Confidential Close Call Reporting in the Railroad Industry: A Literature Review to Inform Evaluation

4/11/2006

Jonathan A. Morell*
Carl Hanssen
David Thompson
Richard Wallace
Barbara Wygant

*New Vectors, LLC
3520 Green Court, Suite 250
Ann Arbor, MI 48105-1579
jonny.morell@newvectors.net
734 302-4668

Confidential Close Call Reporting in the Railroad Industry: A Literature Review to Inform Evaluation

4/11/2006

Jonathan A. Morell*
Carl Hanssen
David Thompson
Richard Wallace
Barbara Wygant

*NewVectors, LLC
3520 Green Court, Suite 250
Ann Arbor, MI 48105-1579
jonny.morell@newvectors.net
734 302-4668

This work was performed under GSA MOBIS Contract Number GS-10F-0261K, Delivery Order Number DTRT-06-F-50001, "Formative and Summative Program Evaluation Technical Expertise to Assess FRA's Close Call Reporting System Demonstration Project". The work was performed by the Altarum Institute with assistance from the Evaluation Center, Western Michigan University. This document is a deliverable item of Task 1, Program Evaluation Plan.

Acknowledgements

Thanks go to Jordan Multer, Joyce Ranney, and Jane Saks of the Department of Transportation's Volpe National Transportations Systems Center for their advice and guidance as this document was developed. Thanks also go to Quentin Witkowski at Western Michigan University's Evaluation Center for his efforts at reviewing literature and organizing material. Interviews with the Steering Committee of the C^3RS program provided much needed information and wisdom about success factors in close call reporting system. They gave us the perspective we needed to interpret the literature in a fashion that served the purposes of this evaluation. Tom Raslear at the FRA's Office of Research and Development provided the funding and the inspiration needed to implement and evaluate C^3RS. We thank him for his vision and his persistence.

Author Biographies

Jonathan A. Morell, Ph.D.

Dr. Morell is an organizational psychologist whose work focuses on how organizations behave, change, and interact. He researches and writes about their internal dynamics, and the influence of their political, social, economic and technological environments. He conducts projects aimed at discovering current realities, evaluating innovation, researching trends, and developing methods. In addition to publishing widely, Jonny serves as Editor-in-Chief of the international journal *Evaluation and Program Planning* .

Carl E. Hanssen, Ph.D.

Dr. Hanssen is a Senior Research Associate at The Evaluation Center, Western Michigan University. His experience includes multi-year studies for private foundations, large federal government projects, a diverse group of survey research and data collection efforts for businesses, and evaluation of multiple training programs and curricula for a professional services firm.

David Thompson

Mr. David Thompson is a Corporate Scientist in the Enterprise Solutions Division of Altarum. He has 38 years of experience in the evaluation and modeling of healthcare delivery, logistics, and defense systems. He is currently directing Altarum's support of the Scientific Registry of Transplant Recipients to study national policies for the allocation of cadaveric organs for transplantation. His logistics experience includes the development of risk assessment methods for defense supply networks. His defense experience has focused on the development of mathematical methods for the evaluation of tactical, doctrinal, force structure, and materiel alternatives.

Richard Wallace

Richard Wallace is Senior Transportation Analyst with the Altarum Institute in Ann Arbor, MI. He has 14 years of experience in the transportation field and served as a principal in several large transportation system evaluations, including both roadway- and transit-based implementations. He recently completed a study for the Transportation Research Board that examines the cost-effectiveness of providing non-emergency medical transportation for transportation-disadvantaged populations, and he is currently working on a project concerning the application of high resolution, remotely sensed data to transportation system needs He is also heavily involved in a project to evaluate the environmental effects of agricultural conservation programs managed and administered by the U.S. Department of Agriculture's Natural Resources Conservation Service.

Barbara Wygant

Ms. Barbara Wygant is a project manager at The Evaluation Center at Western Michigan University. Ms. Wygant has a B.S. in Mathematics and is currently pursuing her Master of Public Administration from Western Michigan University. She currently coordinates many of the various data gathering and synthesis activities involved in evaluation projects; helps lead training workshops involving evaluation reporting and data analysis; conducts numerous site visits to observe and interview evaluands; performs literature reviews; and provides extensive report editing and writing of evaluation materials.

Executive Summary

In the field of safety and accident prevention, there is an accepted belief that "close calls" are reliable indicators of unsafe conditions, and that decreasing the incidence of close calls will decrease the incidence of accidents. Given the potential of close call reduction to improve safety, and given the difficulty of implementing a close call reporting system, the question arises: Can an effective close call reporting system be implemented in the railroad industry in the United States? The Confidential Close Call Reporting System (C^3RS) is the FRA's effort to answer this question. Because C^3RS is a test, it is being heavily instrumented, i.e., many aspects of its functioning are being evaluated, as is a wide range of possible impacts that the program may have. In preparation for the evaluation, a literature review was conducted with respect to critical aspects of the evaluation plan.

Structure and Functioning of Close Call Reporting Systems

Close call reporting systems have been implemented in numerous industries and in many different parts of the world. These industries include nuclear power, aerospace, chemical processing and production, healthcare, and several within the transportation sector (aviation, railroads, and marine). These systems share various characteristics, including: 1) incentives and other factors to promote reporting (such as confidentiality or anonymity) and protect reporters, 2) rules dictating who is eligible to report, 3) types of incidents that can be reported, 4) guidelines for how (or where) to report, 5) guidelines for the use of data derived from reports, and 6) identification of who reviews the reports and system ownership.

Problem Solving and Data Analysis

C^3RS problem solving teams will face two challenges. The first is to determine why a particular problem occurred, and what solutions might solve the problem. Continuous improvement (CI) methods are appropriate in these cases. Some problems, however, will derive from tight coupling of many different elements and related systems, and will require analysis in terms of complex systems and normal accidents.

Problems can be arrayed on a continuum from simple, to complicated, to complex. Both simple and complicated problems are amenable to traditional accident analysis and CI methods. However, the reasons for simple problems will be easily apparent and solutions will be obvious. Complicated problems will require more rigorous assessment. The prevalence of problems along the simple–complicated–complex continuum will follow a Pareto pattern, with most of the problems (at least early in the C^3RS process) falling at the simple end of the scale.

The second problem solving challenge is to choose among multiple possible solutions, each with advantages and disadvantages. Some of tools and techniques needed to help make these choices are the standard group process and group decision making methods that support any group process. A second useful group of tools are those that help evaluators discern the implications of choices in program design. These include logic models, scenario planning, backcasting, and assumption based planning.

Group Process

This section examines cross-functional and small problem-solving teams in large organizational settings. The literature review focused on how teams successfully

operate in relation to analyzing accidents, accident prevention and mitigation issues, labor and management relations, continuous improvement in organizations, and critical decision-making. This chapter examines the 1) composition of teams; 2) characteristics of successful teams; 3) obstacles to effective teamwork; 4) group process techniques; and 5) evaluation tools related to team measures.

Implementing Change

The critical link between problem solving and C^3RS impact is the ability to implement change based on the recommendations of Peer Review Team (PRT). Research on innovation adoption provides strong guidance as to the conditions under which innovations will be adopted. Stage models involving people, organization, and characteristics of the innovation are available, as are scales and assessment tools that can guide customization of instruments for determining why and when solutions proposed by PRTs are adopted. Given the nature of "simple" problems, the innovation literature suggests that most solutions proposed for the early problems received by C^3RS will be adopted. As time goes on fewer, but more difficult problems (complicated and complex one) will arise, and adoption rates may decrease.

Consequences of Close Call Reporting Systems

Close call reporting systems have been shown to provide both safety- and non-safety-related improvements. Numerous evaluations have shown that incident reporting systems in general produce substantial net benefits to the organizations involved. These benefits arise from reductions in the number of accidents and like events, as well as from other operational and managerial improvements. They also can produce less tangible, but valuable, changes in organization culture toward a fairer and more collaborative environment.

Safety Culture and Climate

The terms culture and climate have frequently been used interchangeably, but there are critical differences between the two. Climate is essentially the psychological perception of the state of culture. Also, climate tends to be concerned with intangibles, whereas culture is concerned with real observable acts and behaviors. Finally, climate tends to be temporal and subject to change, whereas culture tends to be enduring and resistant to change. For the purposes in evaluating C^3RS, culture is the more important construct, though it is likely that safety climate will help shape and explain the extent to which safety culture has in fact been impacted by the implementation of C^3RS.

Sustainment of Close Call Reporting Systems

An important aspect of the evaluation is to monitor whether sustainability is being built as C^3RS develops. Two different concepts are at play. "Sustainability" refers to the capacity of an innovation to continue. "Sustainment" is the extent to which an innovation maintains itself after start up funds are gone. For C^3RS, one important aspect of sustainment is rooted in organizational structure and behavior. A second important aspect of sustainability is the extent to which C^3RS is rooted in organizational culture. Organizational behavior and culture interact, with each having the ability to support or weaken the other. Evaluation must track the role of behavior and culture separately, as well as their combined effect.

As sustainment proceeds, it will be important to consider interactions between the nature of the C^3RS innovation and developing sustainability. C^3RS can be seen as a set of core functions wrapped in a larger bundle of form and function. It seems possible (even likely) that as C^3RS matures and adapts to changing circumstances, there will be a need to change the characteristics that support the core functions. An important measure of sustainability is the extent to which that kind of adaptation occurs.

Table of Contents

1.0 Introduction

In the field of safety and accident prevention, there is an accepted belief that "close calls" are reliable indicators of unsafe conditions, and that decreasing the incidence of close calls will decrease the incidence of accidents. Over time and across disciplines, the concept of a "close call" has taken on different names and somewhat different meanings. Reason (1997) and Van der Schaaf (1991) use the term "near miss" for events (or sequences of events) that lead to accidents. In health care, "error" can be an activity that exposes a patient to risk, but does not result in harm (Corrigan, Kohn et al.). But by whatever name, the themes of reliable indicator and causal sequence remain.

The intuitively appealing notion that accidents can be avoided by eliminating events that are related to accidents has empirical justification. Statistical analysis has shown a relationship between reducing the occurrence of "precursor events" and reducing the occurrence of accidents (Kirchsteiger, 1997). Other research has shown that near misses and accidents have common causal pathways (Wright & van der Schaaf, 2004).

The belief that reducing close calls can improve safety has been embraced by the Federal Railroad Administration, but they question whether an effective close call reporting system can be implemented in the railroad industry in the United States. There is good reason for skepticism. To improve safety, the existence of close call situations must be known, the reasons for those close calls must be determined, and change must be implemented. Each of these elements is problematic. Knowing that a close call occurred implies that someone close to the event (and perhaps even responsible for it) is willing to report the incident. Analysis of reasons for the close call requires obtaining potentially sensitive information. Implementing change implies that railroads are willing to invest resources and change procedures.

Given the potential of close call reduction to improve safety, and given the difficulty of implementing a close call system, the question arises: Can an effective close call reporting system be implemented in the railroad industry in the United States? The Confidential Close Call Reporting System (C^3RS) is the FRA's effort to answer this question. The name of the program conveys its logic. Confidentiality is needed to encourage the reporting of close calls, which in turn would lead to corrective action that would impact safety.

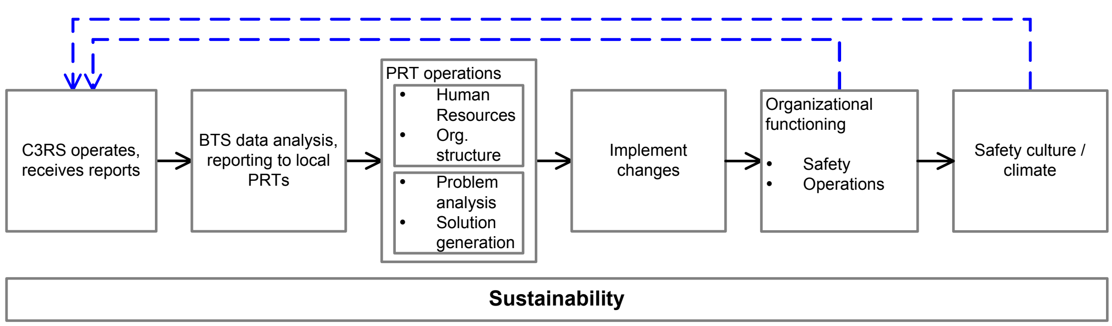
In C^3RS, close call information is reported by workers to the Bureau of Transportation Statistics (BTS), an arm of the Department of Transportation that has legislative protection allowing it to keep information confidential ("CIPSEA, the Confidential Information Protection and Statistical Efficiency Act" 2002). BTS's task is to analyze the close call and forward its findings (in sanitized form) to local Peer Review Teams (PRT) in participating railroads. The existence of the C^3RS is a triumph of perseverance and dedication by its champions, and of cooperation among three groups of stakeholders – the FRA, the railroad unions, and railroad management.

C^3RS is a test that is being implemented in four railroads for the purpose of determining: 1) how such programs should be implemented, 2) what impact they have, and 3) how they can be sustained in the long run as standard operating procedure in the entire railroad industry. Because C^3RS is a test, it is being heavily instrumented, i.e. many aspects of its functioning are being evaluated, as is a wide range of possible impacts that the program may have. That evaluation is being carried out by a team at the Altarum Institute, with assistance from The Evaluation Center at Western Michigan University.

There are two broad sources of information needed to design a powerful evaluation. The first is opinions from stakeholders about how a program is structured, why it should work as it does, and what impacts it may have. As critical as stakeholders' opinions are, however, they suffer from the restriction of representing a program/policy-specific view. It is also necessary to integrate the opinions of outside experts who can provide the wisdom of diverse experience and the guidance of empirical research (Morell, 2005). This document is a distillation of that outside expertise. It was developed by conducting a literature review on topics germane to the C^{3}RS evaluation.

As a prelude to the literature review, the evaluation team, in collaboration with the C^{3}RS Steering Committee, developed an elaborate logic model for the program. (Logic models are used to identify what needs to be measured, identify relationships among measures, develop the overall methodology, and assure consensus on what data needs to be generated by the evaluation. (Rogers, P., 2005) The C^{3}RS logic model was then used as a guide to determine what literature needed to be consulted. For instance, one element in the logic model is "analyze incoming reports about close calls". This element implies that problem solving teams know how to do such analysis. This in turn, requires that the evaluation team identify indicators of group problem solving ability. The full logic model decomposes processes into too fine a level of detail to be useful for guiding the literature review. Therefore a higher level version was used, as shown in Figure 1.

Figure 1: High Level Logic Model



Despite their visual similarity, evaluation logic models are not exactly the same thing as process flow diagrams. Process flow diagrams depict the sequence of business process, information flow, and material flow through an organizational setting. Logic models identify <u>metrics</u> that might be observed. The power of the model is that they show us what metrics are dependent on each other. For instance, Figure 1 indicates that *if* BTS does a good job of analyzing close calls, *then* local PRTs should be able to fulfill their function. To see the value of this insight, consider an evaluation finding to the effect that PRTs are functioning poorly. Is this because the teams are not structured properly, or because they are not getting useful information from BTS? The answer makes a difference in terms of improving the program. If we knew only that the teams were failing, we would have no direction to search for ways to improve the situation.

Figure 1 can be viewed as a pictorial table of contents for literature review. Each of its elements defines a domain of literature that was included. Information was interpreted through the filter of evaluation and measurement. Our focus was on what we were being told about evaluation, and not about the wider implications of the information. This filter was needed to maintain focus on what would otherwise be a far reaching (in fact nearly limitless) scope.

For the purposes of clear exposition, each topic covered in this document is written as a separate, independent chapter. But just as elements in the logic model for the evaluation have links and feedback loops, so too are there relationships among the literatures that are reviewed. In the interest of drawing tighter connections between the literature and the evaluation, these connections are pointed out as the discussion proceeds.

- Structural elements of close call reporting systems, e.g., confidentiality (Section 2.0), drive cultural factors (Section 0) such as a blame free environment.
- Criteria for a well functioning close call reporting system (Section 2.0) require beliefs that the system is having a desirable impact (Section 6.0).
- Understanding differences in the difficulty of change implementation (Section 5.0) has roots in the distinction between the simple and complex problem solving (Section 3.0) that will face process improvement teams.
- Implementing an appropriate logic of problem solving (Section 3.0) will depend on the group process (Section 4.0).
- We see parallels between the group process action review methods emanating from the military (Section 4.0), and continuous improvement methods that come from industry (Section 3.0).
- The impact of close call reporting systems on individual beliefs and attitudes (Section 6.0) affects the development of safety culture (Section 0).
- Improvements in safety and other outcomes of interest (Section 6.0) help make the business case needed for sustainability (Section 8.0).
- Culture (Section 0) and organizational behavior combine to support sustainability (Section 8.0).

2.0 Structure and Functioning of Existing Close Call Reporting Systems

Close call reporting systems have been implemented in numerous industries and in many different parts of the world. These industries include nuclear power, aerospace, chemical processing and production, healthcare, and several within the transportation sector (aviation, railroads, and marine). Within the aviation industry alone, systems have been developed and deployed in the U.S., Canada, Australia, the U.K., New Zealand, Germany, Japan, Korea, and Taiwan (Sullivan, 2001), though these do not all operate at the same level of maturity or success.

While differences between these various deployed close call reporting systems certainly exist, they tend to share some basic structures and functions, even if their specific implementations differ. These generally common characteristics include:

- Incentives and other factors to promote reporting (such as confidentiality or anonymity) and protect reporters
- Rules dictating who is illegible to report
- Types of incidents that can be reported
- Guidelines for how (or where) to report
- Guidelines for the use of data derived from reports
- Identification of who reviews the reports and system ownership

The remainder of this section addresses each of these characteristics of close call reporting systems.

2.1 Incentives for Reporting and Protections for Reporters

At least five factors have been suggested as important for encouraging employees to report close calls. As detailed by Reason (1997), these include 1) confidentiality, 2) protection from disciplinary action, 3) separation of organization collecting reports from the organization with authority to take action, 4) rapid and useful feedback, and 5) ease of reporting. Together, these five factors create an organization climate of trust that promotes reporting and provide some incentive for employees to report. Some existing systems go further into incentives by providing cash awards via a lottery or a simple prize (such as a t-shirt or certificate) for all reporters (Coyle, 2005; Phimister, Oktem et al., 2003)). These award systems, however, can threaten the confidential nature of the system.

While the majority of existing close call reporting systems documented in the literature offer confidentiality to those making reports, few offer anonymity, and some explicitly will not accept anonymous reports. Of the 12 non-medical close call reporting systems analyzed in detail by Barach and Small (2000), for example, ten provide confidentiality to reporters, while only two offer anonymity. Furthermore, the literature provides evidence of system failure when neither confidentiality nor anonymity is present, at least in part due to their absence (Tanaka, 2002). While confidentiality is widely cited as an important, perhaps necessary, element for creating an environment of trust that promotes reporting (Reason, 1997), anonymity is viewed by system designers and managers as an obstacle to appropriate data collection (see, for example, (Phimister et al., 2003), because it prevents potentially important follow-up data collection.

Despite this concern that anonymity can hinder needed data collection, some existing systems do provide it, and at least some of these appear to be successful. In the healthcare arena, for example, Suresh, et al. (2004) have documented a voluntary, anonymous reporting system based on specific medical specialties that succeeded in a neonatal intensive care setting, as measured by growing numbers of reports and promotion of a collaborative learning environment across medical disciplines. Furthermore, many who have studied safety in organizational settings and reporting systems have concluded that anonymity is an important factor for making employees comfortable with reporting (see, e.g., (Wilson-Donnelly, Priest et al., 2005). One study of multiple close call reporting systems (Barach, P. a. S. D. S., 2000)) concluded that anonymity may be required early in the evolution of a reporting system until employees come to trust the system and those reporting see practical results. Within the railroad industry, the well known CIRAS (Confidential Incident Reporting and Analysis System) in the U.K. is confidential, but not anonymous.

Regardless of their treatment of confidentiality and anonymity, most existing systems appear to confer some degree of protection to those employees who file reports. In many cases, confidentiality or anonymity is part of that protection, but even when one or both of these are not present, system designers and managers seek to offer reporters protections against disciplinary actions resulting from reported incidents. As described by Creek (1995), "…employees must trust that reporting close-calls truly represents an opportunity for learning, not discipline."

For systems that offer legal or other protects from discipline, they often impose some limitations on this protection. For example, they may require that the incident be reported within a certain period time after it occurs.

Not only are these structural elements important in their own right, but they also affect the development of safety culture (Section 0), as for instance by helping to establish a "blame free" environment.

2.2 Rules for Who Can Report Incidents

Close calls have the potential to occur in environments in which they are observed by those experiencing the close call, other human observers in the vicinity, and by equipment of various sorts (cameras, black box recording devices, etc.) (Van der Schaaf, 1991). In fact, in some cases, the person experiencing the close call may not even realize that a close call has occurred, but some other party might have observed the event. In response to this situation, existing close call reporting systems have established a variety of rules for defining who is eligible to report incidents, ranging from anyone at all to only directly involved employees with expertise in the associated work. From the literature, the latter choice appears to be linked with more technical fields, such as medicine (Suresh, G., Horbar et al., 2004), while the former is more associated with domains for which knowledge is more widely spread throughout the population, such as the Confidential Marine Reporting Scheme in place in Australia ("Confidential Marine Reporting Scheme (CMRS)," 2004).

In setting guidelines regarding who (or what) can report close calls, existing systems have inherently also affected the nature of the reports received. Accepting observational reports from humans not actually involved in the incidents themselves is likely to greatly increase the number of reports, but these reports will tend to lack the detail provided by involved persons (Van der Schaaf, 1991). Furthermore, automated reports from sensing equipment may allow a system to receive reports about incidents that are hard for humans to detect, but they raise concerns over privacy; this tradeoff may be worth it when incidents are hard for humans to detect but have high potential consequences if they proceed to an actual accident.

In some implementations, close call reporting systems have even been established virtually—that is, simulated events or incidents are created with the intent of observing how errors develop, progress, and accidents averted (or not). One such approach (Lawton & Parker, 2002) revealed that different classes of employees have different standards for making judgments about their work and the work of colleagues and thus vary in their willingness or likelihood to report close calls. In the chemical industry (Van der Schaaf, 1991) and other industrial settings (Masson, 1991), too, process simulators have been used to study how errors are generated, how recoveries occur, and how specific cognitive shortcomings (such as fixation errors) contribute to close calls and accidents.

2.3 Types of Incidents That Can be Reported

One common characteristic of close call reporting systems, across industries, is the delineation between incidents that can be reported via the system and those that should not be (or cannot be). For industries that require mandatory reporting by at least some employees for some serious types of incidents, this distinction often disallows confidential (or anonymous) reporting of incidents that fall under the realm of mandatory reporting requirements. Close call reporting systems also often prohibit reporting of unlawful acts.

2.4 Guidelines for How and Where to Report

Essential to the success of a close-call reporting system, employees eligible to report incidents must know how and where to report these incidents. Typical routes for reporting include phone numbers, web sites (with or without required log in to ensure that only eligible reporters access the system), and forms sent by non-electronic postal systems. Some systems offer a number of reporting options, such as the Confidential Safety Information Reporting Scheme ("Confidential Marine Reporting Scheme (CMRS)," 2004) developed in Australia as means for collecting additional safety data about the nation's train, bus, and ferry services. In many cases, at least for reporting systems that are not anonymous, initial reports are followed up with additional contact between reporters and report takers. In the case of CIRAS used in the railroad industry in the U.K., initial reports are followed by longer, in-depth interviews conducted either in person or over the phone between a trained interviewer and the employee making a report (Wallace, B., Ross et al., 2003). Thus, in this case, the follow-up interview is the primary source of data that is recorded and analyzed.

2.5 Guidelines for the Use of Data Derived from Reports

Close call reporting systems produce a large amount of data, much of it qualitative (or at least narrative based), and how this data is analyzed and otherwise used is a central component of the design of existing systems. This includes concerns related to data security, data handling and coding, data analysis, and distribution of data (generally in some processed form) to parties other than those taking the reports.

While this review addresses data analysis at some length in Section 3.0 (*Problem Solving Methods*), this section addresses this characteristic of existing close call reporting systems at a general level (such as who does the analysis and what data are used), as well as the other data use issues listed above, in general. Section 3.0 addresses the fine detail of precise analytical methods used.

Concerns related to data security must be designed into close call reporting systems to ensure that data are not misused. The data produced by reports must be treated with appropriate security, and reports must be adequately "cleansed" before being distributed. This concern can be particularly relevant when and where reporting systems offer anonymity or confidentiality. These systems tend to offer the most elaborate examples of secure data handling, usually combined with technical procedures that ensure confidentiality. For example, the voluntary, anonymous system described by Suresh, et al. (2004) employs an Internet system that accepts reports only from computers with specified IP addresses for which the standard Internet Information Server web log is deactivated to ensure that no identifying information is captured. Systems that are not anonymous generally rely on post-reporting purging of identifying information (see, for example, (Wallace, B. et al., 2003); ("Confidential Marine Reporting Scheme (CMRS)," 2004)) to ensure confidentiality, sometimes creating a separate data set that does not link to the identifying information in the initial report.

Once reports have been received, someone has to review the data. In some cases, before this occurs, the data are further processed, such as to convert open-ended narratives into coded data more amenable to quantitative analysis. Most notably, CIRAS in the U.K. employs an elaborate scheme for classifying chunks of textual information into one of more than one hundred specific codes (Wallace, B. et al., 2003). Rooted in the realm of hermeneutics and the work of Paul Ricoeur (1981), this method attempts to bridge the divide between qualitative and quantitative traditions. In the healthcare sector, too, data are generally placed into categories or codes (see, for example, (Callum, J. L., Kaplan et al., 2001), though here the tendency is often to have the reporters themselves do some of the selection of relevant categories or codes (Suresh, G. et al., 2004). This may not be surprising in highly specialized sectors of medicine in which only highly trained specialists, who often are the reporters, may be capable of evaluating important characteristics of the incident.

In the realm of data analysis, two general approaches are taken, and these two are quite complimentary. At one level, the details of individual incidents can be analyzed in detail to describe the incident and to identify causes, level of severity, potential consequences, how an accident was avoided, etc. At a second, more aggregate level, many incidents over a relatively long time frame can be examined for trends and the like. To a large extent, the second level of data is dependent on the first to create a detailed case record.

At the case level a variety of specific analytical methods are employed (see Section 3.0), though the tendency is to rely on textual or narrative data. For some reporting systems, attempts have been made to reduce the degree of subjectivity inherent in such analyses by employing rule-based coding schemes as detailed above (Wallace, B. et al., 2003). Conversely, aggregate data is analyzed quantitatively, often with no more sophistication than frequencies and averages (Suresh, G. et al., 2004), though quite elaborate Bayesian approaches have also been employed to study accident precursors (e.g., (Kirchsteiger, 1997).

One last issue associated with use of the data concerns its distribution beyond those who collect the reports. In many existing systems, the data can be viewed (in cleansed form) by many or all employees who participate in the reporting scheme. At the very least, those reporting generally obtain some sort of feedback on what became of their report, and this was true for all 12 systems surveyed by Barach and Small (2000).

2.6 Identification of Who Reviews the Reports and System Ownership

According to the literature (Tanaka, 2002), who reviews the incident reports is a critical design feature of close call reporting systems. Reason (1997), in his classic work, established that separation between who collects the reports and who has the authority to impose penalties or sanctions is important for an effective incident reporting system. Adding additional empirical evidence to this argument, Tanaka (2002), in his study of reporting systems in the healthcare system in Japan, found that lack of such separation was an important contributor to an ineffective reporting system.

Who reviews the reports also correlated highly with who performs the data analysis. In some cases, as for CIRAS, the data are analyzed by a third party (in this case, a university-based analysis group). In other cases, a locally based team reviews the reports (Callum, J. L. et al., 2001); (Suresh, G. et al., 2004). In some cases, a combination may be viewed as desirable, for example, when a local team requires additional expertise to fully evaluate a report (e.g., (Phimister et al., 2003). At another extreme, reports may be viewed by only one or two senior managers or supervisors (Coyle, 2005).

3.0 Problem Solving Methods

C³RS problem solving teams will face two challenges. The first is to determine why a particular problem occurred, and what solutions might solve the problem. For the most part, these skills will be those used in continuous improvement (CI) activities in many different industries and industrial sectors. These problem solving efforts assume a deterministic (often linear) causal path among the root causes and contributing factors behind problems or related groups of problems. For some problems, however, a different perspective will be needed, one based on emergent behavior in complex systems.

The second problem solving challenge is to choose among multiple possible solutions, each with advantages and disadvantages with respect to cost, effectiveness, permanence, time to implement, and the number of different problems that any given solution may affect.

3.1 Problem Solving Analysis Skills

In general, existing close call reporting systems share several characteristics that make their operations similar to those used in CI approaches. They are data based. They employ systematic procedures for identifying problems. They tend to use cross functional teams to analyze reports and form solutions to problems identified by these reports. They employ relationship models to identify root causes and contributing factors. These models are often deterministic, i.e., they rely on unambiguous causal relationships to link a set of determinants, over a predictable path, to a well defined problem or outcome. As an example of parallels between CI and safety analysis, consider the logical similarity between the Six Sigma approach to problem solving (Kwak & Anbarib, 2004), and Van der Schaaf's framework for designing near miss management systems(Van der Schaaf, 1991). (For a more expansive treatment of the structure and functioning of near miss systems, see Section 2 of this report.)

Six Sigma	**Van der Schaaff model**
• Define	Detection Selection
• Measure	Description Classification Computation
• Analyze	Interpretation
• Improve & control	Monitor

Working within the CI tradition has many benefits because CI methods (e.g., Six Sigma, Total Quality Management, Lean Manufacturing) have proven themselves in a wide variety of settings (Taylor & Wright, 2003), (Kwak et al., 2004; Womack, James P. , Jones et al., 1990)

The power of CI notwithstanding, sometimes assessing safety improvement may require analysis that treats incident causation in terms of multiple factors that are tightly linked , which affect each other across multiple domains (e.g. management practice and product design), and over varying time periods (e.g. FRA regulation setting and daily track warrant issuing procedures). These are situations that take on the flavor of "normal accidents" that can occur in complex systems (Perrow, 2001; Sammarco, 2005; Strauch, 2002; Weick, 2004). In these cases an incident may result for many different permutations and combinations of factors, and the whole concept of a "root cause" in the CI sense of the term, becomes meaningless. Many (in fact an infinite number of) causal pathways are possible, and the pathway that caused an accident once may be different from the pathway that caused it a second time. In these cases, analysis must focus on system characteristics that facilitate or inhibit the emergence of particular types of problems.

The literature shows a clear distinction between close call programs and accident analysis programs. The former resemble CI activities, while the latter frequently invoke system level reasons of the occurrence of an accident. (See for example, (Perrow, 2001; Strauch, 2002). Despite the theme of complex systems that is so prominent in accident analysis, it is important to maintain a sense of perspective. The accident ←→ complex system link is by no means perfect, and in fact, many accident analyses employ traditional root cause approaches. Examples include railroad employee fatalities (Office of Safety, 2003), train collisions (National Transportation Safety Board, 2003), and automobile crashes at intersections (National Highway Traffic Safety Administration, 1994, , 2001).

Thus, both close call reporting and accident investigation are heavily tilted in favor of garden variety industrial engineering approaches to quality improvement. This CI bias notwithstanding, it is entirely reasonable to believe that as time goes by, incidents will arise that require explanation in terms of complex systems. This seems especially likely for the BTS, which will be considering many incidents, from many railroads, over an extended period of time.

3.1.1 Continuous Improvement

In the practical, everyday craft of CI, success often comes when empowered teams, representing relevant knowledge domains, take the time to consider basic information. Often in these situations, a problem exists only because nobody has taken the time and trouble to ask why it is there in the first place. When they do ask, the reason for the problem, and the solution, become obvious. Sophisticated data gathering, analysis, and solution trade-offs are not needed.[1]

While simple analysis will often suffice, it cannot be assumed that conclusions and recommendations will always be obvious or easy. To be effective, a CI process must be armed with powerful tools and procedures. Those tools and procedures fall into two broad categories–data analysis and data acquisition. (It would seem that "data acquisition" should precede "data analysis." We are reversing the order to illustrate the paramount importance of data in fueling CI activities.)

[1] We have not found any data that speak to this issue, but our conversations with experienced CI practitioners convince us that this is indeed the case.

In addition to tools and data, successful CI requires that its practitioners organize activity around key concepts. Some of these are generic to all CI activity, while others are unique to specific problem solving domains.

3.1.1.1 Data Analysis

In large measure, the logic of data analysis in CI processes is embedded in CI's tools.[2] Table 3-1 summarizes the most common CI tools and the logic of problem analysis embedded in them.

Table 3-1: Common CI Tools: Description and Analysis Assumptions

Organizing and Displaying Information

Graphing / Visual Display of Information

Description: This category refers to many different ways of visually displaying data. Examples include plots of single variables over time, (e.g. number of derailments over a five year period), and scatter plots of multiple variables (e.g. number of violations by seniority of conductor). The *visual* dimension emphasizes intuitive (i.e. non- statistical) analysis of patters. When statistical analysis is called for, it is irrelevant whether the data are in table or chart form. Because of the importance of interpretation through inspection of the data, the format of the visual displays is important. (Tufte, 1983, , 1997)

Analysis Assumptions: If data are properly displayed, even without statistical analysis, patterns can emerge that help understand why (or when) problems occur.

Pareto analysis/charting

Description: Method of graphing the occurrence of events in order to classify frequency by type (e.g. type of animal – dog, cat, bird, reptile) by the number of people having such pets. (ISix Sigma.com, ; Wikipedia)

Analysis Assumptions: Based of Vilfredo Pareto's observation that 20% of the people own 80% of the wealth. This pattern has subsequently been observed in a great many real world activities. It is valuable in CI because it helps identify those few categories of events that might account for a disproportionately large percentage of an observed problem.

Process/flowchart Mapping

Description: A process map is a visual description of stages and decision points in a process. (ISix Sigma.com)

Analysis Assumptions: Knowing how a process works is important for interpreting information about why that process generates problems. Visual representations of process are useful for developing the necessary understanding.

Deriving Meaning From Data

Cause and effect models/diagram

[2] Some CI tools focus more on group process than on data and analysis. These are omitted here, but are discussed in Section 4.0 – "
Functioning of Problem Solving Groups "

Table 3-1: Common CI Tools: Description and Analysis Assumptions

	Description: Diagram that begins with the problem to be solved and then proceeds to connect prior conditions until root causes are determined. Several logical variants exist, each of which can be transposed into the other because the elements and their logical relationships are the same. The most common form is the Ishikawa diagrams named after its inventor, but more commonly referred to as the fishbone diagram because the model is often represented in form that resembles the skeleton of a fish. (ISix Sigma.com) Analysis Assumptions: 1) Unambiguous causal relationships can link a set of determinants, over a predictable path, to a well defined problem or outcome. 2) Those causal elements can be discovered. 3) If additional data are needed to construct the model, that data can be obtained.
5 Whys	Description: A technique for determining the cause of an observation by identifying the chain of causes behind an immediate problem. (For instance: The signal malfunction was caused by moisture in the circuitry. Moisture was caused by a corroded seal. The seal was corroded because it was let in place too long….) "Five" is a rule of thumb that experience shows to be sufficient for identifying root causes over which investigators can exercise some control. (Lean enterprise institute, 2003; Liker & Meier, 2006; Womack, James P. & Jones, 1996) Analysis Assumptions: Assumptions are the same as those for the "cause and effect diagram"
Statistical Methods	Description: There are a multitude of statistical techniques whose utility falls into three general categories: 1) to discern relationships among variables, 2) to discover underlying commonalities among a large number of variables, and 3) to help make informed decisions about when one variable (or group of variables) is different from another. Analysis Assumptions: Data contain useful information that cannot be discerned with the naked eye.

3.1.1.2 Data Acquisition

CI is fueled by data. Discussions of all the tools and procedures in Table 3-1 assume that the teams using those methods will get data to be organized, displayed, and analyzed. All "how to" discussion of CI are suffused with instructions and exhortations to collect data. The numerous case studies contained in the works cited in this section all devote much of their description to how data gathering teams were organized, and how they went about collecting data. Core tenets of the CI approach are that data based decision making can lead to improvement, that non-data based decision making will lead to poor decisions, and that no problem solving effort begins with a team already having the necessary information at hand.

From the point of view of close call problem analysis, the lesson from CI is not only that information is needed, but that the kind of the information needed does not arise until after a problem solving effort has begun. It is only once problem solving begins that teams can begin to decide what to look for and where to look. Moreover, as they begin to use the information they gather, new awareness may lead the information search down unanticipated paths. This is not to say that the information needed, or the sources of needed information, are always hidden, hard to get, or surprising. As we noted at the beginning of this section, a great deal of CI activity can succeed when teams with the requisite expertise take the time to consider basic information. Thus, we anticipate that much information searching (and subsequent problem solving) will be effective both at the BTS and within local PRTs. Experience with CI leads us to believe that there will be a great deal of low hanging fruit.

The abundance of easy to solve problems notwithstanding, the structure of the C^{3}RS is such that serious constraints may arise in those cases where rigorous information searching is required. BTS is limited in the number of times it can go back to a person submitting a report to ask for additional information. BTS is also limited in the amount of information it can obtain from the railroad that employees the person doing the reporting. Local PRTs may have access to more data, but the sanitized nature of the reports they get from BTS may make it difficult to identify information needs.

3.1.1.3 Key Concepts Guiding CI Activity

Continuous improvement is a philosophy of problem solving combined with deep knowledge about the setting in which process improvement is practiced. Without a commitment to the CI philosophy, and respect for domain expertise, none of CI's tools will lead to substantive change. To the extent that C^{3}RS resembles CI, this is also true for improving safety.[3]

The discussion of problem solving methods described above implies an opportunistic approach to improvement, i.e. a problem is observed, and forces are marshaled to solve it. It is true that a great deal of problem solving of this type takes place as part of CI. However, one of the real strengths of CI is its commitment to actively searching for problems to solve. Thus an important problem solving skill for CI practitioners is the skill of continually finding problems. The importance of this ability has deep roots in the CI movement, as reflected in the Deming Wheel (also known as Shewhart cycle, or the Plan – Do – Check – Act process.) Fundamental to this process is that it is continual, i.e. that each "act" initiates another round of planning, doing, and checking. (In fact, the original graphic for this process was depicted as a spiral in order to emphasize its open ended nature (Wilcox, 2005). A second manifestation of the need to continually find problems is reflected in the fact that Six Sigma is organized around the DMAIC cycle – define, measure, analyze, improve, and control. (Kwak et al., 2004). In the Six Sigma process, an important element of "control" is ongoing monitoring of the improved process in order to keep it functioning and to improve it. (The importance of evaluating improvement is echoed in the literature on innovation adoption, which indicates that organizations perform better when they learn from experience (Carayannis & Alexander, 2002; Carayannis & Turner, 2005). This literature will be reviewed in Section 5: "Implementing Change." We raise the issue here because it represents a research tradition that reinforces the importance of evaluating the consequences of process improvements.) Given the importance of continually and actively scanning for opportunities for improvement, an important question is whether C^{3}RS activity evolves toward an active searching for problems to solve, or whether it remains reactive.

[3] What follows is not a complete analysis of CI principles compared across varying techniques such as Lean, Six Sigma, and Total Quality Management. Such a presentation is far beyond the scope of this literature review. Rather, the discussion reflects our judgment as to which CI principles are particularly relevant to C^{3}RS problem solving skills.

Another aspect of CI that is not captured in the discussion of tools is its powerful commitment to domain expertise. This commitment goes beyond having expertise reside in cross functional teams. It involves organizing information in terms of concepts that experience and theory suggest are important to effecting change in a particular setting. The most highly developed use of domain-specific concepts in CI is found in the application of Lean principles to manufacturing. (This is not surprising because the lean manufacturing approach came out of Toyota, who developed it to improve manufacturing processes (Womack, James P. et al., 1990). Inspection of Lean terms include concepts such as "multi-machine handling", "product family matrix", "production preparation process," and "red tagging" (Lean enterprise institute, 2003). All of these and many more refer to specific activities that make up the manufacturing process. Each can be invoked when trying to understand why a problem exists, or what might be done about it. As data come in about a problem, the implications of that data are assessed relative to concepts like these. One of the reasons Lean has been so successful in manufacturing is that one of the skills its practitioners have is the knowledge of these manufacturing related concepts.

The domain of safety and accident analysis also has a well developed set of specialized terms and theories. Reinach and Viale review micro-level models which focus on human error, and higher level models that focus on both operator errors and contextual issues that may facilitate operator error (Reinach & Viale, 2006). They also review several theory based taxonomies to categorize operator error and contributing factors to accidents. While they choose a particular model for their own study, the important lesson from their review is that such models do exist, that they are based on theories of why accidents occur, and that the models and theories are very much like the domain expertise that drives CI, i.e., they serve as a structure for interpreting data, drawing conclusions, and making recommendations. As with CI, one of the problem solving skills that will be important both for the BTS and the PRTs is the ability to choose from among these models, and to apply them to their analysis activities and choices of solutions.

3.1.2 Problem Solving in a Complex Setting

The population of problem solving instances related to safety can be arranged on a continuum of difficulty ranging from simple, to complicated, to complex. The simple problems are the large number of situations where well constructed teams can easily make good decisions based on readily obtainable data. If the opinion of the CI expert we have consulted is correct, a very large number of C^3RS' problems will fall into this category. (See note 1 on page 10.)

Another large percentage of the problems will be complicated. These are problems for which deterministic analysis will still suffice to yield a solution, but where the play of many factors must be considered, data acquisition may be difficult, and sophisticated analysis tools will be needed.

At the far end of the continuum are truly complex problems that cannot be understood within the framework of CI and the incremental, deterministic problem solving approaches that CI uses. These are problems that exhibit behavior such as emergence, self-organization, and non-linearity (Kauffman, 1995; Marion, ; Wikipedia). These kinds of events have no clearly identifiable root causes in the traditional CI sense of the term, but instead, derive from tight coupling of many different elements and related systems. Complex system dynamics are the underlying reason why normal accidents occur.

From a practical point of view, neither the BTS nor the PRTs will have to contend with too many complex problems. One reason for this assertion is that all our reading on the topic of "close calls" reveals problem solving methods that clearly fall into the CI camp, as do all the examples used in those works. It is only the literature on accidents that invoke complex dynamics, and even then, many of the cases cited are deterministic (Perrow, 2001). Thus given the prevention role of C^3RS, it seems safe to say that only a few of their challenges will require an understanding of complex systems. On the other hand, complex problems may arise, and when they do, problem solving teams will need the tools to deal with them. For instance, it is reasonable to believe that as incident reports pile up, BTS and the PRTs will find apparently unrelated problems that seem to occur together, that the same problem occurs many times because of apparently different root causes, that long life cycle variation changes local behavior, or that non-linear interactions bring about unexpected changes. All of these are indicators that complex system dynamics might be at play. While the literature shows that complexity is not invoked to explain near misses, these examples all seem plausible. As a minimum, the teams should be able to intelligently consider the possibility that a seemingly CI-like problem may in have a complex explanation.

The problem with rigorously analyzing events (accidents and close calls) in terms of complex systems is that the requisite analysis calls for formal mathematical treatments and computer simulation, skills that are outside the expertise of either the BTS teams or the PRTs. Absent these skills, application of complex system principles to real world events requires using those principles only as heuristics and guiding principles. Fortunately, one of the most powerful examples of such application comes from normal accident theory, as first espoused by Charles Perrow (Perrow, 2001; Perrow, C., 1999). In particular, an excellent guide for applying complex system considerations to safety comes from Perrow's famous 2 x 2 matrix that jointly rates accidents on a continuum of loose to tight coupling, and on another continuum of interactions among elements, from linear to complex. The matrix provides a heuristic that allows people to decide if complexity is at play, and if so, what its implications are for understanding why an accident occurred, and what to do about it. As Weick puts it, Perrow's matrix helps to frame problems and link multiple levels of analysis (Weick, 2004).

3.2 Solution Choice Skills

As with initial problem solving, many solution choices will be obvious once basic data about a problem are collected. (See note 1 on page 10.) But it is not safe to assume that this will always be so. In fact, it is almost certain that some problem solving exercises will yield multiple possible solutions, and that the best choice is not readily apparent. Thus, C^3RS problem solving teams will need skills to assess alternatives and make recommendations. The difficulty in mastering those skills is analogous to the problem of formal complex system analysis, i.e. the techniques needed require formal training in esoteric disciplines, e.g. cost benefit analysis (Boardman, Greenberg et al., 2001), multiattribute utility analysis (Edwards & Newman, 1982), analytic network decision processing (Erdogus, Aras et al., 2006). There are, however, more accessible tools to help with choosing from among alternatives. One group of these tools is the standard group process and group decision making methods that support reasoned consideration of ideas and choices. These are the stuff of good group process, and are discussed in Section 4.0 of this report.

A second useful group of tools are those that help groups discern the implications of choices. The better known those implications, the wiser the choices among alternatives. The problem is analogous to that faced by program evaluators when they try determine the likely consequences of programs they are evaluating. (The better understood the consequences, the better the evaluation methodology.) Morell reviews six methods that are applicable to the context of near miss analysis (Morell, 2005). 1) Logic models are pictorial representations that identify relationships between what a program does, and its long and short term goals (Rogers, P., 2005). 2) By working out logic models for a proposed change, it will be possible to identify what impacts the program may have, how long it may take for those changes to be realized, and what events may intervene to prevent the desired changes. 3) An extension of the logic model approach is scenario planning, in which several alternate implementations are compared (Godet, 2000; O'Brien, 2003). 4) Logic models and scenario planning can also be worked in reverse. Using backcasting, one posits the desired future, and works back to see what paths might lead to that desired state (Dreborg, 1996). 5) Assumption based planning first identifies a desired end state, and then systematically identifies four elements that may affect attaining that end state (Dewar, 2002). Load bearing assumptions may easily break under changed circumstances. Signposts are leading indicators of problems. Shaping actions can be employed to support vulnerabilities in implementation. Hedging actions prepare for the possibility of failure. 6) The final technique summarized by Morell is simply an effort to learn from the experience of others. It seems reasonable that any proposed action by the BTS or a PRT has a variant that has been tried elsewhere. It would certainly help choose among proposed solutions if the past history of similar solutions were known.

4.0 Functioning of Problem Solving Groups

This section is based on a review of over 30 articles and 9 books about cross-functional and small problem-solving teams in large organizational settings. The literature review focused on how teams successfully operate in relation to workplace safety issues, analyzing accidents, accident prevention and mitigation issues, labor and management relations, continuous improvement in organizations, and critical decision-making. A wide range of articles and book chapters about close call or near-miss incident case studies in transportation, space exploration, process industries, nuclear and chemical industries, U.S. armed services, and competitive or "lean" manufacturing environments were also reviewed to highlight critical skills that are necessary (and oftentimes lacking) at both a team and leadership level to get to the causes, intervention, and prevention of close call accidents.

4.1 Composition of Teams

Based on a review of the literature, the following three factors were repeatedly cited for composing successful teams:

Functional representation and creative problem-solving by team members. Diversity of team members is important in relation to members having complementary backgrounds and talents. Surowiecki (2004) discusses the value of cognitive diversity and states that this was one of the elements missing most in NASA's space shuttle Columbia Mission Management Team (MMT). He states that unlike early missions such as Apollo where team members had worked in other industries, NASA employees today come to the agency directly out of graduate school, and they are much less likely to have divergent opinions. Along with the right functional mix, members should be open-minded, highly motivated, and creative (Proehl, 1996);(Parker, 2003); (Denison, Hart et al., 1997). Webber (2001) cites studies that show functional diversity may hinder social integration of team members and the positive impact of such diversity will not be realized without a leadership or organizational intervention.

Team size. Parker (2003) states the optimal team size is four to six members, with ten being the maximum for effectiveness. Quinn (1985) found that the most innovative companies limited project team size to six or seven (Holland, Gaston et al., 2000). (Katzenbach & Smith, 1993) studied dozens of teams in a variety of industries and concluded that "large numbers of people usually cannot develop the common purpose, goals, approach, and mutual accountability of a real team. And when they to so, they usually produce only superficial 'missions' and well-meaning intentions." Team tools in decision making, problem solving, and communicating were created to take advantage of small-group dynamics (Parker, 2003)

A skilled team leader. Chapters of books about cross-functional teams and many articles place great emphasis on the importance of choosing (and training) the right leader for a team. Webber (2001) reviews the critical role of the team leader and his or her actions prior to the formation of the team where leaders would typically be responsible for selecting high ability members, gaining top management support to create a high status project (which would be more likely to attract highly capable and talented team members), and selecting team members that have worked well together and/or with the leader in the past. A broad spectrum of literature states that the key tasks of leaders include positive relationship-building and promoting trust with and amongst team members. A leader should also be socially and politically aware of the organization's informal policies. Leaders are also crucial as connectors to key people outside team and should be effective at persuading and influencing others. Leaders should be able to effectively monitor a team's progress and intervene as necessary to keep them on track. However, their general role is often stated to be an enabler to the team (McDonough, 2000).

4.2 Characteristics of Successful Teams

The various components of teams—such as design, internal processes, behaviors, context, performance, environmental factors, effectiveness and outcomes—are highlighted in varying degrees in team literature. These types of elements are grouped or organized differently by authors (such as in outlines or figures) and many components overlap into similar categories. Based on a review of recent team literature articles, including research studies based on hundreds of cross-functional or peer review team participants, the following characteristics are noted as being important to the success of team functioning:

Formal, yet flexible integrative processes. Holland, et al. (2000) state that teams need clarity in direction, decision-making authority, and information in order to be optimally effective. Formal process should include clear definitions of teams' responsibilities, by-laws for the team, and scientific-like procedures for team-based decision making. Meetings should have an agenda and start on time, yet there should be "broad and flexible team process" to allow the team to take collective responsibility for resolving a diverse set of demands (Denison et al., 1997).

Frequent, genuine communication. Several large studies focusing on organizational teams showed that cross-functional communication and cooperation strongly correlate with success (Holland et al., 2000);(Souder, 1988). Cross functional teams are also crucial to weaving information up and down the hierarchal structure of an organization. Modular teaming is a recent strand of literature where the focus is on competitive industries that have widespread communication with dynamic, modular team arrangements. A complex process is divided into smaller parts and various teams focus on simpler parts and tasks that together make up a larger whole. This allows an organization to run multiple, parallel experiments and allows teams to expand and adapt (Evans & Wolf, 2005).

Information and benefits must be shared. A repeated and key element to team success is "transparency" and sharing information readily (Holland et al., 2000); (Parker, 2003);(Evans et al., 2005).

Trust and respect. Trust is a critical factor that is highlighted in all of the reviewed articles and studies on team functions and labor and management relations. Webber discusses the need for leaders to develop trust amongst the team quickly for effective communication, coordination, and cooperation. Parker (2003) states that "trust creates the pathway to open communication." Evans & Wolf (2005) state that "when information flows freely, reputation, more than reciprocity, becomes the basis for trust."

Cooperation. Good relationships and integration across functions is highly linked with common goals of team members (McDonough, 2000);(Webber, 2002).

Commitment. It is essential that team members have a strong sense of purpose and commitment to continuous learning.

Ownership, Empowerment, Accountability. Team-based accountability is a key success factor. Empowerment is defined in many ways and is often related to the "power" team members have in relation to final decisions, the authority to act, and not being hamstrung by special interests of management, labor, or external groups. Holland, et al. (2000) states "two key activities undermine team empowerment: meddling by functional managers and micro-managing by senior managers" (p. 238).

Management support and resource allocation. Team representation may result in conflicting organizational goals, competition for resources, and overlapping responsibilities. The climate or culture of an organization and buy-in and support key executives and management are factors to success of teams.

Training. Training in team-process skills and leadership training are frequently mentioned throughout team literature as essential in preparing team members to function effectively (Holland et al., 2000);(Webber, 2002).

To illustrate many of the characteristics mentioned above, we now turn to a case study that summarizes characteristics of successful teams. Proehl (1996) analyzed the responses of over 134 respondents who participated on various cross-functional teams at one of the largest transit companies in the U.S. Fewer than 50 percent of the participants reported that their project teams were a success, even though all participants received identical training and guidelines. The differences were related less to skills utilized in the team meetings and connected more to the attitudes and priorities of the team participants. From the questionnaires, four factors emerged as significant elements of team success:

1. The teams which succeeded had leaders, members, and sponsors who viewed the project as a priority.
2. These teams were task-oriented, maintaining their momentum and accomplishing their objectives in a timely way.
3. The leaders took an active role in keeping members informed and providing support and recognition to members.
4. Respect, open communication and mutuality among members were factors critical to success.

Based on interviews, the most frequently mentioned factors contributing to the success of the projects were

- the merits of the project;
- a clearly defined project;
- a chairperson with a positive attitude, commitment to the project and effective leadership skills;
- complementary skills of the members; and
- mutual respect and accountability among the members

When participants were asked to identify ways of improving the future performance of teams, the following recommendations were made:

- clear deadlines by which to complete the projects; monthly status reports required;
- greater clarity about the boundaries of the projects;
- more emphasis on selecting and training the leaders;

- communication about the cross-functional team activities; advertise what teams are doing;
- a designated coordinator whose responsibility is to follow up on teams; and
- meetings on company time with supervisors being held accountable for the participation of members.

4.3 Obstacles To Effective Teamwork

In the case study referenced above, the factors *inhibiting* success of teams were identified as the following: distrust between the executive staff and the employees; scheduling problems for team meetings; lack of support by the sponsor and executive staff; the members dropping out and not following through on assignments; the team lacking the resources to complete the project; and an organizational culture of ‘no change wanted here’ (Proehl, 1996).

Ancona (1990) described the role orientation of effective and ineffective consulting teams and described teams that were isolated, passive, or overly technical were far less successful than teams that proactively managed the political dynamics of their client organization (Denison et al., 1997).

Denison, et al., (1996) surveyed over 360 respondents from 43 cross-functional teams. They discuss how different teams include hierarchical, lateral, and inter-team dependencies that require continuous negotiation. Many subsystems may be operating within an organization that they refer to as “chimney” organizations. Chimney representatives may come to a team meeting to make sure their chimney “got what it needed” and didn’t lose out. Power dynamics between teams and functional organizations can greatly limit the autonomy of a team and must be managed proactively and coordination with other cross-functional teams is often required.

The concept of a team being interdependent with multiple departments and hierarchal structures is also reflected in another study. (Kleiner, Leonard et al., 2002) developed a time-series auto-regressive model to examine the impact of TQM and labor-management relations at a large commercial airplane manufacturing plant in the U.S. The overall impact of TQM was a slight reduction in labor productivity and an increase in production costs to the company for the total time period the policy was in place. Their discussions with management and labor leaders suggested that the failure of employee involvement was largely a result of top-down management. There were also attempts made by first-line supervisors to sabotage the TQM program for fear of losing control of the projection stages they oversaw (p. 212). The authors also discuss an industrial setting where TQM training never reached an intended critical mass due to production managers not releasing many key personnel whose absence from the line would have the threatened their ability to meet tight production delivery goals. Additionally, this period also coincided with a time of layoffs due in part to the inability of the company to deliver planes on time, which many in management in the union attributed to the TQM program (p. 203).

Teams should be aware of union procedures, policies, and employee legalisms when suggesting or implementing any safety measures. Section 7.0 provides further discussion on the culture and climate of organizations.

4.4 Group Process Techniques

The C^{3}RS logic model shows how information will be received and processed by PRTs. How the PRT handles data analysis and decision making will be critical. Section 3 discusses the analysis of information while this section highlights how teams interface and use group process techniques effectively.

Menon et al. (1996) found 'functional' conflict to be beneficial and 'dysfunctional' conflict to be harmful related to new products in organizations (see Table 4.1).

Functional Conflict	Dysfunctional conflict
Healthy and vigorous challenge of ideas, beliefs, and assumptions. Individual departments are willing to consider new ideas and changes suggested by other departments, and to volunteer information and ideas. Consultative interaction with useful give and take and opinions and feelings expressed freely.	Unhealthy behaviors such as distortion and withholding of information to hurt other decision-makers, hostility and distrust during interaction, and creating obstacles to the decision-making process. Opportunistic behavior such as departments overstating needs to influence others and information gate-keeping.

Table 4.1. Source: Menon, et al. (1996), from Holland, et al. (2000).

(Katzenbach & Smith, 2005) state that all effective teams develop rules of conduct at the outset including clear rules of behavior. They also recommend that teams get off to a fast and constructive start and to pay particular attention to first meetings and actions. It may be necessary for the team leader to intervene and take action to make sure all members are respectful and cooperative. The *Thomas-Kilmann Conflict Mode Instrument* (TK) is based on five situational approaches to handling conflict: *avoiding, accommodating, competing* (convincing, debating, voting, exerting power, etc.)*, compromising,* and *collaborating* (Huszczo, 2004).

In order to take advantage of functional representation within a group, each team member should regularly get a chance to speak. Surowiecki (2004) states that one of the consistent findings from decades of small-group research is that group deliberations are more successful when they have a clear agenda and when leaders take an active role in making sure that everyone gets a chance to speak. He also highlights the importance of deference and uses illustrations of how it is often violated in group settings. Just because an individual is talkative, has higher rank, or talks first does not necessarily mean that person is more knowledgeable or correct. However, sociologists have shown that these factors often play a large role in influencing group decision-making.

Team charters should define and clarify decision-making processes, including democratic and consensus procedures. Huszczo (2004) favors consensus decision making and states that teams must be educated on and clear regarding the criteria of decision-making processes. Many authors and researchers also warn of the pitfalls of emphasizing consensus over dissent and in extreme cases call it "group think" (Surowiecki, 2004);(Parker, 2003);(Holland et al., 2000).

Team members should always begin the decision-making process with an open mind. Many articles and studies focus on NASA team processes involved in the Columbia space shuttle disaster. Surowiecki (2004) discusses evidence-based juries versus verdict-based juries and illustrates how the Mission Management Team (MMT) operated as strictly the latter. Surowiecki states that "evidence-based juries usually don't even take a vote until after they've spent some time talking over the case, sifting through the evidence, and explicitly contemplating alternate explanations. Verdict-based juries, by contrast, see their mission as reaching a decision as quickly and decisively as possible." Surowiecki also illustrates how the team succumbed to "confirmation bias," which causes decision makers to unconsciously seek those bits of information that confirm their underlying intuitions. With Columbia, the MMT's conviction that nothing was wrong limited discussion and made them discount evidence to the contrary (p. 178).

Building vibrant human networks is a key to effective teamwork and is also key to implementing change in organizations. Evans & Wolf (2005) advise organizations to keep the information process simple and open; deploy pervasive collaborative technology; and keep work visible. Tools should work together through common standards and team participants should learn, share, and have a "rich semantic knowledge" of the subject matter. Teams should meet on a formal basis at regular intervals and there should also be many brief, "inexpensive" meetings that last five minutes or less. The informal meetings do not require a large amount of staff resources and they help facilitate information faster via informal networks versus formal procedures and lengthy written documents.

Table 3.1 in the preceding section lists continuous improvement tools for information processing and analysis. To use these, teams need group techniques to function effectively. The various group process techniques illustrated in this section may include specific training on team decision-making and leadership skills, including the use of videos, workshops, and retreats. All Toyota CI team members are trained in specific communication protocols that enforce discipline in decision making. The company and others following their lead use a one-sheet A3 reporting process to write lessons in a standard format (Liker et al., 2006);(Evans et al., 2005).

(Darling, Parry et al., 2005) examine the after-action review (AAR) method used by the U.S. Army's Opposing Force (OPFOR), a 2,500 member brigade who help prepare soldiers for combat. AAR meetings became a popular business tool after Shell Oil began experimenting with them and they are now used in several other large corporations. Organizations use the reviews to identify both best practices (which they want to spread) and mistakes (which they don't want to repeat). The AAR meeting addresses four questions: What were our intended results? What were our actual results? What caused our results? And what will we sustain or improve? The reviews or reports may generate a lesson during the AAR process, but OPFOR doesn't consider a lesson learned until its members have changed their behavior in response. The process involves trying out different assumptions and strategies so lessons do not grow stale. Units design numerous small experiments or short cycles of "plan, prepare, execute, AAR." A corporate version, called a before-action review (BAR) may also be used. Short BAR and AAR meetings frequently conducted in task-focused groups establishes feedback loops that can help a project team maximize performance and develop a learning culture over time. It is worthwhile to note that these group process techniques developed independently for civilian organizations parallel continuous improvement processes and problem solving methods in Chapter 3.

4.5 Evaluation Tools for Team Processes and Outcomes

There are a wide variety of examples available on surveys and questionnaires used to measure team processes, behaviors, and outcomes. Archival data relating to training or leadership program may also be used. Evaluation tools range from checklists, Likert-scale items, and team diagnostic questionnaires. Questions are included to see how well teams function with regard to items such clarifying goals, defining roles, communicating openly, and fostering creativity. Menkes (2005) cite studies that show strong correlation between intelligence and success and he promotes an executive intelligence test that measures reasoning and critical thinking skills rather than behavioral skills.

Thorough evaluation of team processes also include a qualitative phase or study. Participants in both the qualitative and quantitative phases include team leaders, members, and sponsors. Group meetings may be observed. Surveys may be distributed to all team members with systematic methods in place to obtain maximum feedback and participation.

5.0 Implementing Change

The critical link between PRT operations and C^{3}RS impact is the ability to implement change, i.e. to translate the PRT's recommendations into action. Thus it is incumbent on the evaluation to explain and measure the extent to which the PRT's recommendations are implemented. To do so, we propose to combine two perspectives. The first is "problem analysis" as discussed in Section 4.0 of this report. The second is the research literature on innovation adoption.

Recall that in Section 4.0 we made the point that most problems will be easy to solve, i.e. that root causes will be quickly apparent and that practical solutions will be readily available. The presentation went on to assert that problems can be arrayed on a continuum from simple, to complicated, to complex, with the incidence falling off in Pareto fashion from the simple to the complex.[4]

Research on innovation adoption provides strong guidance as to the conditions under which innovations will be adopted. Rogers' (1995) famous formulation offers a three part model in which innovation adoption is determined by factors related to individual decision makers, the nature of the innovation, and the organization in which change takes place. People go through the states of: 1) learning about an innovation, 2) being persuaded that an innovation is a good idea, 3) deciding to implement, 4) actually adopting the innovation, and 5) accepting the innovation as standard practice. Organizational characteristics are: 1) centralization of power and control within the organization, 2) expertise within the organization, 3) emphasis on following rules and procedures, 4) interconnectedness of subunits within the organization, and 5) organizational slack. (Although generalizations are dangerous, adoption seems to be related to organizations with low centralization, high skill, low emphasis on rules and formal procedure, and high organizational slack.) Important characteristics of the innovation (at least as perceived by potential adopters) are its: 1) advantage over current practice, 2) compatibility with existing beliefs and practices, 3) degree to which people understand the innovation, 4) extent to which it can be tested prior to full implementation, and 5) ability to observe other people using the innovation.

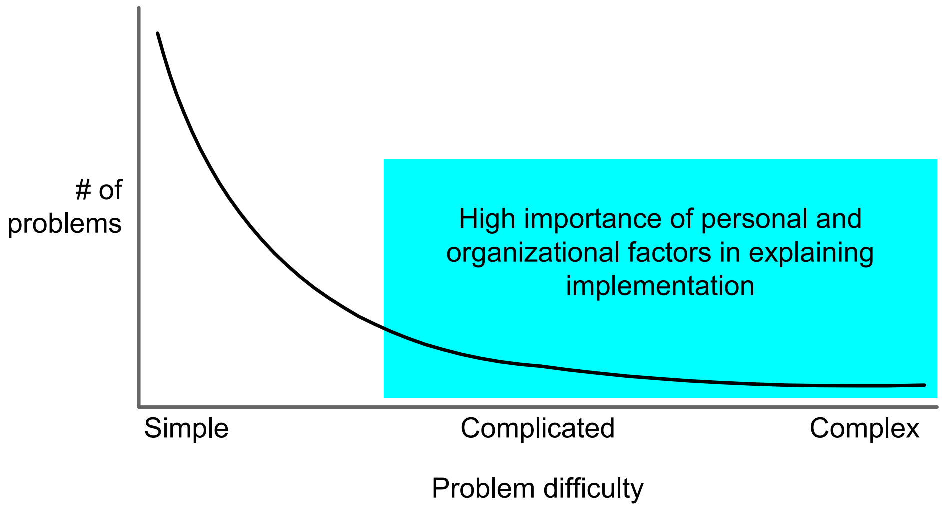
If in fact most of the safety problems discovered fall in the "simple" category, it seems reasonable to expect that they will (tracking against the Rogers categories): have a clear advantage over existing practice, fit with existing beliefs about operations, be understandable, testable on a small scale, and observable. In these cases problems of convincing people to try the innovation, and fit with organizational practice, seem likely to be relatively unimportant. As problems solving moves along the difficulty continuum, however, individual and organizational factors affecting

---

[4] The true shape of the distribution for C^{3}RS problems will be resolved empirically as the evaluation proceeds, but the 80/20 is a reasonable hypothesis based on experience with CI problem solving. There is at least some evidence that reasons for problems tend to cluster in large groups. In a study of errors in a neonatal intensive care unit, 47% of the problems were attributed to "failure to follow policy or protocol", while other factors came in at 27%, 22%, 13%, 12%, 10%, 10% and 9% (Suresh, G. et al., 2004)

innovation are more likely to come into play as factors which explain success.[5] This is the view that will guide evaluation of solution implementation, and is depicted in Figure 2.

Figure 2: Distribution of Safety Problems



The pattern depicted in Figure 2 is not static. It stands to reason that problems dealt with early in C^{3}RS' life cycle will stack up in the "simple" category if for no other reason than because problem solving exercises tend to begin by going after the "low hanging fruit", i.e. opportunities for fast, easy implementation of solutions that will be of obvious benefit. This is almost a restatement of the concept of a "simple problem". As time goes by, however, we might expect the ratio of simple: complicated (and complex) problems to change. This is so for two reasons. First, BTS and the PRTs will dispense with a large number of the simple problems.[6] Second, over time it seems reasonable to assume that patterns among separate problems will come to be observed. While each of these may have originally been traced to its own, insular, root cause, the pattern itself may result from dynamics that are deeper in the organization than each individual problem. For these, the search for causes and contributing factors will shift to organizational and cultural factors. As 0 will explain, these kinds of changes are difficult to effect.

---

[5] As an extreme example, consider NASA's inability to- sustain reforms after the Challenger accident, with the result that root the same problems reasserted themselves in with the Columbia (Columbia Accident Investigation Board, 2003)

[6] Our belief in open systems leads us to conclude that simple problems will never be eliminated, but it is reasonable to expect their incidence to decline.

6.0 Consequences of Close Call Reporting Systems

While generally designed to promote and improve safety-related outcomes, close call reporting systems have also been shown to produce other, non-safety-related outcomes and benefits. These positive outcomes, of course, should be compared to possible negative outcomes (most notably, the costs of operating the systems), giving rise to a summative category of consequences that consists of cost-benefit and other aggregative approaches. Thus, this section addressing the consequences of close call reporting systems is organized into three major subsections:

- Safety-related consequences of close call reporting systems
- Non-safety-related consequences of close call reporting systems
- Overall assessments of the consequences of close call reporting systems

As the first of these subsections will make clear, empirical studies demonstrating safety improvements after the development and implementation of a close call reporting system are very rare. As a result, the full-blown evaluation of outcomes associated with the C^{3}RS being developed by FRA and Volpe is a unique opportunity to rigorously link system with safety outcomes.

6.1 Safety-related Consequences of Close Call Reporting Systems

The primary purposes of a close-call reporting system, as described by van der Schaaf (van der Schaaf, 1991), are to gain qualitative and quantitative insights into errors and how they develop into incidents, as well as to maintain alertness to danger. These purposes all address safety consequences and outcomes. This perspective is deeply rooted in the notion that close calls, at least to some extent, represent steps along a (presumably causal) pathway to an actual accident or injury. Furthermore, this perspective holds that understanding and preventing close calls therefore will ultimately reduce the number accidents—the so-called common cause hypothesis (Heinrich, 1931). If this hypothesis is not valid, then concentrating on close calls may have little or no effect on the rate of injury-causing accidents.

In one systematic investigation of the common cause hypothesis using data from the U.K.'s CIRAS reporting system (Wright et al., 2004), researchers found empirical support for the common cause hypothesis within the railroad industry, though this support dropped when the researchers examined knowledge-based errors, because employees who lack knowledge may not recognize that a close call has occurred. Thus, Wright and van der Schaaf present evidence that safety-related improvements arising from close call reporting systems theoretically should reduce the frequencies of actual accidents and injuries, because these improvements address some of the same precursors that lead to both close calls and accidents. Other authors (Tanaka, 2002) have demonstrated similar findings. These works, however, do not show that the improvements arising from close class reporting systems are the necessary ones for reducing actual accidents, and other work (Hollnagel, 2004) demonstrates that not all precursors, in and of themselves, will lead directly to accidents.

Not only is the literature thin on evidence that close call reporting systems lead to the correct (necessary and sufficient) safety improvements, but empirical evidence of actual reduction in frequencies of accidents and injuries after the implementation of a close call reporting system is even harder to come by. To quote one author commenting on the reporting system implemented in the airline industry in Australia

(Sullivan, 2001), "… how many accidents has the CAIR [Confidential Aviation Incident Reporting] system prevented? No one knows."

Quite often, evaluative articles produce numerous counts of various types of reports, with such counts standing as the measure of successful program implementation (see, for example, (Suresh, G., et al., 2004). Other authors (Callum, J. L., et al., 2001) at least investigate the number of close calls that could have led to death or serious injury, but they still do not look at other incidents that were not part of the reporting system to determine if accident and injury rates were reduced.

While demonstration of reductions in frequencies of accidents and injuries has proven to be difficult, certainly close call reporting systems have led to changes in practices and procedures in many industries that experts deem as "safer." In the aviation industry alone, information derived from close call reporting systems has led to redesign of aircraft, air traffic control systems, airports, and pilot training (Tamuz, 1994). Other articles (Iedema, 2006; Wallace, S. J., 2000) detail numerous case studies of close calls and discuss the lessons learned from these cases. These studies certainly provide insights into what went wrong, and how an actual accident was prevented in certain cases, but they do not detail evidence of subsequent reductions in accident or injury frequencies based on knowledge learned from these close call case studies.

Finally, at a somewhat more philosophical extreme, some literature documents changes in attitudes and awareness that should improve general knowledge of safety issues. This could be described as being safety related, because they demonstrate links between close call reporting and the safety culture or climate (see Section 7.0 for a detailed elaboration of safety culture and climate). In this literature, such changes are documented at various levels, ranging from individual self-awareness (Iedema, 2006) to multi-institutional learning (Suresh, G., et al., 2004).

Given that the evidence that close call reporting systems reduce accidents and injuries is elusive, evidence that they have a positive effect on accident precursors and institutional safety culture is especially important. Indeed, these may well be the most visible signs of change that employees (those who make reports or others) detect in the wake of a reporting system. As explained in detail in Section 2.0, employees must be able to detect some positive change for a reporting system to have any hope of success or long-term sustainability.

6.2 Non-safety-related Consequences of Close Call Reporting Systems

Because close call reporting systems can identify and provide insights into operational problems that lead to close calls, they also have the potential to affect positively other operational parameters beyond safety. As is the case for safety-related outcomes, empirical evidence for these improvements is lacking, though at least one article (Iedema, 2006) provides evidence that close call reporting systems can alter relationships within the workplace, including interpersonal relationships and employee-institution relationships. These authors also provide evidence that reporting systems can cause hurt feelings and lead reporters to question their own self identity (e.g., can lead employees to question their own basic competence as a result of exploring the details of one of their personal failures).

6.3 Overall Assessments of the Consequences of Close Call Reporting Systems

Despite the lack of detailed evidence showing clear reductions in accidents and injuries, the literature presents some evidence (Corcoran, 1998) that close call reporting systems have an overall positive effect—that is, their benefits outweigh their costs. Furthermore, evidence exists that senior managers, including safety managers, believe that reporting systems are cost effective (Langley, Nolan et al., 1996). As a result, throughout many industries, close call and other incident reporting schemes are becoming more common. Indeed, in the U.S., the Institute of Medicine (part of the National Research Council of the National Academies), has explicitly called for the establishment of incident reporting systems (Kohn, Corrigan et al., 2000) as an important approach for identifying, understanding the consequences of, and reducing the frequency of medical errors. Finally, perhaps the most recent broad review of close call reporting systems across industries (Barach, P. & Small, 2000) concludes that such systems are cost-effective, though this review presents no quantitative results, such as cost-benefit ratios or dollars spent per quality adjusted life year gained.

7.0 Safety Culture and Climate

The evaluation logic model (see Figure 1) suggests that changes in organizational culture in general, and organizational safety culture, specifically, should 1) result from successful implementation of the C^3RS and 2) buttress continued operation of the system. Given these assumptions, we have developed the following discussion of safety culture with the goal of helping the evaluation team effectively document changes in safety culture over time.

This discussion is organized in six parts. First, we differentiate between culture and climate. Many researchers have used these terms interchangeably and it thus it is necessary to distinguish between the two. Second, we explore definitions of organizational culture in general and safety culture specifically for the purpose of establishing clear parameters for measurement. Next, we present evidence of a link between "hi" safety culture and "hi" safety, which supports the rationale for incorporating safety culture as a critical outcome for C^3RS. Satisfied that safety culture is important, we then look in turn at steps for creating a strong safety culture and common barriers to establishing such a culture. Last, we examine ways to measure safety culture. This final piece establishes the foundation for evaluation by suggesting practical and proven methods for examining the nature of an organizations safety culture.

7.1 Differentiating between culture and climate

The terms culture and climate have frequently been used interchangeably in the literature and in everyday discussions. Our perspective is that it is important to differentiate the two as our focus for C^3RS is on culture, primarily, and secondarily on climate. (Zhang, Wiegmann et al., 2002) posed the question of whether the two phenomena were in fact different. Through their analysis, they found some critical definitional differences between the two. First, they asserted that climate is essentially the psychological perception of the state of culture. Second, they stated that climate tends to be concerned with intangibles, whereas culture is concerned with real observable acts and behaviors. Third, he suggested that climate tends to be temporal and subject to change, whereas culture tends to be enduring and resistant to change. Gadd (2002) in his review of literature supports this view—he states that climate refers to attitudes whereas culture is more concerned with the underlying beliefs that shape those attitudes.

For our purposes in evaluating C^3RS, we would argue that culture is the more important construct, though it is likely that a measure of safety climate (i.e., the perceptions of workers, managers, and other stakeholders) will help shape our views on the extent to which safety culture has in fact been impacted by the implementation of C^3RS.

7.2 Organizational culture and safety culture

7.2.1 Definitions

Organizational culture is defined in a variety of ways. Gadd (2002) defines culture as the values that influence attitudes and behavior; (Simon & Leik, 1999) define culture as the norms values and assumptions present in an organization; Richter and Koch (2004) suggest that culture is the attitudes, beliefs, and behaviors the pervade an organization; and (Creek, 1995) asserts that culture is behavior based on core values that are critical for achieving an organization's vision and mission. Clearly, this is a loosely defined construct to the extent that some authors can not seem to nail down an operational definition within the same published work. For example, (Krause, 1995) states that culture has four elements—vision, values, goals, and assumptions (p. 11)—and then later states culture is the assumptions, values, and practices (p.17) in an organization.

Our purpose is not to argue for or against any of these definitions but rather to point out that culture is significant for driving organizational behavior and that this behavior emerges out of the values, beliefs, and attitudes commonly held by individuals within an organization. Thus, for the purposes of this work we will assert that culture incorporates shared understandings of these critical elements— a) organizational vision and mission b) values and beliefs, and c) behavior and actions. (This latter factor, though, may be the most important factor as this clearly differentiates culture from the more ethereal concept of climate.)

Given this definition of organizational culture, generally, we are still mostly concerned with 'safety culture,' specifically, because a critical outcome of the C^3RS is to improve safety culture in the U.S. railroad system (which in turn is hypothesized to feed the C^3RS system, improve its operations, and increase its impact and effectiveness). Safety culture, then, can be defined as that part of culture that is related to health and safety (Gadd, 2002). (Glendon & McKenna, 1995) are a bit more precise in stating that safety culture is the embodiment of a set of principals which loosely define what an organization is like in terms of health and safety. Richter and Koch (2004) simply state that safety culture is a focused aspect of organizational culture. We view this final definition as the most usable a) for its parsimony, b) because it implies that safety requires focus and targeted behavior, and c) because it suggests that measuring safety culture may be indistinguishable from measuring organizational culture.

7.2.2 Elements of safety culture

Moving from a general definition of safety culture to a series of observable variables is necessary for measurement. As such, the following discussion outlines what we view as the critical factors for safety culture, i.e., those things that can and should be measured for the purpose of evaluating safety culture change. As with defining culture, various authors have proposed various sets of safety culture factors.

Cooper (2000) for example, lists seven elements of safety culture. First, he states that there must be acknowledgement of the high risk nature of an organization's activities. Second, a blame-free environment must be in place. Third, there must be collaboration across ranks in addressing safety issues. Fourth, resources must be made available by management to address concerns. Fifth and consistent with the idea of a blame-free environment, communication across the organization must be based on trust. Sixth, there must be a shared perception of the importance of safety. Seventh, workers must be confident that preventative measures work. This is a useful set of factors because they encompass elements of climate based on perceptions and attitudes, as well as providing observable indicators of behavior.

Simon and Leik (1999) provide a simpler organizer, listing only three elements. First, they assert that in a strong safety culture, management must lead by example— they must not only talk the talk, but also walk the walk. Second, they argue that team stability is important. Constantly changing schedules, work flow, and team members erodes trust. Finally, and consistent with Cooper, they believe that participation in safety, whether on committees, policy boards, or from an accountability standpoint, must exist across all levels of the organization.

A three factor model is also suggested by Gadd (2002). Communication based on mutual trust must exist, there needs to be shared perceptions of importance of safety, and workers and managers must have confidence in the efficacy of preventative measures.

Pidgeon and O'Leary (1994) list four factors that describe a safety culture—1) senior management commitment to safety, 2) well-defined and accepted practices for handling hazards, 3) continuous organizational learning, and 4) a shared concern for hazards across the workforce.

Last, Reason (1997) gives us four critical elements of a safety culture. They are 1) a focus on reporting, 2) clear principals that differentiate between acceptable and unacceptable behavior, 3) flexibility in patterns of authority based on functional skills to meet changing situations, and 4) organizational learning.

There is clearly overlap across these sets of factors, though a few key elements appear most significant in this discussion. First, there must be shared perceptions that the work of the organization is potentially dangerous and that safety is in fact important. The idea of shared perceptions as a critical element of culture is well-documented. Clarke's (Clarke, 1999) specific example from the British railroad industry supports the assertion that manager and work perceptions need to be aligned for a strong culture to exist. This would seem appropriate when considering the railroad industry in comparison, say, to the accounting profession. Thus, high safety culture becomes more important and a necessary condition in cases where there is risk of death or injury to workers. Second, management behavior is significant. This behavior must consist not only of words, but also of action, such as allocating sufficient resources to address safety issues and putting in place management systems (e.g., hiring practices and training programs) that support safety. Third, all levels of the organization must be involved in safety issues. In the railroad industry, this is likely to require participation by management working alongside the unions and in cooperation with regulators and other stakeholders. Fourth, and perhaps most significant, is the issue of trust across the organization. Trust is important because it is manifested through open communication within and across ranks, occurs when workers do not fear being blamed for incidents, and ultimately is a reflection of core organizational values. This factor, in the railroad context, may be the most challenging to address.

This reduced set of variables provides the C^3RS evaluation with a tangible set of indicators of organizational safety culture. Before allocating resources to evaluate culture change, however, it is important to be satisfied that safety culture is in fact related to positive safety outcomes. The subsequent discussion provides evidence that in fact this is the case.

7.3 The importance of safety culture

Conventional wisdom suggests that organizations with a strong or 'hi' safety culture operate more safely. The common mechanism for how this is thought to occur is described by Krause (1995). His discussion asserts that attitudes lead to behavior that lead to consequences, which in turn lead to reinforcing attitudes that lead to behavior and so on and so forth. In this cycle, attitudes and behavior are clearly elements of organizational and safety climate and culture, respectively, as defined above. Consequences are the results of these attitudes and behavior—in our context of railroad safety, consequences may be punishment for rule breaking, a severe accident, or the successful resolution of a systemic problem through the C^3RS process. Regardless, these consequences reinforce attitudes (both positive and negative) which drive continued (or new) behaviors. The simplest analogy to this theory is that of a snowball gathering mass and speed as it rolls down hill. For C^3RS the task is to ensure that the snowball's mass is comprised of the 'right' things—i.e., positive values, beliefs, and behaviors about safety.

So if we accept the theory, what evidence exists that the theory holds true. There are multiple examples in the literature that show positive relationships between safety culture and safety outcomes.

Zohar is perhaps the leading researcher linking safety climate and safety outcomes. His collective work extends from 1980 to the present and approaches the question in several ways. For example, a recent study (Zohar & Luria, 2005) demonstrated that organizational climate is linked to safety behavior, but that linkage is fully mediated by climate levels within organizational subgroups. This is important because it highlights the reality that while there may be an overarching organizational culture, that there are also likely to be sub-cultures that play an important role in safety.

In 2003, Zohar conducted a study that focused specifically on the role of leadership in safety. This is consistent with the generally accepted view that leadership action (or inaction) is an important safety culture factor. Zohar (Zohar, 2002) found that aspects of leadership pertaining to concern for group members' welfare that stemmed from close relationships with workers was related to leadership practices that promoted a high safety culture. In turn, this high safety culture led to safer behavior as evidenced by lower accident rates.

In a third example, Zohar (Zohar, 2000) showed that employee perceptions of safety climate predicted micro accident records over a 5-month period. Combined, these three studies from Zohar lay an important framework for evaluating the possible effect of safety culture on actual safety behavior.

Other published work is consistent with Zohar's research. For example, (Hoffman & Stetzer, 1996) showed that both organizational and individual variables were related to unsafe behaviors. Organizational variables included group process, safety climate, and intent to approach co-workers engaged in unsafe acts. The individual variable was perception of role overload. This research also showed that at a team level, safety climate and unsafe behavior were significantly associated with actual accidents.

Hoffman, this time working with Morgeson (1999) also showed that the exchange relationships between individuals, leaders, and the organization has safety related implications. At the same time, these leadership relationships were combined with perceptions of organizational support and shown to be related to safety communication, safety commitments, and actual accidents. Finally, (Zacharatos, Barling et al., 2005) showed in two separate studies that high performance work systems were related to occupational safety and that trust in management and perceived safety climate mediated the relationship between high performance work systems and safety performance and safety incidents. These final studies are particularly interesting because they applied structural equation modeling techniques to the multivariate relationships. This technique might prove valuable in the evaluation of C^3RS as hypothesized relationships are known to be complex and multifaceted.

As with most empirical literature, these studies were narrowly focused and addressed limited hypotheses. But, the sum of this work suggests that there is merit to the idea that positive safety culture is related to positive safety outcomes. What is missing from this literature is evidence that a positive safety culture is likely to contribute to more pervasive and effective close call reporting like that proposed by the C^3RS. However, if we accept that reporting close calls is in fact one of the cultural behaviors that drives desired consequences (think about the Krause model), then it is entirely reasonable to hypothesize that positive safety culture will support the close call reporting system. Thus, measuring safety culture as a component of the C^3RS evaluation becomes an important activity.

7.4 Creating a safety culture

Krause (1995) argues that the key to establishing a safety culture is establishing the necessary management systems. In reality, these management system elements are similar to management systems in a variety of settings. As with the definition of safety culture as a focused aspect of organizational culture (Richter & Koch, 2004), creating a safety culture may be simply a function of incorporating safety concepts into the general management systems of an organization.

While Krause argues for incorporating 8 specific elements into management systems to support safety—a) training, b) safety measurement, c) facility design, d) consequences of unsafe behavior, e) accountability for actions, f) making safety a priority, g) allocating resources, h) demonstrating positive attitudes, g) measuring culture, and h) modeling behavior—these elements might be categorized into more general domains to facilitate understanding.

First, the issue of human resources must be considered. Krause includes training on his list, but human resource management processes might also incorporate recruiting practices, job descriptions, and reward systems. As an example, it is plausible that requiring certain safety credentials for new employees would be a reasonable way to bolster safety culture.

Second, measurement is needed. Krause indicates that both culture and more general safety measurement is needed to promote a safety culture. This is reasonable as conventional wisdom suggests that people will act in accordance with measurement.

Third, management behavior is critical. These behaviors may include modeling safety practices, e.g., wearing safety glasses, or participating rigorously in reporting systems. An opportunity for modeling behavior may also exist through participation on cross-functional safety teams in a manner consistent with involving all levels of an organization in safety. Modeling behavior also may become the most significant way to demonstrate positive attitudes about safety.

Fourth, resource allocation is critical. Again, this is a way to demonstrate management commitment and, importantly, little in a management system can be accomplished without sufficient resources. Attention to more expensive facility design that promotes safe workplace practices may be one way this principal is carried out.

Finally, the issue of personal accountability remains important. According to Pidgeon and O'Leary (1994), the most significant barrier to establishing safety culture may be a culture of blame. But, equally important is that a blameless culture is not appropriate. In other words, there must be a balance between consequences for unsafe action, personal accountability for action, and trust between employee groups.

Given these dimensions for establishing a safety culture, it would seem logical that significant barriers would exist in many organizations against establishing these types of systems.

7.5 Barriers to creating a safety culture

Circling back to Pidgeon and O'Leary (1994), the most significant barrier to establishing a safety culture may be a culture of blame. This type of culture might simply be described as one where workers and managers have a tendency to want to place blame on someone when something goes wrong. Thus, this is a culture of punishment and workers in this type of environment are likely to fear punishment and thus are unlikely to want to be associated with any type of safety incident—whether as the cause, victim, or reporter. Pidgeon and O'Learly, though, go on to say that a blameless culture, in contrast, is not the answer. There must be differentiation between tolerable and intolerable events and behavior. Thus, this raises the complexity of understanding what other barriers may exist to establishing a strong safety culture.

In a GAO study of the Veteran's Administration Patient Safety Program, a series of barriers to implementing a close call system was suggested (2004). These included both workflow barriers and attitudinal barriers. Workflow barriers included a) not knowing how to access the reporting system, b) unfamiliarity with the reporting system, and c) not having enough time. Attitudinal barriers were a) perceiving limited value in reporting, b) fear of blame, and c) fear of shame.

When workflow issues are considered, these would appear to be directly related to having the necessary management systems in place to support a close call reporting system. For example, unfamiliarity with the system may be a direct result of lack of communication, training, staff turnover, or resource allocation. Similarly, lack of time might be a function of work schedules and the process for reporting.

Attitudinal barriers relate directly to perceptions of culture held by workers. Consistent with Pidgeon and O'Learly (1994), the fear of blame appears as a prominent barrier. The perception that little will be done, however, may be more significant. This perception might be caused by observing inaction in the past, hearing managers talk about safety but not observing any follow-through, or simply a lack of involvement of various types of workers on safety committees and work teams—in other words, lack of participation may contribute to lack of communication which may then lead to the perception that little is happening or will happen if close calls are reported.

In summary, it would seem important to consider both attitudinal and workflow/management systems barriers in evaluation of culture change due to the C^3RS. The mechanisms for measuring the degree, to which these barriers are being overcome, in turn, provide the basis for evaluating safety culture change due to implementation of C^3RS. Next, we turn to a discussion of ways to measure culture for the purpose of supporting this important evaluation question.

7.6 Measuring safety culture

There are two leading approaches for evaluating safety culture. The first relies on surveys of employees based on the ideas that properly designed surveys more accurately measure effectiveness than procedural engineering (Petersen, 2001) and that employee perceptions of management are the most useful measures of culture(Gadd, 2002). We don't want to quibble about the fact that measuring attitudes may in fact be more akin to measuring safety climate rather than culture. But, as acknowledged above in section 7.1, it would seem appropriate to incorporate a measure of climate as a means of evaluating the degree to which culture has shifted. The second approach is based on observational studies for organizations. Combined, these two approaches form the foundation for what may be the most reasonable means of evaluating safety culture—the safety scorecard. This approach is akin to the balanced scorecard business measurement approach where multiple measurement activities are synthesized to present a holistic perspective or safety culture.

7.6.1 Safety climate surveys

A range of surveys have been used for measuring safety climate (or culture, depending on the author). For our purposes, we will collectively refer to these as safety climate measures as most rely on personal perceptions. Flin, et. al. (2000) reviewed 18 different scales used to assess safety climate. They demonstrated that the most commonly assessed dimensions were management, the safety system, and actual or perceived risks. As discussed previously, these are significant elements of safety culture.

In addition to the dimensions outlined above, a number of other organizers have been used, but most remain consistent with the elements of culture already described—employee-management relationships, communication, trust, personal perceptions of actual safety, safety processes, and organizational values. The following are descriptions of selected instruments intended to highlight these various constructs. It should be noted that these instruments are more recent than those reviewed by (Flin, Mearns et al., 2000), with the exception of the classic 40-item Zohar survey (1980).

For example, (Molenaar, Brown et al., 2002) proposed a structure that accounted for a) people, b) processes, and c) values. Each of these survey domains was comprised of multiple items where respondents were asked to rate their agreement on a scale of 1-6. For analysis, a particular attribute was only judged to be part of the corporate culture if there was strong agreement between the responses of various employee groups—e.g., managers, employees, and contractors. This requirement for agreement across ranks is consistent with the definitional factors of culture which call for shared perceptions ((Cooper, 2000);(Gadd, 2002; Leape, 1988) and a shared concern (Pidgeon & O'Leary, 1994) about safety and hazards.

DeJoy, et. al. (2004) used a 43-item survey to study safety climate in the retail sector. The instrument was organized into 8 scales that contained a single item (for the coworker support scale) to 9 items (for the organizational support scale). A specific scale labeled "Safety Climate" contained 7 items, though, based on our definitions discussed above, the variety of different scales would all seem relevant to measuring safety culture.

A similar approach was used by Evans et. al. (2005) in their study of organizational culture and safety at four lumber manufacturers. Their safety climate scale contained

17 items adapted from Zohar (1980) that asked employees to rate agreement with various statements on a 5-point Likert scale.

Lastly, the climate survey developed by Zohar (1980) deserves acknowledgement as so many instruments in use appear to have at least adopted its basic principals. This instrument contained 40 items where employees rated their perceptions of various aspects of safety culture from highly positive to neutral. Consistent with the definitional elements of safety culture described above, workers' perceptions of management attitudes toward safety and perceptions about the importance of safety in everyday work were found to be the most important factors for defining safety climate when compared with safety audit results.

Combined, these instruments share common characteristics. First, they are essentially perceptual instruments, i.e., they ask for agreement with a set of statements such as "my supervisor is concerned about my safety and health on the job." There is no attempt to validate how the supervisor actually feels and thus these instruments seemed based on the premise that perception is reality. Second, each contains multiple scales or dimensions. As suggested above, employee-management relationships, communication, trust, personal perceptions of actual safety, safety processes, and organizational values appear in some form on most instruments. This is significant because these dimensions appear closely aligned with the definitional elements of safety culture—though they are being measured as indicators of climate.

7.6.2 Behavioral Audits

While the range of safety climate surveys discussed above represent alternatives for measuring perceptions of culture, these perceptual measures should be partnered with measures of observed organizational characteristics that shed light on the level of safety culture within an organization. Combined, this approach can provide multiple perspectives on safety culture and should lead to a more accurate evaluation of the degree to which C^3RS has moved the meter in this area.

The behavioral audit proposed by Mol (2003) represents an approach based on observable phenomena. Mol's behavioral audit focuses in 6 domains and leads to the development of a 'cultural health scorecard.' Those domains are 1) management commitment and leadership, 2) shared ownership, 3) supervision and decision making, 4) safety issues and resolution, 5) safety capacity reservoir, and 6) employee commitment and resourcefulness. For each domain, observable, concrete metrics are suggested—though it is reasonable to presume that in practice these metrics might be altered to meet the particular context where the tool is being used. For example, under the shared ownership domain, two measures are provided as examples—the number of position descriptions that incorporate safety and the number of hazardous area employees involved in safety forums. Under the capacity reservoir domain, measures include the percent of employees with appropriate certifications and the amount of training provided versus training planned. From these examples, it is clear how this measurement approach deviates from measures of perception.

For each measure, several variables are addressed. First, there are target, baseline, and periodic 'actual' measures that enable the user to track progress toward organizational goals. Second, each metric is assigned a weight with the total weights equaling 100. To provide an organizational 'score,' each actual measurement is compared to the target. This proportion is multiplied by the assigned weight to establish a weighted score for that measure. The sum of the weighted scores then represents a total cultural health score. Lastly, progress required is calculated as the portion of the discrepancy between the target and observed values that should be addressed in the upcoming report period. So, if two years remain in a planned measurement period, the deficiency between target and observed might be divided by 2 to show the expected progress for the coming year.

Overall, this approach combines several important features. First, there is a focus on real observed phenomena that provides a nice counter to worker perceptions. Second, the approach recognizes that all measures are not equal and allows the user to assign weights as appropriate for that context. Third, the approach allows scores to be tracked over time. Fourth, the tool supports decision making by allowing managers to allocate resources to address the most important deficiencies and toward the areas where the greatest progress is needed and expected.

While this tool appears most relevant in site-specific contests, i.e., within a single factory, this is not, from our perspective a requirement. We would argue that many of the measures and artifacts advocated by this approach such as job descriptions, training records, and turnover statistics, can be collected at a distance. Thus, within the context of this evaluation of C^3RS, the behavioral audit may prove to be an extremely useful tool for examining culture at the various test railroad sites.

7.6.3 Safety Scorecards

Moving from the specific example of a behavioral audit to the more general example of a safety scorecards is needed because of the reality that individual measures of safety culture, for example a safety audit, are rarely found to be correlated with safety outcomes (Peterson, 1995). The difficulty in designing and implementing a scorecard, of course, is figuring out what should go on a scorecard and then convincing stakeholders about the appropriateness of those measures. Overall, though, most scorecards will have the following elements—1) a measure of incidents or safety outcomes, 2) results of a safety audit, 3) results from perceptual surveys, 4) some type of process measure indicating steps taken toward addressing safety, and 5) financial measures, e.g., claim costs. Petersen cites three examples—from Navistar, Kodak, and the National Safety Council that incorporate various measures in some or all of these categories.

In many ways, adopting a safety scorecard approach may satisfy the desire to evaluate safety climate through a perceptual survey and the more pressing need to examine safety culture through a behavioral audit approach. Our recommendation, then, would be to incorporate both techniques into the C^3RS evaluation.

8.0 Sustainment of Close Call Reporting Systems

C^{3}RS is an innovation that is receiving both a good deal of funding from outside of the railroad industry, and a great deal of special attention from a Steering Committee consisting of champions from labor, management, and government. These champions realize that success in their endeavor has two dimensions. The first is that C^{3}RS improve safety in the test sites. The second is that C^{3}RS establish itself as an ongoing process even after the outside funding and special attention is gone. The ability of an innovation to transition to standard practice is known as "sustainability". Sustainability is not something that happens at the end of a program. Rather it is a state that flows from a process that begins almost at the time an innovation is first implemented. An important aspect of the evaluation is to monitor whether sustainability is being build as C^{3}RS develops in its test sites.

8.1 Sustainment as a Function of Organizational Behavior and Culture

To understand sustainability, a bit more rigor in the use of terms is in order. Two different concepts are at play. "Sustainability" refers to the capacity of an innovation to continue. "Sustainment" is the extent to which an innovation maintains itself after start up funds are gone.(Schroeter & Morell, 2005). Two other terms related to sustainment are "institutionalization" , and "routinization" (Johnson, Hays et al., 2004; Yin, 1979). They are often used interchangeably, and both refer to embedding an innovation as standard practice in an organization. Because C^{3}RS is designed as a change within an organization (as opposed, say, to compliance to seatbelt usage laws), "institutionalization/routinization" is a particularly important aspect of sustainment. Thus from the point of view of this evaluation, the critical question is whether C^{3}RS' follows a path that moves it from an idiosyncratic activity to accepted standard operating practice within railroads.

Yin (1979) identified three distinct stages in the journey from innovation to acceptance.[7] "Passage" is the initial transition into routine practice. Examples might include making an activity an officially recognized part of a department's operations, or providing a line item in a budget. "Cycles" refer to repetitions of events that embed an innovation in routine practice. Examples include continued inclusion in budgets, or replacing a director after an incumbent leaves. "Niche saturation" refers to the integration of the innovation into other systems in an organization. Another way to look at niche saturation is to think of it as the degree of cross linkage between the innovation and other parts of the organization.

Many indicators of sustainment must be unique to C^{3}RS. For instance, particularly important elements of niche saturation are relationships between C^{3}RS and a railroad's other safety efforts, and between C^{3}RS and the railroad's larger CI mechanisms. Development of these unique indicators, however can be guided by scales, instruments and frameworks that have been developed to measure institutionalization (Barab, Redman et al., 1998; Goodman & Steckler, 1989).

[7] Actually, many stage models exist. Johnson et. al. (2004) found sixteen. But the Yin model stands up as useful and reflective of the logic embodied in the others.

Measures of sustainment can be used as indicators of whether sustainability is developing over the course of C^3RS' implementation. For instance, if we know that links with other CI activities are important, ongoing evaluation can detect whether those links are being developed.

The discussion above places a strong emphasis on organizational behavior, e.g. functions allocated to departments or line items included in budgets. However, as we have seen in Section 0, culture, (i.e. a commonly accepted set of beliefs and related normative behaviors), also plays an important role in determining how an organization functions. The role of culture in sustaining C^3RS is important because it is entirely possible for organizations to subscribe to procedure in form, but not in substance. For instance, a railroad may have an institutionalized safety program that never deals with truly important issues because employees don't believe that substantive problems are worth reporting, or because teams do not recommend powerful corrective actions because they believe that management will not take them seriously.

Thus "culture" as a concept will play two roles in the evaluation. One role is as an outcome variable that mediates the feedback loop between initial reporting of problems and improved safety. The second role is as a variable that affects sustainability.

To get a full understanding of sustainability, culture and institionalization must be seen as more than just two factors whose vector sum moves sustainability in one or another direction. Rather, they must be analyzed as two factors that interact with each other. Institionalization drives standard behavior, and given our behavioral based view of culture, it is reasonable to assume that as behavior changes, so too will the organizational safety culture. Further, as the common beliefs and associated normative change that comprises culture takes place; the willingness of the organization to implement formal change should increase.

8.2 Sustainment and the Evolution of C^3RS

As sustainment proceeds, it will be important to consider interactions between the nature of the C^3RS innovation and developing sustainability. C^3RS can be seen as a set of core functions wrapped in a larger bundle of form and function. The core functions are reporting, problem analysis, and change implementation. Other elements support those core functions. For instance, BTS is involved because of the belief that absent such an organization, problem reporting would not occur. One PRT per railroad is, at present, considered the best way to install a problem solving function in each participating company. (There could after all, be several PRTs per railroad, or one PRT that cut across all participating railroads.) It seems possible (even likely) that as C^3RS matures and adapts to changing circumstances, there will be a need to change the characteristics that support the core functions. An important measure of sustainability is the extent to which that kind of adaptation occurs. For instance, consider the possibility that the success of C^3RS in one of a railroad's service units spurs the company to implement the program across its other service units. In such a case, the railroad may have a strong desire to cut the learning curve at other locations by adding knowledge transfer and mentoring roles to the duties of its existing PRT. If this scenario were to unfold, the nature of the C^3RS program would change because in contrast to their original mission, the PRTs' success would now be measured in terms of knowledge transfer and mentoring. If they rose to that challenge, sustainability, i.e. the capacity for C^3RS to become institutionalized, would increase. If they failed at knowledge transfer and mentoring, sustainability would decrease.

9.0 References

Barab, S. A., Redman, B. K., & Froman, R. D. (1998). Measurement characteristics of the levels of institutionalization scales: Examining reliability and validity. *Journal of nursing measurements, 6*(1), 19 - 33.

Barach, P., & Small, S. D. (2000). Reporting and Preventing Medical Mishaps: Lessons from Non-medical Near Miss Reporting Systems. *BMJ, 320*, 759-763.

Barach, P. a. S. D. S. (2000). Reporting and Preventing Medical Mishaps: Lessons from Non-medical Near Miss Reporting Systems. *BMJ, 320*, 759-763.

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimar, D. L. (2001). *Cost-benefit analysis: concepts and practice*. New Jersey: Prentice Hall.

Callum, J. L., et al. (2001). Reporting of Near-miss Events for Transfusion Medicine: Improving Transfusion Safety. *Transfusion, 41*.

Callum, J. L., Kaplan, H. S., Merkley, L. L., Pinkerton, P. H., Fastman, B. R., Romans, R. A., et al. (2001). Reporting of near-miss events for transfusion medicine: improving transfusion safety. *Transfusion, 41*, 1204-1211.

Carayannis, E. G., & Alexander, J. (2002). Is technological learning a firm core competence, when, how and why? A longitudinal, multi-industry study of firm technological learning and market performance. *Technovation, 22*(10), 625-643

Carayannis, E. G., & Turner, E. (2005). Innovation diffusion and technology acceptance: The case of PKI technology. *Technovation*, In Press.

Clarke, S. (1999). Perceptions of organizational safety: Implications for the development of safety culture. *Journal of Organizational Behavior, 20*(2), 185-198.

Columbia Accident Investigation Board. (2003). *Report of Columbia Accident Investigation Board, Volume I*, from http://www.nasa.gov/columbia/home/CAIB_Vol1.html

Confidential Information Protection and Statistical Efficiency Act (2002).

Confidential Marine Reporting Scheme (CMRS). (2004). In A. M. S. Authority (Ed.) (Marine Notice 10/2004 ed., pp. 4): Australian Maritime Safety Authority.

Cooper, M. D. (2000). Toward a model of safety culture. *Safety Science, 36*, 111-136.

Corcoran, W. R. (1998). *The Phoenix Handbook: The Ultimate Event Evaluation Manual for Finding Profit Improvement in Adverse Events*. Windsor, CT: Nuclear Safety Review Concepts.

Corrigan, J., Kohn, L. T., & Donaldson, M. S. *To Err Is Human: Building a Safer Health System* Washington DC: National Acadamy Press.

Coyle, G. A. (2005). Designing and Implementing a Close Call Reporting System. *Nurs Admin Q, 29*(1), 57-62.

Creek, R. N. (1995). Organizational Behavior and Safety Management. *Professional Safety, 40*(10), 36.

Darling, M., Parry, C., & Moore, J. (2005). Learning in the thick of it. *Harvard Business Review*, 84-92.

Denison, D. R., Hart, S. L., & Kahn, J. A. (1997). From chimneys to cross-functional teams: developing and validating a diagnostic model.

Dewar, J., A. . (2002). *Assumption based planning: A tool for reducing avoidable surprises* Cambridge: Cambridge University Press

Dreborg, K. H. (1996). The essence of backcasting. *Futures, 28*(9), 813 - 828.

Edwards, W., & Newman, J. R. (1982). *Multiattribute evaluation*. Thousand Oaks CA: Sage.

Erdogus, S., Aras, H., & Koc, E. (2006). Evaluation of alternative fuels for residential heating in Turkey using analytic network process with group decision making. *Renewable and sustainable energy reviews, 10*, 269 - 279.

Evans, P., & Wolf, B. (2005). Collaboration Rules. *Harvard Business Review*, 96-104.

Flin, R., Mearns, K., O'Connor, P., & Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*, 177-192.

Gadd, S. (2002). *Safety culture: A review of the literature*: Health Safety Laboratory.

Glendon, A. I., & McKenna, E. F. (1995). *Human safety and risk management.* London: Chapman and Hall.

Godet, M. (2000). The art of scenarios & strategic planning: Tools & pitfalls *Technological Forecasting & Social Change, 65*, 3 - 22.

Goodman, R. M., & Steckler, A. (1989). A framework for assessing program institutionalization. *Knowledge in society: The international journal of knowledge transfer, 2*, 57 - 71.

Heinrich, H. W. (1931). *Industrial Accident Prevention*. New York: McGraw-Hill.

Hoffman, D. A., & Stetzer, A. (1996). A cross-level investigation of factors influencing unsafe behaviors and accidents. *Personnel Psychology, 49*, 307-339.

Holland, S., Gaston, K., & Gomes, J. (2000). Critical success factors for cross-functional teamwork in new product development. *International Journal of Management Reviews, 2*(3), 231-259.

Hollnagel, E. (2004). *Barriers and Accident Prevention*. Burlington: Ashgate.

Huszczo, G. E. (2004). *Tools for Team Leadership: Delivering the X-Factor in Team eXcellence*. Palo Alto: Davies-Black Publishing.

Iedema, R., et al. (2006). Narrativizing Errors of Care: Critical Incident Reporting in Clinical Practice. *Social Science & Medicine, 62*, 134-144.

ISix Sigma.com. *Cause & Effect Diagram*, from http://www.isixsigma.com/offsite.asp?A=Fr&Url=http://www.skymark.com/resources/tools/cause.htm

ISix Sigma.com. *Pareto Chart*, from http://www.isixsigma.com/library/content/c010527a.asp

ISix Sigma.com. Process Mapping and Flow Charting: ISix Sigma.com.

Johnson, K., Hays, C., Center, H., & Daley, C. (2004). Building capacity and sustainable prevention innovations: a sustainability planning model *Evaluation and Program Planning, 27*(2), 135-149

Katzenbach, J. R., & Smith, D. K. (1993). *The Wisdom of Teams: Creating the High-Performance Organization*. Boston: Harvard Business School Press.

Katzenbach, J. R., & Smith, D. K. (2005). The Discipline of Teams. *Harvard Business Review*, 162 - 170.

Kauffman, S. (1995). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity* Oxford: Oxford University Press.

Kirchsteiger, C. (1997). Impact of accident precursors on risk estimates from accident data bases. *Journal of Loss Prevention in the Process Industries, 10*(3), 159 - 167.

Kleiner, M. M., Leonard, J. S., & Pilarski, A. M. (2002). How Industrial Relations Affects Plant Performance: The Case of Commercial Aircraft Manufacturing. *Industrial and Labor Relations Review, 55*(2), 195-218.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To Err Is Human: Building a Better Health System*. Washington, DC: National Academy Press.

Krause, T. R. (1995). *Employee-Driven Systems for Safe Behavior: Integrating Behavioral and Statistical Methodologies*. New York: Van Nostrand Reinhold.

Kwak, Y. H., & Anbarib, F. T. (2004). Benefits, obstacles, and future of six sigma approach. *Technovation*, 1-8.

Langley, G., Nolan, K., Nolan, T., Norman, C., & Provost, I. (Eds.). (1996). *The Improvement Guide*. San Francisco: Josey-Bass.

Lawton, R., & Parker, D. (2002). Barriers to incident reporting in a healthcare system. *Qual Saf Health Care, 11*, 15-18.

Lean enterprise institute. (2003). *Lean Lexicon: A graphical glossary for lean thinkers*. Brookline, MA: Lean Enterprise Institute.

Leape, L. L. (1988). Promoting patient safety by preventing medical error. . *JAMA, 280*(16), 1444-1447.

Liker, J. K., & Meier, D. (2006). *The Toyota Way Fieldbook*. New York: McGraw Hill.

Marion, R. *The Edge of Organization: Chaos and Complexity Theories of Formal Social Systems*. Thousand Oaks CA: Sage.

Masson, M. (1991). Understanding, Reporting and Preventing Human Fixation Errors. In T. W. van der Schaaf, D.A. Lucas, and A.R. Hale (Ed.), *Near Miss Reporting as a Safety Tool*. Oxford: Butterworth-Heinemann.

McDonough, E. F. (2000). Investigation of Factors Contributing to the Success of Cross-Functional Teams. *J PROD INNOV MANAG, 17*, 221-235.

Molenaar, K., Brown, H., Caile, S., & Smith, R. (2002). Corporate Culture. *Professional Safety, 47*(7), 18-27.

Morell, J. A. (2005). Why are there unintended consequences of program action, and what are the implications for doing evaluation? *American Journal of Evaluation, 26*(4), 444 - 463

National Highway Traffic Safety Administration. (1994). *Examination of Intersection, Left Turn Across Path Crashes and Potential IVHS Countermeasures* (No. DOT-VNTSC-NHTSA-94-4). Washington DC: National Highway Traffic Safety Administration.

National Highway Traffic Safety Administration. (2001). *Analysis of Crossing Path Crashes* (No. DOT-VNTSC-NHTSA-01-03). Washington DC: National Highway Traffic Safety Administration.

National Transportation Safety Board. (2003). *Collision of Burlington Northern Sana Fe Freight Train With Metrolink Passenger Train Placentia, California April 23, 2002* (No. NTSB/RAR-03/04). Washington DC: National Transportation Safety Board.

O'Brien, F. A. (2003). Scenario planning - lessons for practice from teaching & learning. *European Journal of Operational Research, 152*, 709 - 722.

Office of Safety. (2003). *1998 railroad employee fatalities: An analytical study*. Washington DC: Federal Railroad Administration.

Parker, G. M. (2003). *Cross-Functional Teams: Working with Allies, Enemies, and Other Strangers*. San Francisco: Jossey-Bass.

Perrow. (2001). Normal Accidents. *International Encyclopedia of the Social & Behavioral Sciences*.

Perrow, C. (1999). *Normal Accidents, With an Afterword and Postscript on Y2K*. Princeton NJ: Princeton University Press.

Petersen, D. (2001). The safety scorecard: Using multiple measures to judge safety system effectiveness. *Occupational Hazards, 63*(5), 54-58.

Peterson, D. (1995). The Safety scorecard: Using multiple measures to judge safety system effectiveness. *Occupational Hazards, 63*(5), 54-58.

Phimister, J. R., Oktem, U., Kleindorfer, P. R., & Kunreuther, H. (2003). Near-Miss Management Systems in the Chemical Process Industry. *Risk Analysis, 23*(3), 445-459.

Pidgeon, & O'Leary. (1994). *Need to get this*.

Proehl, R. A. (1996). Enhancing the effectiveness of cross-functional teams. *Leadership & Organization Development Journal, 17*(5), 3-10.

Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Hampshire, England: Ashgate.

Reinach, S., & Viale, A. (2006). Application of a human error framework to conduct train accident/incident investigations. *Accident Analysis and Prevention, 38*, 396-406.

Richter, A., & Koch, C. (2004). Integration, Differentiation and Ambiguity in Safety Cultures. *Safety Science, 42*, 703-722.

Rogers, E. M. (1995). *Diffusion of innovations*. New York: Free Press.

Rogers, P. (2005). logic model. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 232-325). Thousand Oaks, CA: Sage Publications.

Sammarco, J. J. (2005). Operationalizing normal accident theory for safety-related computer systems. *Safety Science, 43*, 697 - 714.

Schroeter, D. C., & Morell, J. A. (2005, Oct 24th - 30th). *Criteria for evaluating program sustainability within a multi-site context: A longitudinal approach* Paper presented at the Joint Meeting of the American Evaluation Association and Canadian Evaluation Society, Toronto CA.

Simon, S. I., & Leik, M. (1999). Breaking the Safety Barrier: Implementing Culture Change. *Professional Safety, 44*(3), 20-25.

Souder, W. E. (1988). Managing relations between R&D and marketing in new product development projects *Journal of Product Innovation Management, 5*, 6 - 19.

Strauch, B. (2002). Normal accidents - yesterday and today. In C. Johnson (Ed.), *Workshop on the investigation and reporting of incidents and accidents (IRI! 2002)* (Vol. GIST Technical Report G2002-2, pp. 10 - 18). Glasgow, Scotland: Department of Computing Science, University of Glasgow, Scotland.

Sullivan, C. (2001). Who Cares About CAIR?, *Australia and New Zealand Society of Air Safety Investigators*. Cairns, Queensland.

Suresh, G., et al. (2004). Voluntary Anonymous Reporting of Medical Errors for Neonatal Intensive Care. *Pediatrics, 113*(6), 1609-1618.

Suresh, G., Horbar, J. D., Plsek, P., Gray, J., Edwards, W. H., Shiono, P. H., et al. (2004). Voluntary Anonymous Reporting of Medical Errors for Neonatal Intensive Care. *Pediatrics, 113*, 1609-1618.

Surowiecki, J. (2004). *The Wisdom of Crowds*. New York: Doubleday.

Tamuz, M. (1994). *Developing Organizational Safety Information Systems for Monitoring Potential Dangers.* Paper presented at the Proceedings of the Physical Sciences Annual Meeting II, San Diego.

Tanaka, K. (2002). Learning by Reporting System of Organizational Accidents in Japan. In C. Johnson (Ed.), *Investigation and Reporting of Incidents and Accidents* (Vol. GIST Technical Report G2002-2). Glasgow: Department of Computing Science, University of Glasgow, Scotland.

Taylor, W. A., & Wright, G. H. (2003). A longitudinal study of TQM implementation: Factors influencing success and failure. *Omega: The International Journal of Management Science, 31*, 97-111.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire CT: Graphics Press.

Tufte, E. R. (1997). *Visual Explanations*. Cheshire, CT: Graphics Press.

Van der Schaaf, T. W. (1991). A Framework for Designing Near Miss Management Systems. In T. W. van der Schaaf, Lucas, D.A., and Hale A. R. (Ed.), *Near Miss Reporting as a Safety Tool* (pp. 27 - 34). Oxford: Butterworth-Heinemann.

van der Schaaf, T. W. (1991). Introduction. In T. W. van der Schaaf, D.A. Lucas & A. R. Hale (Eds.), *Near Miss Reporting as a Safety Tool*. Oxford: Butterworth-Heinemann.

Wallace, B., Ross, A., & Davies, J. B. (2003). Applied hermeneutics and qualitative safety data: The CIRAS project. *Human Relations, 56*(5), 587-607.

Wallace, S. J. (2000). Catching Near Hits. *Professional safety, November*, 30-34.

Webber, S. S. (2002). Leadership and Trust Facilitating Cross-Functional Team Success. *Journal of Management Development, 21*(3), 201-214.

Weick, K. E. (2004). Normal accident theory as frame, link and provocation. *Organization and Environment, 17* (1), 27 - 31.

Wikipedia. *Complex system*, from http://en.wikipedia.org/wiki/Complex_system

Wikipedia. *Pareto distribution*, from http://en.wikipedia.org/wiki/Pareto_distribution

Wilcox, M. (2005). Deming, Walter. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 631-636): Elsevier.

Wilson-Donnelly, K. A., Priest, H. A., Salas, E., & Burke, C. S. (2005). The Impact of Organizational Practices on Safety in Manufacturing: A Review and Reappraisal. *Human Factors and Ergonomics in Manufacturing, 15*(2), 135-176.

Womack, J. P., & Jones, D. T. (1996). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*. New York: Simon and Schuster.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *The Machine that Changed the World*. NY: Harper Collins.

Wright, L., & van der Schaaf, T. (2004). Accident versus near miss causation: A critical review of the literature, an empirical test in the UK railway domain, and their implications for other sectors. *Journal of Hazardous Materials, 111*, 105-110.

Yin, Y. K. (1979). *Changing urban bureaucracies: How new practices become routinized*. Lexington: Lexington Books.

Zacharatos, A., Barling, J., & Iverson, R. D. (2005). High-performance work systems and occupational safety. *Journal of Applied Psychology, 90*(1), 77-93.

Zhang, H., Wiegmann, D. A., Thaden, T. L. v., Sharma, G., & MItchell, A. A. (2002). *Safety Culture: A Concept in Chaos?* Paper presented at the 46th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA.

Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology, 85*(4), 587-596.

Zohar, D. (2002). The effects of leadership dimensions, safety climate, and assigned priorities on minor injuries in work groups. *Journal of Organizational Behavior, 23*(1), 75-92.

Zohar, D., & Luria, G. (2005). A multilevel model of safety climate: cross-level relationships between organization and group-level climates. *Journal of Applied Psychology, 90*(4), 616-628.