets would be. We determined that before we could assess whether we had a sufficient workforce that was progressively getting better, we needed to put some work into determining goals to measure against. A side project was created to focus on determining skill sets for each of our organization roles, consider each person's capabilities against the roles they were filling, and from that generate assessments of two factors:

- Skill levels vs. roles
- Capable resources vs. future need

Note: a valuable bi-product of going through this process was information for each manager as to the development opportunities for each of their direct reports.

For #3, assessing employee satisfaction, a survey approach was deemed most appropriate. A survey was developed to gather employees' feelings toward their understanding of SQA's goals, their role within SQA, their feeling of support, and opportunities available to them to develop in the ways they wanted.

8 A First Assessment

The goals of taking a first assessment were to:

- Gain an understanding of the lower-level metrics and/or data acquisition processes/tools that were necessary to generate our higher-level metrics.
- Acquire data that could be analyzed to understand the validity of our measurements so we could determine if we were, in fact, measuring what we intended.
- Create baseline metrics values to compare against as we moved forward.

It quickly became apparent that we did not have sufficient processes and tools for the collection of the raw data needed for some of our metrics. For example, the readiness metric required tracking of deliverables (specification documents, test cases, product functionality) relative to milestones. Although we already had good practices in place for determining whether a milestone deliverable was met, we had no means of accurately or consistently recording these activities. Maintaining a spreadsheet checklist for each product was easily instigated as a short-term solution. Ultimately, a more robust web-based tool that allowed for setting up and tracking of deliverables on a product-by-product basis across all teams within SQA would be put in place.

In some cases, there was no short-term solution for a metric, and a bigger effort was required to reach the point where we had a necessary foundation for our metrics. The spoilage effectiveness metric is such an example. One of the key components to this metric is classifying a defect's root cause. Although our defect tracking system supported assigning a root cause to a defect, this was not a required field; plus, there was great inconsistency when this field was used. It became clear that simply changing our tool to require an indication of root case would not result in reliable data for our metrics. To achieve this, we needed to instill a cultural change in our partner development teams to realize the intent of the field and to adopt a behavior of determining "why" the code change was necessary. Was it due to:

- An incorrect implementation of a design?
- The correct implementation of a flawed design?
- A problem with the requirement?
- …

Influencing our partner development teams to encourage investigation into the true root cause of each defect did not happen overnight. It required a behavioral shift that would mature over time. Additionally, the task of going through tens of thousands of previously-resolved defects that had been created over numerous years and dozens of product cycles was deemed an unrealistic endeavor. As a result, it would be a significant period of time, on the order of a few product lifecycle generations (about one year), before a set of data would exist that could yield meaningful results for this metric.

Consequently, some metrics would be a "work in progress" until such a time that processes or tools matured, behaviors changed, and raw data became available.

9 Setting Thresholds and Actions

For those areas where valid, reliable data was available and metrics could be calculated and assessed, the next step was to determine what constituted acceptable results.

Given the broad span of metrics we chose, normalizing measurement units across the different metrics categories was unrealistic. We determined that each metric needed to be assessed independently and then acceptable vs. non-acceptable results reported in a common manner across all categories. Additionally, simply reporting a quantitative value for a metric was not sufficient. The current trend for each metric also needed to be determined and reported to fully convey whether improvement was being realized or whether we were losing ground in particular areas.

To normalize the assessment of both quantitative and trend results across all metrics, we adopted a simple color-coded reporting scheme:

Color	Meaning	Action
Green	Current quantitative results are within desired thresholds and trending indicates they will remain that way for at least the next reporting cycle.	None needed
Yellow	Current results are good but trending shows we may miss our threshold in the next reporting cycle. OR Current results are not within the target threshold but trending indicates it will be within the threshold by next reporting cycle.	Minor corrective action needed
Red	Current results are far outside the target threshold. OR Current results are not within the target threshold and trending says it will get worse.	Major corrective action required

This approach gave us the ability to use different units and/or scales and to track trends differently for each metric, yet use a consistent reporting technique for communicating overall results.

An attempt was made when initial thresholds were set for each metric to define the corrective action that would be taken for yellow and red conditions. Over time, however, we realized that the corrective actions we initially defined were not always appropriate to address particular situations. Consequently, we changed to an approach of determining appropriate corrective actions if, and when, a yellow or red situation occurred. Actions could then be designed and employed to address the exact root cause.

We encountered situations where there was an inverse relationship between metrics. This was particularly true in the area of operational efficiency. For example, often a positive trend for an effectiveness metric would result in a negative trend in an efficiency metric. At first seeing such results was disconcerting, but after understanding how improvements in one area can cause inefficiencies in other aspects of our operation, we realized that although the metrics results were not where we wanted them in all cases, it was a true reflection of where SQA had the opportunity to improve in the future.

The outcome was a realization that a narrative around our metrics was just as valuable, if not more valuable, than the status of the metrics themselves. The work that went into understanding why the metrics were what they were and why they were trending in a certain direction yielded the most useful information of all, both to us and the recipients of our assessment reports. It gave us the ability to:

- understand if we were doing the right thing
- assess whether we were getting better over time
- realize our opportunities for improvement

… which is what this endeavor was all about: assessing the health of our SQA organization.

10 Continuing to Focus on the Right Thing

Determining the period of time for generating and assessing SQA organization metrics turned out to not be as simple of an effort as we anticipated. Some of our metrics were driven by product life-cycles (which varied greatly depending on the complexity of products), some by product release dates (i.e. peak selling seasons), some by financial milestones (i.e. fiscal quarters/years), etc. There was no one-size-fits-all period for assessment.

At the same time, selecting a fixed period (i.e. monthly) of collecting data, generating metrics, and doing assessments did not make sense either. A specific metric does not necessarily yield good results based on a periodic calendar date. For example, calculating the number of field escape defects for a product on the first of each month does not make sense while the product is in development. That kind of metric needs to be generated only after the product is released.

We chose to collect data and calculate each metric separately, based on the period for that metric which yielded a meaningful result. Assessments, however, were made based on a fixed period of time; we chose quarterly. Although not all metrics were updated each quarter, we did have the ability to report our organization assessment regularly. Plus, this ensured that what we reported accurately reflected the results for the most-recent time the metric was generated.

11 Summary

If there is one truth we learned as a result of this effort, it's that complacency in a QA organization leads to lower quality over time.

A company that produces a market-leading product will succumb to eventual competition if it does not continue to innovate and become more operationally efficient in how it brings products to market. The same is true with complacency in a software quality organization. Without continuous innovation in our approaches to quality, and on-going improvements across all facets of SQA, the quality of products will degrade over time due to the ever-increasing demands of the product's technology and feature set, and the ever-shrinking budget and time-to-market constraints.

Using an organization assessment approach such as the one described in this paper will help in determining if there is, in fact, improvement over time. However, we must continuously ask whether the improvements we are making are really what the business needs. So in addition to looking at our metrics to determine whether we are getting better, we also need to continuously ask whether we're measuring the right things.

Adopting an approach to regularly ask ourselves the fundamental question "Are we doing the right thing?" will help determine if the metrics accurately reflect what the organization needs to know as to how it effectively and efficiently it delivers to the needs of the company.