

Technology for Team-Based Quality Assessments

Craig Kaplan

Abstract

Periodic, team-based quality assessments were a key element in a TQM effort that allowed IBM to reduce service costs and defects while increasing customer satisfaction and revenue per employee. Team-based assessment software is a highly cost-effective approach to periodic quality assessment. By averaging scores and merging comments from many assessors, team-based software increases the validity of a self-assessment at the same time that it increases involvement and understanding. Combined with leadership and other widely available technologies, team-based assessment software has helped organizations ranging from the U.S. Air Force to AT&T address seven major challenges to quality assessment efforts.

Author Biography

Craig Kaplan is CEO of I.Q. Company (<http://www.iqco.com>), a software development and consulting company based in Santa Cruz, California serving more than 175 client organizations including IBM, AT&T, Siemens, and the US Air Force. Dr. Kaplan received his M.S. and Ph.D. degrees from Carnegie Mellon University and co-authored the book, Secrets of Software Quality: 40 Innovations From IBM (McGraw-Hill 1995). He is a member of the ASQC, ACM, APA, AMA, IEEE, and the Scientific Research Society, Sigma Xi.

CEOs invest in Total Quality Management (TQM) programs because they want to improve the bottom line. They are betting that TQM will lead to higher profits, either by reducing costs or increasing revenue from more satisfied customers. For example, TQM efforts at IBM's largest software development site began in earnest when a new General Manager discovered that 60 - 80 cents of every development dollar was being spent on software maintenance and rework. TQM represented an opportunity to reduce these costs.

In IBM's case the TQM investment paid off. Exhibit 1 summarizes the results achieved between 1989 and 1993. The primary goal of cost savings was realized with service costs decreasing by 20%. At the same time, revenues grew as customer satisfaction with key products increased 14% relative to competitors. Underlying these financial results were more than 40 leadership, process, and technology improvement techniques which enabled IBM to reduce its software defect levels by 46% and increase revenue per employee by 56%. Together, these factors helped IBM's largest software site remain profitable at a time when much of the rest of the company was experiencing financial distress.

EXHIBIT 1: Quality Results at IBM's Largest Software Site

Team-based quality assessments were probably the most important element of IBM's successful TQM effort.

Why You Need Computerized, Team-Based Assessments

Quality assessments help you form a clear idea of where you are, and of what you need to improve. Team-based assessments provide a more accurate view than a single assessor can provide. Team-based assessments also foster buy-in because they involve more people. Periodic assessments help you track progress over time and maintain focus. Finally, computerized assessments yield time, resource, and effectiveness benefits compared with low-tech assessment methods.

Choosing a Quality Assessment

Which quality assessment criteria you choose is less important than your commitment to involve the entire organization in the assessment, and to follow up each assessment with concrete quality improvement actions. You should choose an assessment that matches your goals. If your goal is quality improvement on a broad scale, you might consider the Malcom Baldrige National Quality Award criteria or other state quality award criteria that are based on Baldrige. If your focus is primarily on process, or if you need quality certification, the ISO 9000 series might be a logical choice. Finally, there are quality assessments and models for specific industries, such as the Software Engineering Institute's CMM model for the software industry.

IBM chose the Malcom Baldrige National Quality Award as its primary quality assessment because the Baldrige criteria are so comprehensive. The seven Baldrige categories cover all aspects of a business ranging from leadership to process improvement to business results and customer satisfaction. Although IBM's largest software site also sought (and obtained) ISO 9001 registration, ISO is much narrower in scope than Baldrige. In fact, at the time of the IBM site's ISO registration most of scope of ISO 9001 could fit within the scope of a few of the 1994 Baldrige categories, (Exhibit 2).

EXHIBIT 2: Comparison of ISO 9001 and 1994 Baldrige Criteria Scope

Once you have chosen an assessment, the real work begins. Sound management practice combined with appropriate use of technology can help you overcome obstacles and sustain a successful quality improvement effort.

Technology and Seven Challenges to Quality Assessment

Most large organizations will face assessment challenges in the following seven areas: 1. Commitment, 2. Communication, 3. Skills & Training, 4. Validity, 5. Costs, 6. Making Improvements, and 7. Staff Burnout & Assessment Transitions. Combined with strong leadership, information technology -- specifically team-based assessment software-- can help an organization tackle each area.

Note: For an annually updated listing of specific software vendors that can supply some of the technology discussed below, check the March issue of *Quality Progress* magazine.

Gaining Commitment

Leadership plays the central role in gaining buy-in, while technology plays a supporting role. In order for a large scale quality improvement effort to have a reasonable chance of success, senior management must be committed. Typically commitment at the senior level comes because these executives realize that cost savings and other competitive advantages will accrue from sustained quality improvement efforts. But even once benefits are understood intellectually, true buy-in doesn't occur until everyone begins actually implementing and taking personal responsibility for quality improvement.

Commitment follows involvement. Technology simply makes involvement easier in the same way that it is easier to convince people to help you build a house if you allow them the option of using power tools instead of insisting that everything be done by hand.

Periodic assessments are a tried and true method of getting everyone involved. At IBM, one site General Manager insisted on quarterly management reviews of quality improvement activities. Similarly, the U.S. Air Force conducts periodic quality self-assessments of Air Force units world-wide. Other organizations ranging from the city of Austin, TX to AT&T all use periodic quality assessments not only to gain information, but to raise awareness and gain buy-in to improvement efforts.

The most specialized power tools in the assessment world are computerized team assessment products. Most assessment software supports automatic scoring and graphing of results as well as an easy way to enter comments and other information that supports the scores (Exhibit 3). These standard features reflect an emphasis on increasing efficiency by eliminating the drudgery involved in calculating scores or creating reports with comments and graphs. However, new assessment software programs are also beginning to address issues which affect effectiveness such as relevance, and team participation.

EXHIBIT 3: Sample Question Screen from an Assessment Software Product

Unfortunately, there is a relevance gap between the generic wording of most popular assessment criteria (e.g. Baldrige or ISO 9000 series) and the specific day-to-day activities of an organization. Unless you bridge this relevance gap in a meaningful way, you get resistance instead of buy-in.

Training is an effective but expensive way of bridging the gap. However in addition to teaching the organization the language of the criteria, many government and private organizations are beginning to adapt the criteria's language *to the organization*. The trend can be seen in the proliferation of state quality awards, almost all of which are modeled on the Baldrige award, but many of which have been adapted to fit the specific needs of a state's applicants. Companies do the same thing. For example, at IBM, the sometimes obtuse language of the Baldrige criteria was routinely translated into terms software developers could understand.

Even Federal organizations which are committed to the Baldrige award, often modify the wording to better fit their constituencies. For example, the U.S. Air Force uses Baldrige as the basis for its "Unit Self-Assessments." However, the wording is often customized to reflect the military character of the Air Force (e.g. substituting "Base Commander" for "Senior Executive").

New assessment software can help bridge the relevance gap in two ways. First, some assessment software includes online notes and advice that help interpret the criteria. This trend parallels the proliferation of books by consultants aimed at helping organizations make sense out of the different assessment criteria. Second, some assessment software packages can be rapidly customized to include notes, changes to question wording, or even new questions and point values in accordance with an organization's needs. Software assessments that include extensive online help and that allow customization, help increase buy-in to the overall assessment process by bridging the relevancy gap.

Since buy-in follows involvement, assessment software that supports participation by teams also helps deepen commitment within the organization. One of the difficulties with many assessment software packages is that they implicitly assume that a single assessor will be completing the assessment. Unfortunately, even if a single assessor were

able to complete an accurate assessment, the single assessor approach implies that quality is the responsibility of a single individual (e.g. the Quality Assurance manager). In reality, quality improvement efforts are much more successful if they are approached as team-based efforts requiring everyone's involvement.

Team-based assessment software allows you to merge the scores and comments from many assessors, and even from many team of assessors, into a combined assessment. Team-based software also allows you to read comments from different assessors, view average scores from all the assessors, and export the comments and scores in a form suitable for editing into a final report. Where organizations previously used a combination of spreadsheets, databases, and word processors to conduct team assessments, many groups from the U.S. Air Force, Siemens, and AT&T, are now using team-based assessment software to achieve the same effect more easily.

To recap, assessment software that is team-based and relevant (e.g. customizable, and friendly in terms of providing tips and advice online) has the best chance of helping an assessment effort gain wide-spread acceptance and buy-in.

Communication Technologies

The good news is that many of the communications technologies that facilitate quality assessments are already being used by most organizations. For example, Email is a natural way to share information among team members conducting assessments -- especially if the people involved are in different physical locations.

Many organizations approach the task of a Baldrige assessment by dividing the seven categories among seven team leaders. Each team leader works with several people and information is shuttled back and forth using Email and electronic transfer of word processing documents and spreadsheets.

Within this general team-based approach, variations are possible. For example, IBM used many cross-functional teams, each responsible for conducting a complete Baldrige Assessment in a particular product or process area. Then all of the assessments were rolled up into one overall assessment for the entire site.

As the scope of the assessment effort increases, communication technology becomes increasingly important since it becomes increasingly difficult to communicate directly, face to face. Thus, in addition to Email, IBM set up electronic bulletin boards to allow anyone at the site to comment on quality improvement issues. The site General Manager often personally responded to posts on the bulletin boards.

Everyday communication technologies such as teleconferencing, video taping, and photocopying can also play an important role in improving communication in large scale assessment efforts. AT&T used teleconferences to gather input from team members in different locations. One IBM site used to videotape its quarterly Baldrige assessment presentations, and went so far as to broadcast several of them live, throughout the site.

Surprisingly, one of the most effective vehicles for raising awareness about specific quality innovations at IBM was a humble newsletter produced by one person who knew lots of programmers with good ideas. He got the programmers to write-up what they were doing. Then he edited the articles, photocopied them, and distributed them throughout the site. The articles proved to be a valuable source of information that helped assessment teams understand what activities were going on in parts of the site outside of their normal experience.

Beyond the general communication technologies just mentioned, organizations may want to consider two less common approaches that can greatly enhance and focus communication between assessment team members. The first approach, Computer Supported Team Workspaces, was originally developed by IBM to help teams of programmers accomplish their work more effectively. The concept worked so well that it was borrowed by teams of all sorts including IBM's assessment team members who used it to help write the site's 70-page Baldrige report.

IBM's approach was to computerize a meeting room by adding a computer with the capability to project the computer screen on the wall. Assessment teams gathered in the room to review drafts of the assessment document which were projected directly from a computer to a large wall screen. People's comments, suggestions, and notes were added directly to the document during the meeting. Consensus was achieved on phrasing, and in some cases missing data or information could be directly brought up on the computer from other sources that were electronically linked.

An internal IBM study estimated that the site could save up to 15% of its direct labor costs by conducting design sessions, code reviews, document reviews, and other team-oriented meetings in the Computer Support Team Workspaces. Much of the savings resulted from the fact that changes could be made and consensus reached during the meeting instead of requiring additional meetings to review changes later. Subjectively, there is something powerful about seeing your ideas immediately captured and documented on a big screen. Some of the brainstorming sessions went on so long that the joke was you ran out of air (because the computers generated heat in the enclosed room) long before you ran out of ideas.

Although IBM eventually built special ergonomic rooms for its Computer Supported Team Workspaces, you don't need a fancy room to try out the idea. At your next meeting, just hook up a computer to a LCD panel projection panel on top of an overhead projector. Open the assessment document on your computer, and ask a member of the group to make changes to the document and type comments from the group as they arise in discussion. The fact that the group can see the changes immediately on the projected screen creates an effective feedback loop. The miscommunication and misunderstandings that inevitably develop when each person takes his or her own notes disappear as group members debate over what is on the projected screen, clarifying points, and hopefully reaching consensus. Perhaps the most satisfying aspect of this approach is that the team

can leave the meeting have made actual changes in the work product -- something that is rare in many corporate meetings.

A second approach to improve communication between team members is to cross reference assessment items. For example, one of the communication problems that almost all Baldrige teams encounter is the issue of determining linkages between different items of the Baldrige criteria. One team member might be the leader for Category 5 which includes the item, "Management of Supplier Performance," while another team member might be focusing on Category 6 which includes the item, "Supplier Performance Results." These two team members need to talk, so they can avoid duplication of effort and share information.

The U.S. Air Force solved the linkage problem by creating a series of tables that cross-referenced related Baldrige items. The same approach has been incorporated into team-based assessment software used by the Air Force. Nicknamed "golden threads," hypertext links allow assessors to instantly jump to descriptions of all the items related to the one they are currently working on (Exhibit 4). In a merged assessment containing the scores and comments of an entire team, the golden threads allow team members to quickly compare everyone's scores and comments for all related assessment items.

EXHIBIT 4: Golden Thread Linkages in the Air Force Quality Assessment Software

To recap, many existing communication technologies such as email, electronic bulletin boards, teleconferencing, and even photocopied newsletters can facilitate large scale team assessment efforts. Computer Supported Team Workspaces and electronic cross-referencing of assessment items are other examples of specific technology applications that can improve communication between team members.

Skills and Training

Even if you customize the assessment criteria to make it maximally relevant to your organization, you are going to have to provide some level of assessment training. Although it's hard to beat a good human instructor for effectiveness, good instructors also tend to be costly. For example, a two day class with a Baldrige Examiner can easily run \$10K plus expenses.

Less costly alternatives include joining non-profit associations that make training available to their members (e.g. The Council for Continuous Improvement), purchasing videotaped training, or relying on a variety of computerized options ranging from computer-based training programs to online tutorials and the online help built into assessment software programs.

Internet surfers might want to peruse Baldrige information that is available free of charge from the National Institute of Standards and Technology (<http://www.nist.gov>). Other useful sites are the American Society for Quality Control (<http://www.asqc.org>) which

points to a variety of other quality improvement sources, and I.Q. Company's site (<http://www.iqco.com>) which features a free searchable database of quality-related resources and referral information.

A common training strategy is to provide team leaders with intensive formal assessment training, and then have these leaders informally train the other team members. Since the quality and quantity of training varies with this approach, it is a good idea to ensure that the assessment software you use includes adequate scoring guidelines and assessment help. Otherwise, you'll find that the scores and comments provided by the team members will vary so widely that consensus becomes difficult to achieve.

In short, you want to ensure that enough training is provided (electronically or otherwise) so that there is consistency in the way teams are interpreting the criteria and scoring the organization.

Validity

Related to the issue of scoring consistency, is the concept of validity. For a self-assessment to be valid, it must capture the views of a representative sample of the organization's population. It should also be based on objective data rather than on subjective opinion or "gut feel." Usually anchoring scores in objective data (i.e. increasing the validity of the score) also leads to more consistent scoring. For this reason, an improvement in scoring consistency (when accompanied by data supporting the scores) can indicate an improvement in validity as well.

Senior management, especially, is notorious for scoring organizations as they think the organization should be, rather than as the organization actually is. Tempering the "vision" of senior management with concrete data, and the views of others in the organization, helps provide an accurate assessment of the organization's true quality maturity.

Sometimes it is even helpful to ask assessors from outside the organization to participate since they tend to have a less biased view. By tracking scores of various groups over several periodic assessments you can see whether the scores begin to converge. If they converge, as they did in IBM's case (see Exhibit 5), it is probably a good indication that the teams conducting the assessment are developing a better understanding of the criteria and are beginning to score more consistently. This is especially true if one of the groups doing the scoring consists of highly trained individuals (e.g. Baldrige Examiners) whose scores can be assumed a priori to have a relatively high degree of validity.

EXHIBIT 5: Convergence of Scores From Two Groups of Assessors at IBM Over Four Successive Baldrige Self-Assessments

Technology can help by automating the process of combining many diverse points of view and by providing mechanisms for capturing objective data. In years past, spreadsheets and word processing programs were the primary means of averaging scores,

capturing comments, and producing reports and graphs. Today, team-based assessment software performs the many of the same functions in one integrated package.

For example, team-based software can help focus attention on areas where assessor disagree by plotting the scores of different assessors (or teams of assessors) on the same graph (Exhibit 6). Items with sharply divergent scores probably need more examination. Some assessment software not only averages scores from many assessors but also calculates statistics that show quantitatively where assessors agree and disagree. The advantage of such statistics is that they, like the graph in Exhibit 6, help focus discussion and attention on the areas where there seems to be the most disagreement.

During discussion and the process of reaching consensus assessors learn from colleagues who scored the organization differently. Ideally, the consensus process improves the validity of the assessment not simply through compromise, but by bringing to light new facts and information known previously to only some of the assessors.

EXHIBIT 6: Graphing Scores From Multiple Assessors Using Team-Based Software

The best assessment software goes beyond merely asking for a numerical score. Most criteria require assessors to document and justify scores by providing data and evidence that a particular criterion has been satisfied. The assessment software should provide a means (e.g. a large-capacity comment box) for capturing notes, and other data that support the score. Ideally the software should support the capability to incorporate graphics, images, and tables directly into the assessment. In the absence of such features however, it should be possible at least to export assessment information to other programs that support graphical linking such as Microsoft Word.

To recap, gaining several points of view and insisting that scores be supported by data are the primary methods for increasing validity. Software assessments that allow merging assessments, comparing assessors statistically and graphically, capturing comments, and exporting assessment information help increase both usability and validity.

Cost

The major cost of any large scale assessment effort is the time of the people involved. The cost of technology that can increase the efficiency of the assessment effort typically pales by comparison. For example, consider the case of IBM's largest software site. Using estimated figures published in the book, *Secrets of Software Quality*, and the assumption of a burdened labor rate of \$100K/per person-year, we get the following estimated time cost:

# of Person-Years Expended in Four Quarterly Baldrige Assessments:	72
# of Additional Person-Years Expended to Secure ISO 9001 Registration:	50
Total # of Person-Years:	122
Total Estimated Time Cost = 122 P.Y. X \$100K/P.Y. = \$12.2M	

Compare this figure of \$12.2 Million with typical costs for Team-Based Assessment Software for Baldrige plus the incremental cost for adding an ISO 9001 module:

Cost for a Single User Team-Based Baldrige Assessment Software License:	\$600
Cost for a Single User License for additional ISO 9001 assessment:	\$200
Total Software Cost/User:	\$800
# of Licenses Required (10% of Site Population):	180
Total Estimated Software Cost = 180 X \$800 = \$144K	

The cost of Team-Based Assessment Software is roughly 1% of the cost of the overall time involved in conducting four assessments. Even if the software provided only a 10% increase in efficiency (which is low by most estimates), an organization could save ten dollars in personnel costs for every dollar invested in assessment software. More importantly, the assessment would be easier to conduct and more valid because it could include more people.

Improvement Steps

Strictly speaking, assessments are designed to tell you where you are, not necessarily what you should do to improve. In fact, Baldrige makes a specific point of emphasizing that the criteria are not prescriptive. And yet, pragmatically, there is not much point in doing an assessment unless you plan to take the next steps to actually improve.

The ISO 9000 series incorporates the notion of Corrective Action Requests which address this issue, but still the organization is left to its own devices when it comes to figuring out what corrective actions to take.

As a response to the need to take next steps towards quality improvement, several software companies have developed products which target process improvement. There are flowcharting tools, brainstorming tools, database tools for managing teams, Reengineering tools, tools for creating cause and effect diagrams, and tools for creating control charts and graphs. Many of these tools have been integrated into quality improvement suites. It is possible to buy quality improvement toolkits separately, or bundled with assessment software.

Some assessment software even includes a built-in database of quality improvement resources. Based on your assessment score, the software automatically retrieves recommended titles of books, articles, software, or other resources from the built-in database. You click on titles of interest to view resource abstracts (see Exhibit 7).

The software even allows the addition of new resources to the database that might be specific to the organization using the assessment. For example, the U.S. Air Force uses assessment software that includes a customized database with material specific to the Air Force. Other organizations can request inclusion of descriptions of their key quality policies and procedures as related to various categories of an assessment.

EXHIBIT 7: Abstract of a Quality Improvement Resource

It is possible to use both team-based assessment software with built-in improvement resources and specific follow-on tools such as flowcharting or brainstorming software. The key point is that for an assessment to be most effective, there should be a link between the result of the assessments and the next steps which will lead to actual improvement.

Handling Staff Burnout and Transitions to New Assessments

One of the most serious challenges to IBM's TQM efforts, was the "burnout" of employees who had worked many months on quality improvement without being able to see clear results. Unfortunately, there is a lag period between when a process is first improved, and when new products created with the improved process finally reach customers. Depending on the length of the product cycle, that time lag can be two years or more -- a long time to wait in Corporate America.

Periodic assessments help ameliorate this problem because the assessments typically give credit for changes in processes well before financial results of these changes are apparent. During IBM's four-year quality improvement journey, there was a point, between year 2 and 3, when the main thing that kept the site motivated was the steadily increasing score on the Baldrige criteria. Without that internal gauge of progress, many managers and professionals would have given up since it took several years (see Exhibit 1) for financial results to appear.

A related challenge is managing the transition from one assessment to another. Transitions can be minor (e.g. the shift from 1994 Baldrige to 1995 Baldrige). Or they can be major (e.g. going from Baldrige to the ISO 9000 series). In both cases there is likely to be some resistance to the idea of change, particularly when a lot of effort and training has gone into learning a particular set of assessment criteria.

Technology can help by automating the switch from one set of criteria to another as much as possible. For example, the U.S. Air Force has developed a formula (which can be implemented on a spreadsheet) for estimating your 1995 Baldrige score from your 1994 score. Some assessment software programs have gone a step further by incorporating an estimation capability that automatically estimates scores on a one assessment from your scores on another related assessment.

For example, you might estimate performance on an ISO 9001-based assessment from scores on 1996 Baldrige. The advantage, of course, is that you are able to leverage past work to help you with new assessment. You needn't start completely from scratch.

However there is an important caveat that must be applied to all such estimation approaches: the accuracy of the estimated results depends critically on the degree of overlap (correlation) between the two assessments in question. Thus, as we saw earlier (in Exhibit 2), it would be foolish to expect an accurate estimate of performance on

Baldrige's Business Results Category based on an ISO 9001 assessment, because ISO 9001 has very little to say about business results.

Provided that assessment software includes estimates only where they are appropriate, the estimation feature can add value and help reduce the feeling of frustration that occurs when an organization switches from one assessment to another. In the case of two assessments which have a high degree of overlap (e.g. 1995 Baldrige and 1996 Baldrige, or 1996 Baldrige and a Baldrige-based state or city quality award) estimation capability may be especially useful.

Finally, it is worth emphasizing that much of the data and supporting comments which are gathered for one assessment, often apply to other assessments as well. For example, preparing for ISO 9001 registration at IBM was much easier because of all the work that had been done previously for the periodic Baldrige assessments. It is healthy to try a variety of assessments, **provided** that you stick with each long enough to actually get results. The spirit of continuous improvement is to never rest content. No assessment is perfectly complete, so if the organization seems to be doing quite well by one set of criteria, perhaps it is time to try another set.

Conclusion

Periodic, team based assessments should be at the heart of any serious quality improvement program. They are an excellent way to gain buy-in across the organization. They help the organization focus its improvement efforts. Perhaps most importantly, they provide the in-process feedback that gives people confidence to stay with the quality improvement journey, even though financial results may be several years away.

Team-based assessment software is a highly cost-effective approach to periodic self-assessment. A typical cost analysis shows it is reasonable to expect at least ten dollars of cost savings in labor for every dollar invested in team-based software. Team-based software can also improve the effectiveness of the assessment. By averaging scores and merging comments from many assessors, team-based software increases the validity of a self-assessment at the same time that it increases involvement and understanding.

Customizable assessment software, particularly if the software has extensive online help and assessment advice, can help team members easily see the relevance to their day-to-day jobs. Finally, the ability to estimate scores on one assessment from scores on a related assessment can leverage assessment effort and ease the transition from using one set of assessment criteria to another.

Computerized team-based assessments are complementary to a large number of existing technologies that have greatly helped past assessment efforts. These include: Email, electronic bulletin boards, the World Wide Web, transferring spreadsheet/database/word-processing files electronically, teleconferencing, and use of Computer Supported Team Workspaces.

Although computer-based team assessments can be a tremendous help in carrying out an assessment program, it is important remember that assessments can be successful only if they are part of long term committed effort to improve quality, as opposed to a “program of the month.”

The cartoonist, Scott Adams, captured the distinction very well in a “Dilbert” comic strip that appeared recently in the journal of the American Management Association. In the strip, Dilbert has been asked by his boss to help the company apply for the “Millard Bullrush Quality Award.” Under duress, Dilbert pops open his laptop computer, teams up with Dogbert, and says: “I have to submit my project for a ‘Quality’ award. I’ll need your help on the dishonest parts.”

Like Dilbert, we can team up and use the latest technology to help with our assessments. However our toughest challenges may be whether we have the courage to be objective and the commitment to stick with quality improvement long enough to reap the rewards.