The use of surveys for Program/Policy or Initiative Evaluation

Jeffrey M. Saltzman
The use of surveys for Program/Policy or Initiative Evaluation

Jeffrey M. Saltzman

There have been a large number of statements made in various textbooks and articles that are along the lines of “if you can’t measure it, you can’t manage it” or “you can’t manage what you don’t measure”. It is a cute statement; one used to justify various management behaviors, and is quoted over and over ad nauseum. The only problem is that it is wrong. There are plenty of things that we don’t measure and yet manage quite well. The statement itself as it is commonly used is actually misquoted, with the original statement being “you can’t control what you can’t measure” taken from a text on software development.

Do you measure your degree of hunger before putting some food in your stomach? Do you measure the length of your hair prior to getting a haircut? Do you measure your pleasure at attending a baseball game, a Broadway show, a concert, or a movie prior to deciding if you should attend another? Do you measure your likes and dislikes of various food choices before deciding what to have for dinner? Do you measure the buildup of plaque on your teeth or the bacterial content of your mouth before deciding if it is time to brush those pearly whites? Do you contrast various aspects of your car’s performance by the gasoline brand, its miles per gallon, its horsepower etc. before deciding which brand of gasoline to fill your tank with? We make choices everyday in our lives that are not rigorously measured and we seem to manage many of those decisions just fine and to exert the appropriate amount of control over them.

Medical doctors use a variety of techniques, among them clinical judgment, in deciding a course of action for our various ills. This is often done without rigorous measurement of the presenting condition, but rather by looking at the panoply of signs and symptoms being displayed. And remarkably people get better every day. Teachers when reviewing essays from their students are using similar judgments in deciding whether a paper deserves an A+ or a C. Artists of various kinds also use judgments when deciding how to proceed on an artistic creation. Those judgments are based on their experiences and training as professionals in their respective fields. The list of non-measured judgments with more often than not positive results could go on and on.

Evaluating what you can’t measure

How do these non-measured judgments get made? And how do they get made in such a way as to lead to fairly consistent positive performance? How do these types of decisions differ from the kind that would benefit from more formal or rigorous evaluations? These decisions are based on rules-of-thumb or heuristics. Experience, training, logic, folk-wisdom, categorization including bias and bigotry are all used to generate rules that allow people to make decisions with less than complete knowledge of the situation. You can decide if you need a haircut without actually measuring the length of your hair by operating with certain rules-of-thumb. Maybe your trigger
point for a haircut is calendar based – every 4-5 weeks. Or maybe it is when the hair on the sides of the head touches your ears or for some when there is any sign of stubble on an otherwise gleaming sphere. Some simply look in the mirror and “know” that it is time. Whatever the rule-of-thumb is, you can apply it to the decision point regarding the need for a haircut without actually measuring the length of your hair.

Managers in organizations apply rules-of-thumb or heuristics daily in their jobs. There are oftentimes rule books and procedure manuals that can be referred to, formal information flows that can be queried, but with the large number and variety of decisions to be made including many times regarding the unforeseen, rules-of-thumb are often utilized. Some managers have a good set of rules-of-thumb that tend to yield positive results and others with poor sets or those that apply good rules inconsistently are more or less flying blind.

Value statements developed by organizations (and often hung on a wall), if actually utilized, can help frame-up the paradigm that should be used to determine or help guide decision making. For instance Google lists 10 things that they “have found to be true” including “#6. You can make money without doing evil.” Merck lists 5 values on their website including “#2. We are committed to the highest standards of ethics and integrity.” Citigroup talks about 3 shared responsibilities with the overall goal of being “the most respected global financial services company.” Many corporations have these Value statements and they represent core priorities of the culture, the overarching rules by which people would be expected to make decisions within the organization.

The downside of these rules-of-thumb is when stereotypical concepts or bias are the basis for their development and then those bias laden heuristics are employed when making decisions. Without having actual information, or in an attempt to quickly sort through or condense a large amount of information available, people have a tendency to base decisions on preconceived or stereotypical notions and sometimes those notions could be called erroneous at best, despicable at worst. Rules-of-thumb or heuristics though are not inherently bad and do not inherently lead to poor decisions, but when erroneous, bias, bigoted or just plain stupid rules-of-thumb are adopted poor decisions will follow.

One application of these concepts comes about when voters are deciding whom to vote for in an election. An election can be thought of as a type of program evaluation. Each candidate develops a program (a campaign strategy) designed to get them elected. The candidate with the best program, the one that resonates most with the voters, receives a very clear evaluation, they get more votes and hence assume office (unless of course if you are in Zimbabwe or Florida). For the voter responsible for evaluating the candidate’s program two possible paths can be taken when deciding where to cast their ballot. One path, a rational approach, is to analyze the candidates on each of their positions and then to select the candidate that most closely matches your own views. The second possible path is to employ heuristics, general rules-of-thumb to allow you to select your candidate of choice without having to analyze each of their positions. It has been found that in mock elections, that voters who try to collect and analyze as much information about each and every candidate as they can make poorer decisions about whom to support. In other words they choose candidates who do not necessarily come closest to mirroring their own
positions on issues. Those who use rules-of-thumb made better decisions. The rules-of-thumb employed included party affiliation, endorsements, position in political polls, even physical appearance. One conclusion of this work was that a typical voter uses party affiliation similarly to how consumers use brand loyalty as a short cut in the decision making process on which products and services to purchase.

You can manage what you are not measuring through various heuristics; the question is when does measurement add appreciably to your ability to make better judgments? There are times when rigorous measurement of programs, processes or choices can greatly aid in the decision making process and surveys can be used in that evaluation. For instance, in organizations when the amount of resources that can be brought to bear are limited and difficult decisions need to be made about the deployment of those resources, a measurement process that gives insight into the expected benefits of action A vs. action B can be especially helpful. Consider the following situations:

- A school system wants to know if the investment it is making in advanced teacher training is improving educational attainment among its students

- The GAO wants to know if efforts in information dissemination by government agencies can be documented as furthering government agency goals

- A corporation wants to know if changes made in a reorganization are achieving the desired impact in terms of increasing organizational effectiveness and customer satisfaction

- A zoo wants to know if it should proceed with increasing elementary school outreach by investing in an animal travel program, developing an on-site hands-on children’s zoo, or investing in bringing in more exotic animals and building new exhibition space in order to maximize attendance

- A retailer wants to know if it should invest in more physical retail locations, beef up its website or send out additional catalogs with its limited budget dollars

- A benefits departments wants to know how the company’s redesigned benefits package is being perceived by employees and how well the execution of the benefits program is being conducted by the new outsourced provider

- A refiner wants to know if it should invest in a new Greenfield plant (a tangible asset) or if there would be greater ROI if it invested the equivalent dollars in additional training for all of its workers that would be expected to increase efficiency and throughput in its existing manufacturing assets (an intangible asset)

- A sales department wants to measure its sales funnel of potential deals, to determine if its marketing and sales efforts are having the payback it desires and to predict future sales volumes for the organization.

All of these decisions or potential decisions can be enhanced by an effective effort aimed at program evaluation. Clearly having effective measures is invaluable to those trying to manage these kinds of decisions. Where to invest, how much to invest, when to invest are critical decisions and managers would be well served to seek out systematic measures to enhance their judgment. The
American Evaluation Association defines evaluation as “assessing the strengths and weaknesses of programs, policies, personnel, products, and organizations to improve their effectiveness.” While organizations vary in their efforts at program evaluation one interesting statistic comes from NASA-Goddard which has determined that for them between 7% and 9% of a total program’s budget should be spent on evaluation, and a recent Air Force initiative devoted 15% of the project budget to a rigorous evaluation of whether their new inventory control system was working as planned. Program evaluation though has been around long enough and has been practiced by a wide enough variety of people that not only have myths sprung up about program evaluation but other documents that attempt to debunk the myths of program evaluation are also in existence.

Traditionally when organizations have thought of program, policy or initiative evaluation what springs to mind is a list somewhat like the following as areas to be evaluated:

2. Views of Pay & Bonuses
3. Health Safety and Environment (HSE) Emphasis and Policy,
4. Effectiveness of Diversity Programs,
5. Work Life Balance Initiatives,
6. Psychological Recognition Efforts,
7. Development & Training Opportunities,
8. Advancement Systems,
9. Physical Conditions,

The traditional definition should include a broader array of various types of programs, policies and initiatives that an organization may undertake and would benefit from a rigorous evaluation methodology.

**Method Impact**

Programs can be delivered through various methodologies, outreach, a computer interface, a classroom-type experience, a hands-on experience, reference books and guides, telephone and other methods. The same content with differing delivery mechanism may not have the same impact. Method impact represents a potential confound in determining the impact of a program. You might have great content which is simply being delivered with an inappropriate mechanism.

**Field Experiments**

“I got this body while eating pizza, hamburgers”, and she then leans forward whispering into the camera “even chocolate!” Before and after pictures are displayed on the TV screen. While the shapely product promoter is telling you how she lost all that weight and became irresistible by eating the foods she is hawking there is a line in small letters on the right-hand bottom corner of the screen, “Results not typical”. I wonder what result was not typical. That people who use this product end up on TV promoting it? Or that people who use the product actually lose
weight? Or that people who eat the advertised food actually think it is any good? Maybe “Results not typical” refers to all three. One has to wonder given the schlock nature of the ad just what warning their fine print is conveying.

Of course, sadly, the ad would not be running, and it has been around awhile, if it did not work in attracting people to use the product. We should all be aware of the weakness of the case study approach as well as approaches that claim success without appropriate experimental design. How would you evaluate the above statement if you treated her claim “eat this food and lose weight” as an organizational program or initiative and wanted to determine if in fact you could place any stock in those claims?

Program evaluation suffers a history of skepticism often due a history of poorly conceived evaluation methodologies. One of the most widely used designs for program evaluation is one in which 1). a single group is given a baseline measure, then 2.) the program is implemented and then 3.) a post implementation measurement is made to determine the effect of the program. This approach is fraught with problems. Let me illustrate. Let’s use one of the examples listed from above:

• A school system wants to know if the investment it is making in advanced teacher training is improving educational attainment among its students.

In this example a baseline measure regarding standardized student achievement test scores is collected prior to the implementation of the new teacher training program. In addition surveys can be done of the students asking about their comfort and mastery with the subjects covered by the teacher training program. After the teachers receive their training and the next class of students comes through, the measurement process gets repeated. Students on average now report that they feel more comfortable with the targeted subjects and test score in fact are moving up. The teacher training program is determined to be a success and additional funds are poured into teacher training. What was not taken into consideration using this approach is the fact that the students got new textbooks with vastly improved course material and the class schedule was redesigned so that the students spent more time on the targeted subjects. What was thought to be an outcome of teacher training was actually better course material and more time spent devoted to the subject.

There are various approaches that could enhance the ability of the school system to improve its evaluation. Here is one. If the above field experiment had been carried out as follows the results would have been much clearer. In this alternative approach the same setup is used, but only half the teachers in the first round receive the training. When the next round of students come through the program the students whose teachers received the training are compared to those whose teachers did not, holding everything else constant. Holding everything else constant in this case means that all students received the new texts and all students had their schedules changed. Now when we compare the student whose teachers received the training to those who did not, we might find that all student scores improved but the students whose teachers received the advanced training had even higher scores and felt even more comfortable regarding their mastery of the subject than those students whose teachers did not. In this case a more confident determination can be made that the
advanced teacher training did help to improve test scores.

The criticism of this approach is that it is not right to withhold a potentially beneficial experience from those in the control group. And the answer to that is, in certain circumstance it is not right to withhold the potential benefit (experimental drugs being used in life threatening circumstances is one case). However in the above example it is truly unclear whether there is any benefit to the students whose teachers went through the advanced training vs. those who did not. And there certainly would be a long-term benefit to the school system by knowing the true impact of that training experience.

**Types of Evaluation**

Compounding the poor track record noted above, many program evaluations have been sneered at as wasteful efforts due to another factor and that is whether a program gets continued funding may be somewhat contingent on issues outside of the effectiveness or the evaluation of the program. Common problems facing program evaluation consist of unclear program goals (so how do you know if you have accomplished the goal?), difficulty in collecting data, a lack of cooperation among those responsible for implementation (those whose plates are already full) and the politics of and vested interests surrounding various programs. Yet with a methodical approach insights can in fact be gained without too much pain into whether the program is having the hoped for impact.

Program evaluation can take place within a specific window in time or it can be an ongoing activity. For instance, an organization might implement a phased-in approach to a new program, or a new product/service etc. and decide that after 6 months the program will be evaluated to determine its impact, and whether funding should be continued at the current level, increased or diminished. If there is a positive finding to the evaluation effort, the program would be rolled out to the rest of the organization. Other organizations may not use a phased in approach and may simply have two or more programs being performed simultaneously in the organization and then conclude possibly that it cannot afford all of them, so which ones are benefiting the organization more? An evaluation might be done once, at a single point in time to help to make that decision. Other organizations might decide that long-duration type programs, ongoing efforts, need to be constantly or periodically monitored to ensure that they are still delivering the desired for outcomes or that processes associated with them are not deteriorating in their effectiveness.

Surveys can be used for both single points in time and on-going long duration reviews as well as to measure program processes, content, and outcomes and there are appropriate uses to all in program evaluation. Some of the questions that program evaluation efforts can be used to answer include:

**Process**
- Is the program being carried out well, and how can it be improved?
- How can you measure the impact of various programs on organizational processes?

**Content**
- Are the people targeted by the program understanding, retaining and actually utilizing the content being provided?
- Is the content covering all critical areas?
Outcomes

• Will the program have an impact on organization effectiveness, increasing the organizations ability to perform?

• Which programs are the most cost effective, providing the biggest bang for the buck?

• How can I measure the return on investment (ROI) of various programs?

What types of program evaluation can surveys assist with?

<table>
<thead>
<tr>
<th>Evaluation Type</th>
<th>Single Point in Time</th>
<th>Long Duration Processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Evaluation</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Content Evaluation</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Outcome Evaluation</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Process

The evaluation of the processes of a program, how the program is being executed, can range from being simple (sometimes simpleminded) to the information being quite difficult to obtain and providing very meaty insights. Sometimes unfortunately simple process questions are all that is used to evaluate a complex program and conclusions drawn about processes are used to evaluate the entire effort.

Two very common forms (and often the simplest) are when trainees attend a class or when participants attend some kind of conference. Typically a form is passed out at the end of the class or the conference asking questions about the facility, the quality of the food, the enthusiasm of the speakers, the room itself or the sound systems etc. If you want people not to feel at all pressured into giving a positive response, because the instructor is collecting their forms, and participants may feel that the instructor is able to tell who is responding which way, the form can be sent to the participant a few days after the meeting or the survey can be done via the web.

Sample process type questions for a class or conference

How would you rate, on a scale of “Very Good” to “Very Poor”, each of the following?

1. Ease of enrolling in the class?
2. Ease of getting to the location of the class?
3. Amount of parking available?
4. Class starting at the expected time?
5. Snacks provided during the breaks?
6. Being able to hear the instructor?
7. Being able to see the charts and graphs shown?
8. The enthusiasm that the instructor displayed when presenting the material?
9. Overall the effectiveness of the instructor in presenting the material?
10. Start time of the class (Scale: “Much Too Early”, “Too Early”, “Just Right”, “Too Late”, “Much Too Late”)?
11. Number of breaks during the class? (Scale: “Too Few”, “Just Right”, “Too Many”)?
12. Temperature in the room? (Scale: “Too Cold”, “Just Right”, “Too Hot”)?

Process evaluation becomes much more important when it is about critical business processes or the evaluation of a redesigned business process. If we think of what the future might hold for some organizations regarding how this might be done, one possibility is to have a survey embedded and to become a critical part of a business process providing feedback in a real-time manner.

**Sample Process Evaluations**

A supplier of food products to a restaurant chain may ask each restaurant manager to take a few moments and respond on a website regarding today’s shipment. This on-going evaluation would be an example of a long duration process evaluation. Did the shipment contain the correct products in the right quantities? Did the products arrive in good condition? Were they fresh? Was the invoicing that accompanied the products correct? Any response to the survey outside of an acceptable limit can trigger an immediate follow-up response by the supplier’s sales representative who can take care of and resolve issues for a customer having an unhappy experience. Building the survey into the business process itself, potentially the ultimate in program evaluation, can result in greatly increased levels of customer satisfaction.

A customer at a web enabled self-service gas station pump can be asked to respond to a very short series of questions while pumping gas. The list of questions would vary each time the pump was used so that while each customer would only be asked to respond to 3-4 questions a larger overall picture can emerge. For instance question sets could be built around:

- Perceived quality of the product
- Perceived value
- Physical condition of the station (cleanliness, lighting etc.)
- Environmental programs
- Other image/reputation related questions
- Location
- Staff
- Convenience store products
- Any special programs or initiatives underway.

Each of these questions sets would share one final outcome question such as, “what is the likelihood of you repurchasing the product at this station that you are purchasing today”. By having this one item shared across all the question sets, it is possible to combine the question sets, using that common item as the link, so as to derive which results and programs and at what levels will have the greatest impact on repurchase intentions (the ultimate goal of initiating programs within a retail environment). By building this system with a real-time interface the store manager can receive immediate information about customer perceptions that can influence the sales at the store. For instance, if the store normally receives a score of 75% favorable on cleanliness, and the score today drops to 60%, the manager should go check conditions. Meanwhile the store data can be rolled up so that regional managers at the chain can see real-time how customers in aggregate are viewing the products and services delivered in their stores and as needed take corrective action.

Say an HR department overhauls its offering and instills a campaign of customer service. By conducting a pre-change baseline survey
and a post-change follow-up the department would begin to gain some insight into the effects of the changes it is enacting. This insight would be greatly enhanced if the changes were to be phased in. With a phased in approach you are essentially providing a control group so that at the same points in time you can contrast the responses coming from people under the new systems or programs against those who are still experiencing the old methods. Here is a short questionnaire covering process type questions that if asked of its customers, might provide HR some insight in terms of how it is doing as it undertakes the change process.

On a scale of “Very Good” to “Very Poor” and with an escape “Does Not Apply”:

**Payroll**
1. The accuracy of your electronic payroll deposit (correct amount, deductions correct etc.)
2. Receiving your pay on time – that is, when it was supposed to arrive
3. Ease with which you get accurate and timely answers to payroll questions

**Compensation** – For managers who make salary decisions
4. Ease of access to the information and tools you need to make salary review decisions
5. The accuracy of decision implementation

**Benefits**
6. Ease of getting information you need to make decisions about which benefits you should select for you and/or your family
7. Ease with which you can make changes after making your initial benefit choices
8. The accuracy with which your changes were implemented

**Resourcing / Recruiting / Job Posting**
9. As an applicant, the ease with which you can use the company’s on-line job posting system

**Managers**
10. Ease of the process to fill the job openings
11. Effectiveness of the process used to fill the job (e.g. timeliness, candidate volume and quality)

**Employee Records/Information**
12. Ease in getting access to or updating your personal employee information (e.g. address, marital status, job and pay history)
13. (Managers) Ease in obtaining job-related information on employees within your team

**Performance Review**
14. Effectiveness of the Performance Review Process (e.g. feedback coaching, the scheduling and tracking of performance reviews/appraisal process)

**Training**
15. Ease in getting information about available training courses
16. Ease of enrolling for a course

**Personal Development**
17. Access to tools that help you in deriving your personal development plans
18. Effectiveness of tools available to help you derive your personal development plan

**Severance**
19. (Managers) The support you receive during a severance process (e.g. delivery of terms and conditions, awareness training, support)

**Relocation Administration**
20. Being able to obtain information needed about relocation process (including expatriate assignments)
21. Effectiveness of the overall relocation package
22. Resolution of any issues regarding relocation (e.g. visa, bank accounts, work permits, home disposition, shipping of household goods)

**Compliance**
23. Ability to obtain information regarding compliance with employment laws, regulations and company policies

**Outcomes**
24. How would you rate your overall satisfaction with the products and services that you are provided from the HR department
25. Overall have the quality of the products and services provided by the HR department since the implementation of the “new XYZ” program has (1. Greatly Improved, 2. Improved, 3. Neither Improved nor Declined, 4. Declined, 5. Greatly Declined)

**Program Content**
Surveys can also be used to evaluate the content of programs and initiatives. Different type of programs can have radically differing types of content. Evaluating the content of an educational or informational program might be the most straightforward. Did the material cover the knowledge base which it was supposed to? In general if you think of programs as potentially impacting a participant’s knowledge, skills or abilities, content evaluation of various types of programs can proceed as an evaluation of whether the program has the content necessary to alter the participant’s knowledge, skills, ability or any combination of the three, enhancing their capabilities. For instance, if the sales representative of the food supplier who receives information from the restaurant manager on the status of today’s shipment were to find that the new program was supplying information that enabled him/her to perform their job better, and increased their knowledge of customer perceptions and ability to react to dissatisfied customers they would deem the program a success.

The criteria used to evaluate the content of a program are extremely dependent on the audience for whom the content is targeted, but in general the issues to be considered when evaluating the content of a program include:
Relevance – is the content important to achieving the programs goals;

Breadth – does the content cover the material as thoroughly as it should;

Appropriateness – is the content conveyed in such a manner that it will be understood and accepted by those who are to receive it;

Quality - is the content information accurate and clearly presented;

Impact – did the content provide information in such a way that it will change/improve the outcomes deemed important (e.g. change in behavior, change in understanding) immediately and for longer durations.

The above issues regarding content can be measured by asking those impacted by the program to complete surveys and to rate their reactions to a series of statements that flesh out the dimensions listed above. This can be done immediately after the individual comes into contact with the program, treating the content evaluation as a transactional evaluation or it can be done after a period of time passes to determine if the content has staying power with the individual and if it translates to longer term impacts and on-the-job performance. A combination of an immediate reaction review as well as a longer duration review can give additional insight into the effectiveness of the content.

Outcomes

When evaluating the outcomes of a program one issue to be considered in the evaluation is the time duration of the impact will you be evaluating. Impacts of programs range from the short- to long-term and a question surrounding your evaluation is the meaningfulness of measuring one or the other or a combination of short, medium and long-term impacts. From a short-term perspective, did the audience consider the information delivered to be credible and was it delivered in an acceptable fashion? Where there immediate changes in the audience’s knowledge, attitudes or behavioral intentions? On an intermediate scale, did the expected outcomes transfer into other environments, did behavior change, or did organizational performance improve? And long-term did the desired impacts stick?

Knowledge assessment is another way to evaluate the outcome of a program. If the program was to impart specific knowledge to the individual, whether it be knowledge regarding an academic course topic, how to run a cash register, what the definition of acceptable customer service is, the risks associated the XYZ, how to find information on relocation assistance, or how to enforce the law etc. you can evaluate the participants knowledge and mastery of the facts and processes that they should now be familiar with. While individual performance will vary depending on ability and skill, a determination can be made regarding the group as a whole achieving an acceptable level of comprehension on the content that was imparted.

In one employee attitude survey conducted on a law enforcement group 6 months after they attended a training class on statute enforcement, in addition to the typical questions surrounding perceptions of the work environment and organizational effectiveness, embedded in the survey was a series of knowledge questions regarding the statutes that this particular group was to enforce. A typical survey question has no right or wrong answers it is simply a response to how the employee feels about certain issues, however these knowledge questions did in fact have a right and wrong answer and about a third of this law enforcement population could not correctly
answer questions regarding the statues they were to enforce. Six months after the program on how to enforce the statutes the correct information was rapidly fading from memory.

Overview of Steps for using Surveys in Conducting a Program/Policy or Initiative Evaluation

Planning:
1. Review organizational mission, vision, values, products, customers etc.
2. Review background material on specific program, policy or initiative for goals and objectives,
3. Define clear goals of what would constitute a successful program,
   - Programs must have clearly specified goals and objectives before an evaluation can take place. A program goal is a broad statement of what the program hopes to accomplish or what changes it expects to produce,
4. Define metrics that will determine if goals have been achieved,
5. Determine if a single point in time or long-duration evaluation is appropriate, if single point at which point will data be collected,
6. Establish definition of population eligible for evaluation,
7. Determine if census or sampling approach will be used,
8. Determine best approach for reaching that population, (e.g. paper survey, web, IVR),
9. Define project logistics, major milestones, timeline, costs,
10. Define challenges that might arise preventing a successful evaluation and prepare contingencies to address those that arise.

Orientation:
1. Orient those whose help you need for successful evaluation, explain their roles and what help you need, timelines, methodology etc.

Interviews:
1. Conduct focus groups and/or interviews of key stakeholders, explain what is going on, listen to their concerns, gather input for defining questions to be used as metrics, (in addition to content this step is intend to help achieve buy-in to process),
2. Review background material on program, policy or initiative for potential survey content.

Creating the survey
1. Based on population definition determine needed demographics (e.g. gender, ethnicity, tenure, location, occupation),
2. Based on orientations, background material, focus groups and defined critical metrics craft the survey questionnaire.

Data Collection/ Fielding of the Survey
1. Begin rollout of communications plan
2. If sampling create sampling plan and tie to actual program participants
3. Enable as many people to complete the survey in as easy a fashion as possible. Consider multiple administration methodologies, language translations, extended administration windows etc.
Analysis of the results

1. Analysis should consist of various views of the results; a combination of the following often leads to robust conclusions. Enough time should be allotted for this step:

- Descriptive – what are the results overall, by each demographic category, over time (if available)
- Inferential – what seems to be causing what
- Forward looking – by an examination of variance and though the use of simulation, what would likely happen if certain parameters were to change
- Linkage to outcome metrics – demonstrating that those parts of the organization that scored more favorably on the program achieved higher levels of performance
- Conclusions and Recommendations

Write up of finding and recommendations

On-site Presentation of Findings

Follow-up training and assistance for implementation of findings

Cautious embrace of variance

Within the world of Six Sigma there is a paradox that is articulated as follows: to attain Six Sigma performance we must minimize process variability, slack and redundancy by building variability, slack and redundancy into our organizations. In other words, in order to constantly improve performance (the goal of Six-Sigma), room must be made in the organization for the investigation, the vetting of alternative methods and procedures.

Wait a minute, today’s organizations are lean and mean. You have to constantly do more with less. Organizations are downsized, resulting in fewer people but the amount of work required does not often get adjusted. You can’t be efficient and get all this work done if you build in extra resources in order to test new procedures and methods. But you must. Long term process improvement and organizational success are dependent upon it.

With the goal of improving the organizational performance in mind, the path does not begin with the concept that what we need to do is to stamp out all variability, to make everything conform to a certain standard. Variability needs to be understood (through rigorous measurement), it needs to be controlled (in order to minimize defects or errors), but removing all variability eliminates the ability of the organization to learn from itself and eliminates opportunities to improve. Variability is needed in order for improvement to occur.

Here is an example to illustrate that point. If you are examining 200 departments within an organization, and there is no variation whatsoever, you cannot learn much. You cannot learn that in these 15 departments that do “A”, “A” leads to more positive outcomes such as lower levels of employee turnover compared to the 15 departments that do “B”. In this case “A” and “B” can be the same thing but at different levels. For example “A” might be a high level of employee engagement, whereas “B” might be a low level of engagement. By examining these differences the organization can learn, it can improve. Without variation you
cannot learn, you cannot determine that “A” leads to one outcome and alternative “B” to another.

In addition to the concept of variability, add the concept of redundancy. Redundancy is needed in order to allow experimentation. Redundancy is when two different approaches are available to achieve a desired outcome. For instance using a stamping machine to form a part from a roll of sheet metal, or using powdered metal to form the same part in a mold under pressure. Which procedure is better, leading to lower costs, less waste, fewer defects, and better part performance?

By being able to breathe through the nose and mouth, a redundancy, nature was freed up to experiment with noses for other potential uses – such as the elephant’s trunk. If the elephant could not breathe through its mouth it would be in a precarious position if it filled it’s trunk with water and then needed a breath. By having people perform a task using more than one method it is possible to determine if one method is more advantageous than another. Upon standardizing around the more advantageous method, you immediately begin investigating other methods (in a controlled fashion) to see if those new methods yield even more improvement. This requires slack, the controlled embrace of variability and redundancy.

**When programs cannot be allowed to fail**

Sometimes programs or initiatives are mission critical to the organization or involve safety or potentially issues such as national security and they simply cannot be allowed to fail. The goal of a program, policy or initiative evaluation effort in these cases is not to determine if the program is not working once implemented, but rather to evaluate at critical milestones whether the program is working “up-to-that-point”, a phased approach to implementation and evaluation. If the answer is affirmative then the program can proceed onto its next phase and eventually reach a successful total implementation. If the answer though is negative, the program will repeat the current phase until it is done at a level deemed acceptable. Successful performance at each stage is a requisite condition for moving onto the next phase of the project.

Program evaluation is not all that difficult. But what is required to successful evaluate a program is a scientifically based rigorous approach that if followed will lead to better, more impactful decisions and a better bottom-line for organizations.
References


