



Hard Measurements for Soft Science: Behavior-Based Safety Has Evolved

Author – Terry L. Mathis
Founder & CEO – ProAct Safety

In the early 1980s many safety professionals were excited about the possibilities of using new advances in the behavioral sciences to improve organizational safety. Among the technologies being investigated was the idea of behavioral observation. Behavior is by definition “an observable act” and therefore measurable by workplace observation. If a statistically-significant connection could be made between certain behaviors and accident probabilities, measuring these behaviors through observation might provide a more accurate measurement of workplace safety.

One reason this technology was viewed to have potential applications was because traditional safety metrics had been almost totally comprised of lagging indicators, i.e., accident investigation data. For many years organizations had measured safety by their failure rates. These practices in lagging-indicator measurement led to reactive management practices in which organizations did little to eliminate risks until those risks actually turned into a detectable pattern of workplace injuries. Workers identifying risks would sarcastically ask “who wants to get injured on this so we can get it fixed?” Organizations reacted to tragedies, but failed to prevent them.

Not only was accident data a lagging indicator, it was reported data. Employees, including those injured and those who witnessed the injury, were required to report accidents; and reported data was problematic. First, not all accidents were reported, and secondly, the reports were not always accurate. Organizations began to realize all the influences that impacted reporting and non-reporting of accidents. Even well-intentioned programs such as safety bonuses could be a significant factor in discouraging the reporting of accidents by workers who could cost their fellow workers a bonus by reporting their injury. Also, injured workers and witnesses often were unclear on details or distorted the facts to avoid blame. Reported data was potentially both incomplete and inaccurate. Behavioral observations could solve both of these problems by adding new, proactive metrics gathered through observational sampling to supplement lagging indicators gathered through employee reporting.

The problem with this new technology was that it was largely subjective. Observers were asked to make a value judgment on the relative safety of the behaviors they observed. Some systems asked the observer to “rate” the safety on a scale of 1-10 while others called for a dichotomous rating such as “safe” or “unsafe” which would be scored in separate columns. Such evaluations would vary wildly among observers and early attempts to norm the evaluations met with limited success. About the time these problems began to dampen the enthusiasm for the new technologies, two new approaches were developed to address the problem of subjectivity: one was called “pinpointing” and the other was called “operational definition.”

Pinpointing simply broke a complex behavior down to single movements which could be observed in isolation. Rather than asking an observer how safely a butcher was cutting meat, it would ask about a single action such as “cutting away from (rather than toward) your other hand”. The observer would consider the worker in compliance if he or she were doing this pinpointed behavior and would consider them not in compliance if they were not. Obviously, this technique required someone to make another judgment on which behavioral pinpoints were important for observers to focus on. Many checklists for observers were still compiled from the perceptions of those involved about which behaviors were most important in preventing accidents.



Hard Measurements for Soft Science: Behavior-Based Safety Has Evolved

Operational definition allowed for a more complex behavior to be observed as a whole by defining which multiple criteria of the behavior were required to rate the behavior as safe. An operational definition might have several parts such as "is the worker using an approved cutting device; is the device free of blade chips; is the worker cutting away from his or her other hand?" If the answer to every section of this definition is "yes," the worker is considered in compliance. If the answer to any section is "no," the worker is marked in non-compliance. Operational definition, while allowing for more complex observations, also requires more knowledge on the part of the observer. This means more training and more time to observe depending on the number of items being observed. Like pinpointing, operational definition raised the question of whether or not the items being observed were the most important.

The solution to this issue with both approaches was found in a simple application of Pareto Analysis. Using a worksheet of the most common pinpointed or operationally defined behaviors, an organization can Pareto Analyze their accident and near-miss data to determine which behaviors have the most significance in potential accident reduction based on historical data. Obviously, past accidents may not be a perfect indicator of future accidents, but in most organizations it is significantly predictive. Using a checklist for observations based on the significant few from the Pareto Analysis and allowing observers to "write in" non-checklist behaviors that are observed proved to be a formula for determining the most statistically significant behaviors and for continuous adjustments of the checklist as behavioral issues progress and change. The first 100 checklists developed using this methodology had an 88% or higher correlation to ongoing accidents for three years.

Early Pareto worksheets were crude and often did not include important behaviors to be considered in accident data analysis. As more and more sites and industries took this approach to checklist development, the list of behaviors grew and several studies in the old "super computers" of the day provided a much more comprehensive list of the behaviors most often critical to accident reduction. As the list grew, practitioners began to divide the behaviors specifically by industry and workplace type. The statistical significance of these lists was verified by the increased accident reduction produced in observation processes using the improved worksheets for checklist development. As variations in terminology started to become the norm, the practitioners using the worksheets were able to downsize the number of behaviors to more generic and common models. The most commonly used worksheets in Behavior-Based Safety now have less than 40 generic behaviors on them. Specific checklists developed from these worksheets seldom need to include non-worksheet items other than specific, procedurally-oriented behaviors connected to machinery or processes specific to the industry or site.

Early checklists often utilized every behavior identified in the Pareto analysis, making them average between 15 and 28 behaviors. Observations using these checklists were lengthy and required extensive, ongoing training for observers. It was not uncommon for observations to take 30-45 minutes each. A lot of work was also done to norm the observers by pairing them up and using observation data to do comparative analysis of the variation between observers. Another problem arose with the old checklists: workers did not internalize the behaviors and became dependent on the observations. When the number of observations was diminished in year two or three, the frequency of the behaviors diminished proportionately and accident rates increased to near pre-observation levels. Theoretically, a year or two of observations should have installed some stability of the behaviors which would perpetuate without continued observations. This did not hold true. Analysis of the problem eventually revealed that the long checklists overloaded the workers causing the behaviors not to become habitual, but dependent on the reminders from observers.

This problem was eventually solved by shortening the checklists to the significant few (usually 6 or fewer) behaviors. The shortened checklists facilitated quicker internalization of the behaviors. Workers could name the behaviors after 5-8 observations and the behaviors became habitual after 12-14 observations. The shortened checklists also made it possible to shorten observation times to 5 minutes from the original 30-45 minutes and also to shorten observer training. Many sites focused on the significant few for the first 18 months and then begin to evolve the checklist behaviors further down the Pareto list.



Hard Measurements for Soft Science: Behavior-Based Safety Has Evolved

It became obvious after the first 2-3 years that items removed from checklists could begin to become recurrent problems. Common practice now is to move items to the bottom of the checklist or create a “new items” list and an “old items” list for observers. Growing checklists to more items after starting with an initial 5-6 item list has not proven to cause the kinds of problems the old, longer lists caused. Apparently, after workers have internalized a handful of behaviors, they are able to focus on new ones while maintaining the old ones. Adding to the list after the original behaviors are addressed does not cause the same level of overload that was caused by the longer lists.

As sites began to develop software applications to analyze observation data, several methods were used to report the data. Over time, the prevailing metric became simply Percent Safe. This metric is calculated by dividing each observation into a dichotomous measurement of safe or not safe. The total number of safes is divided by the total number of observations to determine Percent Safe. “Safe” is common terminology but the other category is still referred to as “not safe”, “unsafe”, “at-risk” and “concerned” depending on the consultant or method used. Most software applications compile Percent Safe the same way although some allow for multiple safes and not safes while others either have one safe or multiple not safes and will not allow for multiple safes on the same behavior on the same observation. In addition to the Percent Safe, most software applications also measure the sample size and the raw number of “not safes” to help those analyzing the data to determine the statistical significance of the metric and the magnitude of the risks measured.

If a worker works 40 hours per week times 4 weeks per month, the total hours worked equals 160 hours or 9,600 minutes per month. If this worker is observed 5 minutes per month, the sample size is .00052083. The number of “not safes” observed can be divided by this number to determine the actual number of risks taken in the behavioral category during the month. As data accumulates over months and years, it can be compared to accident data to determine the actual probability that a “not safe” behavior will turn into an accident and the severity can also be determined as a probability. This data can eventually build the site’s accident pyramid (according to E.G. Heinrich’s strategy) and determine the overall probability of accidents per risk taken. Obviously, the more data, the quicker the actual probabilities can be statistically determined for the site.

Other sampling strategies have also found their way into BBS methods. Sites develop observation strategies using sampling techniques, and software verifies sampling through analyzing the distribution of observations over locations, shifts, tasks and other variables. Observers are often assigned to observe workgroups, tasks, specific workers, upset times, or other criteria and the overall assignments are designed to sample a representative group. Such strategies vary in specificity and method, but often result in a statistically significant sampling of the workforce on a monthly basis which allows both causal analysis and trending over time. Trends have proven to be one of the most significant metrics for determining observational effectiveness. When the strategy for observations results in increasing Percent Safe and corresponding decreasing of Total Recordable Rate (Total Recordable Rate is a ratio defined by OSHA as the number accidents meeting OSHA’s criteria for reporting per 200,000 hours worked), that is a good indicator that Behavior-Based Safety (BBS) is having the desired effect on the site culture. A study of 500 sites implementing BBS showed that the increase in Percent Safe averaged 12% the first 12 months and the corresponding decrease in Total Recordable Rate averaged 37.5%.

An ongoing issue with observation data is that observers are tampering with the measurement while they are taking it. That is, the observers are announcing the observations or, in some cases, asking permission to observe. This practice almost undoubtedly compromises the metric, but has been found to enhance the impact on behavioral change. When observations are taken candidly, they almost always create mistrust and resistance. Observations that are open and friendly actually speed the behavioral change even though they artificially increase the Percent Safe. Those analyzing observational data should realize that the Percent Safe is actually what workers are capable of when they know they are being watched vs. what they usually do when they are not being observed. Although this is a serious breach of measurement protocol, it has proven to be a key element in promoting behavioral change at the cultural level and is common practice among BBS practitioners.



Hard Measurements for Soft Science: Behavior-Based Safety Has Evolved

Almost any good data base can report observation data and several have been customized to meet the needs of a BBS process. The examples used in the illustrations are from a firm called NuDatum Software who built their product on a Microsoft Access database platform. This software can provide the big picture of the observation data in an "Overview" report. This report provides Percent Safe, number of safes and unsafes, and sample size, and can be reported as a table (ref 1) or a graph (ref 2). The software will also trend data over a period of up to 12 months, will report observer activity by observer as well as observer team, and will report observer comments sorted by behavior, as well as additional comments (not related to checklist behaviors) from observers or workers. All these reports can be run for all data or filtered by variables such as time, location, day of the week, and other programmable variables which usually vary from site to site. These reports can be queried by multiple variables and comment fields can be searched for key words or phrases.

BBS Steering Teams are taught to regularly report and analyze this data to develop action plans for improving safety. The percent unsafe data represents potential accidents, and action plans target specific behaviors and influences on behavior which contribute to these risks. A combination of Percent Safe and original Pareto analysis help to determine the risks most likely to produce accidents. These high impact risks are prioritized according to impact, but also according to ease of solution. Steering Teams are taught to aim for high impact but to not miss opportunities for quick wins. The quality of action plans is often directly proportional to the quality of comment data. The numbers help to prioritize issues, but the comments provide the profound knowledge and insight needed to solve the problems. A format often used for these comments on unsafe behaviors is the "what/why" format. Observers marking unsafes are instructed to discuss with the workers observed the rationale for taking risks. So a comment would state briefly the risk followed by the reason the worker gave for taking the risk.

Action plans developed from observation data fall into two categories: action plans to improve safety and action plans to improve the observation data. Action plans to improve safety are aimed at the influences on risk taking. Some of these influences are simply worker perceptions or habits which can be impacted by information or reminders and by the focus and frequency of the observations themselves. Others, however, are often conditional or organizational in nature. Conditional influences can include workstation design, the location of tools and equipment in relation to the tasks, and the availability of needed supplies, among others. Organizational influences can include such issues as training, availability of supervisors, and the clarity and focus of safety communication. Steering teams can address some of these issues themselves. Other issues require help from management or engineering. Such help is requested with attached documentation of why the Steering Team views this as a critical issue. Obviously, good trust and cooperation between the Steering Team and other organizational levels is critical to success.

Observation data can also reveal its own problems. Observation frequency and number are critical to ensuring an adequate sample to give the data statistical significance. The distribution of the data across times, days of the week, days of the month, locations, and tasks can also affect the validity of the data. Reporting the data and utilizing the reports to analyze the sampling strategy is another important job of a BBS Steering Team. The team often communicates to the observers needed changes in the observation frequency or distribution and can actually use observation frequency as a tool to impact safety. Increasing observations in an area can actually result in quicker behavioral change as well as provide additional data to understand a behavioral issue. Additional variables are often added as Steering Teams see the inadequacy of observation data. Variables can be used to ensure that the sample include observations of new employees, upset conditions, special tasks, contractors, or other issues that represent increased risks.



Hard Measurements for Soft Science: Behavior-Based Safety Has Evolved

Ongoing Pareto Analysis of accident data is necessary to maintain the correct focus on behavioral issues. As accidents decrease in frequency, checklists may need to be changed to reflect the behaviors most critical in continued accident reduction. Checklist changes also require communication with management and with the workforce at the site. Good Steering Teams become excellent communicators and they keep everyone focused on risk-reduction strategies and informed about new focus, past successes, and the status of the process and its key activities. The basic key process indicators (KPIs) of BBS include rates of participation among Steering Team and observers, hitting of target number of observations, quality of observations, number of action plans initiated and number completed, and increasing of percent safe. When these KPIs are right, BBS almost always produces a decrease in accident frequency and severity proportional to the increase in percent safe.

For an effort that began with almost total subjectivity, BBS has evolved into a process that has sophisticated and statistically-significant metrics. Since it is virtually impossible to improve anything you cannot measure, the measurement of human behavior is a necessary element in many improvement initiatives. In BBS, we are measuring behaviors related to safety, but the same techniques also enable the identification, definition and measurement of behaviors related to other goals. Any process that has a human element can potentially benefit from the measurement techniques developed during the evolution of Behavior-Based Safety.

Author – Terry L. Mathis
Founder & CEO - ProAct Safety, Inc