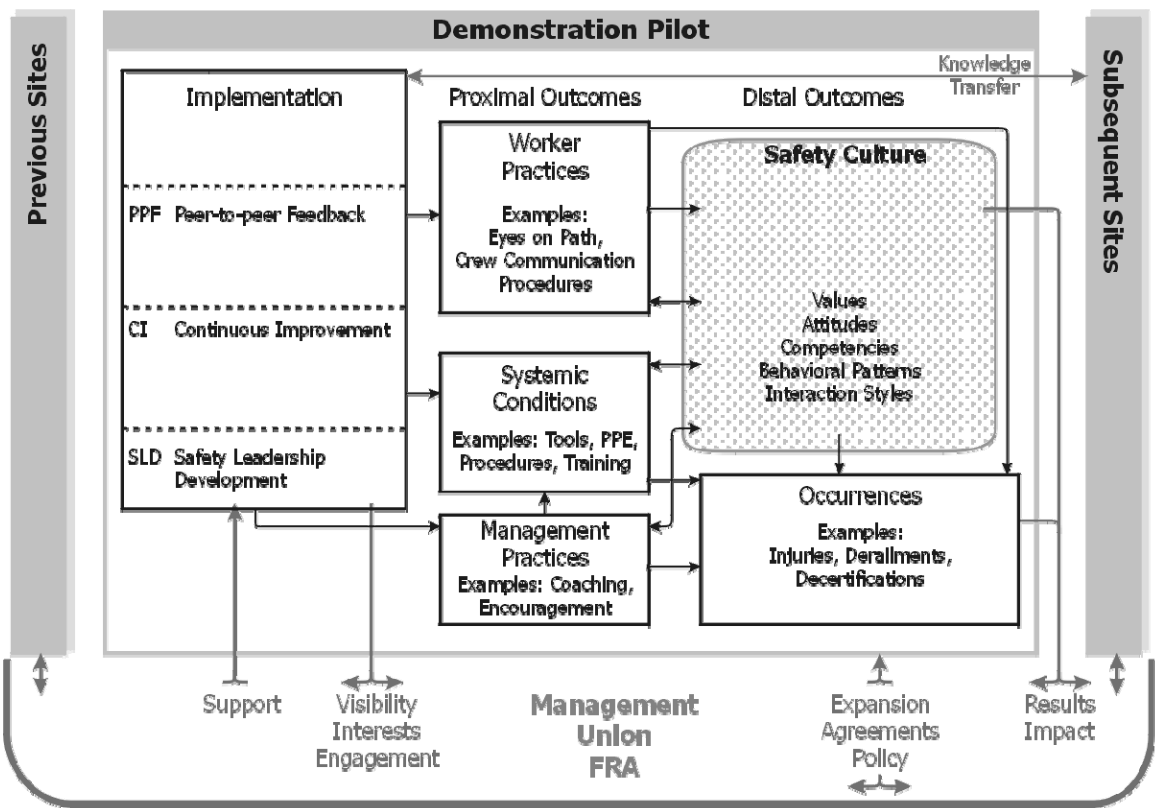
U.S. Department of Transportation

Federal Railroad Administration

Office of Railroad Policy and Development
Washington DC 20590

Transformation of Safety Culture on the San Antonio Service Unit of Union Pacific Railroad



DOT/FRA/ORD-12/16

Final Report
October 2012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2012	Final Report

4. TITLE AND SUBTITLE
Transformation of Safety Culture on the San Antonio Service Unit of Union Pacific Railroad

5. FUNDING NUMBERS
RR04A2 HG384

6. AUTHOR(S)
Michael K. Zuschlag, Joyce M. Ranney, Michael K. Coplen, and Michael A. Harnar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
John A Volpe National Transportation Systems Center
55 Broadway
Cambridge, MA 02142-1093

8. PERFORMING ORGANIZATION REPORT NUMBER
DOT-VNTSC-FRA-10-07

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Department of Transportation
Federal Railroad Administration
Office of Research and Development, Mail Stop 20
1200 New Jersey Avenue, SE
Washington, DC 20590

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DOT/FRA/ORD-12/16

11. SUPPLEMENTARY NOTES
Program Manager: Michael Coplen

12a. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public on the FRA Web site at http://www.fra.dot.gov.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
The Federal Railroad Administration conducted a pilot demonstration of Clear Signal for Action (CSA), a risk reduction process that combines peer-to-peer feedback, continuous improvement, and safety leadership development. An independent formative and summative evaluation of the pilot using qualitative and quantitative measures found that CSA can be implemented on the railroad despite the historical mistrust between labor and management. It is helpful if the site for CSA is open to change and local and external leadership support the process. Over two years, the site with the CSA process experienced improved labor-management relations and an approximately 80 percent decrease in at-risk behaviors. The CSA process was associated with a 79 percent decrease in engineer decertification rates, and an 81 percent decrease in the rate of derailments and other incidents. Comparison locations showed no decreases on these safety measures. Sustaining CSA at a site depends on ongoing cooperation between labor and management to effectively resolve sensitive issues related to the CSA implementation, both protecting the integrity of the worker's process and addressing management concerns. The experience of demonstration suggests that CSA can be effective in promoting a transformation in the broader organization toward more proactive, nondisciplinary approaches to safety.

14. SUBJECT TERMS
Changing At-Risk Behavior, CAB, Cab Red Zone, CRZ, Continuous Improvement, Implementation Outcomes, Initial Context, Peer-to-Peer Feedback, Program Evaluation, Road Safety, Safety Culture, Safety Leadership, Sustainability, Switching, Transformation

15. NUMBER OF PAGES
211

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in)	= 2.5 centimeters (cm)
1 foot (ft)	= 30 centimeters (cm)
1 yard (yd)	= 0.9 meter (m)
1 mile (mi)	= 1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2)	= 6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2)	= 0.09 square meter (m^2)
1 square yard (sq yd, yd^2)	= 0.8 square meter (m^2)
1 square mile (sq mi, mi^2)	= 2.6 square kilometers (km^2)
1 acre = 0.4 hectare (he)	= 4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz)	= 28 grams (gm)
1 pound (lb)	= 0.45 kilogram (kg)
1 short ton = 2,000 pounds (lb)	= 0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp)	= 5 milliliters (ml)
1 tablespoon (tbsp)	= 15 milliliters (ml)
1 fluid ounce (fl oz)	= 30 milliliters (ml)
1 cup (c)	= 0.24 liter (l)
1 pint (pt)	= 0.47 liter (l)
1 quart (qt)	= 0.96 liter (l)
1 gallon (gal)	= 3.8 liters (l)
1 cubic foot (cu ft, ft^3)	= 0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3)	= 0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

[(x-32)(5/9)] °F	= y °C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm)	= 0.04 inch (in)
1 centimeter (cm)	= 0.4 inch (in)
1 meter (m)	= 3.3 feet (ft)
1 meter (m)	= 1.1 yards (yd)
1 kilometer (km)	= 0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2)	= 0.16 square inch (sq in, in^2)
1 square meter (m^2)	= 1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2)	= 0.4 square mile (sq mi, mi^2)
10,000 square meters (m^2)	= 1 hectare (ha) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gm)	= 0.036 ounce (oz)
1 kilogram (kg)	= 2.2 pounds (lb)
1 tonne (t)	= 1,000 kilograms (kg)
	= 1.1 short tons

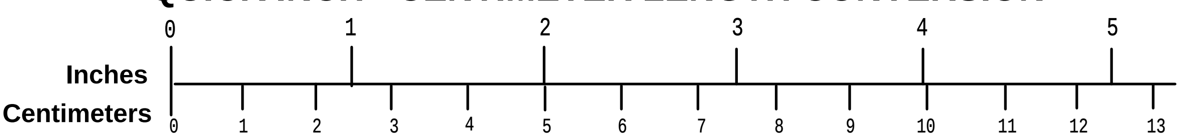
VOLUME (APPROXIMATE)

1 milliliter (ml)	= 0.03 fluid ounce (fl oz)
1 liter (l)	= 2.1 pints (pt)
1 liter (l)	= 1.06 quarts (qt)
1 liter (l)	= 0.26 gallon (gal)
1 cubic meter (m^3)	= 36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3)	= 1.3 cubic yards (cu yd, yd^3)

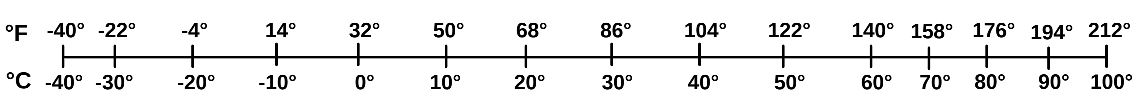
TEMPERATURE (EXACT)

[(9/5) y + 32] °C	= x °F

QUICK INCH - CENTIMETER LENGTH CONVERSION



QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSION



For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50 SD Catalog No. C13 10286

Updated 6/17/98

Acknowledgments

This evaluation would not have been possible without the cooperation of many managers and employees at Union Pacific Railroad.

The authors would especially like to thank Brian Gorton, Michael Byars, and Kelvin Phillips for their considerable assistance, along with Lance Fritz, Joe Santamaria, Roby Brown, Michael Mitchell, Mark Barnum, Ted Lewis, Shane Keller, Ronald Tindall, Greg Burger, John Dunn, Russell Elley, Mike Araujo, Paul Dillon, Carl Eddington, Jose Gutierez, Wil Hardiman, Chad Jistel, Oscar "Doctor" Mayfield, Fernando Nanez, Pat Pino, Martin Vacca, Mario Valadez, John Paul Schuster, and Andy Wright. Thanks also to George Wollard and Jay Finney of Behavioral Science Technology, Inc., for providing education and insights into implementing methods like Clear Signal for Action in the railroad industry. Along with Kelly Johnson and others at BST, George also gathered data for us from a survey customized to our requirements. Jonny Morell from NewVectors and Demetra Collia from the Bureau of Transportation Statistics provided additional technical assistance. Special thanks to Wayne Nelson for instructing and advising on gap time analysis and survival analysis. Shuang Wu of Computer Sciences Corporation assisted in processing and analyzing the data. Katherine Blythe and the staff of MacroSys provided editorial assistance.

The work is being performed under an interagency agreement between the Federal Railroad Administration's Human Factors Research & Development Program and the John A. Volpe National Transportation Systems Center's Human Factors Division.

Contents

Figures

Tables

Executive Summary

A Need for New Safety Approaches

Despite continuous efforts by the Federal Railroad Administration (FRA), industry, and unions, accident rates in the railroad industry have remained flat since 1986 (FRA, 2001, 2008). After declining between 1976 and 1986, accidents, injuries, and fatalities have held at a steady rate for nearly 25 years, indicating that the current system has reached its limit and that further advances in safety will require new approaches.

Clear Signal for Action for Improving Safety and Culture

FRA is evaluating potential risk reduction methods that focus on the unsafe conditions and practices that might lead to events, rather than taking the traditional approach of focusing on the events themselves. These new methods have already proved effective in other industries and FRA is optimistic they can be successfully tailored to the railroad industry.

Clear Signal for Action (CSA) is one such risk reduction method. CSA integrates three approaches that have been previously applied in other industries to improve safety proactively: Peer-to-peer Feedback, Continuous Improvement, and Safety Leadership Development.

- *Peer-to-peer Feedback (PPF).* A steering committee develops a site-specific checklist of safe and at-risk behaviors and conditions contributing to injuries. Employees then use the checklist to observe their anonymous peers' safety behavior, confidentially provide coaching feedback to the peers on safety, and gather feedback from the peers for improving conditions.
- *Continuous Improvement (CI).* Within a safety context, CI establishes data collection methods to identify systemic causes of observed at-risk behaviors and conditions (in this case, completed PPF checklists with feedback from workers), then implements corrective actions to address the causes. At-risk behaviors include those behaviors the worker can control. Potential systemic causes include organizational policy, training, tool design, environmental conditions, procedures, and cultural characteristics. Data collection continues after a corrective action is taken to allow its effectiveness to be evaluated.
- *Safety Leadership Development (SLD).* SLD provides management training to promote proactive safety practices, such as PPF and CI.

Changing At-Risk Behavior: A CSA Demonstration Pilot at Union Pacific

Between 2005 and 2008, FRA sponsored a CSA demonstration pilot in the transportation department at the San Antonio Service Unit (SASU) of Union Pacific Railroad (UP).

Behavioral Science Technology, Inc. (BST), a company that has implemented CSA-like programs in a broad range of industries, instructed and advised the implementation of the CSA pilot called Changing At-risk Behavior (CAB). Local management and workers, along with their unions (Brotherhood of Locomotive Engineers and Trainmen, and the United Transportation Union), worked together to implement CAB.

The CAB process began in August 2005 with the initiation of regular peer-to-peer feedback sessions. CAB initially focused on behaviors to improve alertness and teamwork among locomotive cab personnel on the road. This was limited to practices performed when operating under constraining signals, a situation that UP calls Cab Red Zone (CRZ), for which there are specific rules in the *General Code of Operating Rules*. Fourteen months after its origination, CAB expanded to include safety in yard-switching operations.

A Comprehensive Evaluation of CAB

FRA also sponsored an independent evaluation of CAB to address the following questions:

- *Implementation*. What are the characteristics of a CSA implementation that allow its activities to be performed as planned in the railroad industry? How should the CSA process be improved to more effectively establish it in railroad settings?
- *Outcomes.* What are the effects of the CSA process on safety and safety culture?
- *Sustainability.* What characteristics of the implementation and organization (e.g., economic and political) are required to sustain CSA once it has been successfully implemented?
- *Transformation.* How does the CSA process influence the broader organizational environment of the pilot CSA demonstration, if at all? Specifically, does the process make the stakeholder organizations more receptive to a proactive, nondisciplinary approach to safety, encouraging the spread of CSA-like processes to other parts of the organization or industry?

In addition to answering the above questions, the evaluation also gathers information on the *context* of the organization and pilot site that affects their inclination toward attempting a CSA implementation.

The evaluation questions were answered by assessing implementation activities, manager and worker practices, work conditions, safety culture, occurrences, and transformational influence on the broader organizations. The evaluation was both formative and summative because the findings informed both implementation as it developed and the conclusions in this present report on CAB's successes and shortcomings.

The evaluation assessed CSA in both semidispersed (rail yard work) and dispersed (road) work environments. In the absence of strict experimental controls, the evaluation used a multimeasure, mixed-method (qualitative and quantitative), quasi-experimental design.

Qualitative measures included:

- Evaluating team members' field notes and project records such as reports, news releases, and newsletters.
- Periodic interviews with workers and managers.

Quantitative measures included:

- CAB process metrics, including the percent of the targeted workforce that completed training for conducting peer feedback sessions, and the number of feedback sessions conducted.
- Statistics from the PPF feedback checklists.
- Corporate safety data such as engineer decertifications and incidents (mostly derailments).
- Close-ended attitude and behavior surveys of workers and managers, administered in 2005 and 2008 to track changes.

This use of multiple measures and data sources enabled confirmation of any observed effects. Comparison data were used to isolate factors not associated with the CAB process to verify that any observed effects were likely due to CAB and not to other service unit or regional efforts.

The evaluation in this report was subjected to two independent external reviews. First, a national expert on the analysis of time-to-event data reviewed the advanced statistical procedures used for the safety data in this evaluation, and found the analytical procedures to be suitable. Secondly, in the course of receiving the American Evaluation Association's (AEA) 2011 Outstanding Evaluation Award, evaluation experts appraised the evaluation against documented standards.

Evaluation Findings

Context

A site is most suitable for a CSA implementation if it has the following characteristics:

- Openness to actions that are inconsistent with the organizational culture, which facilitates the establishment of CSA and cultural change in general.
- Local leadership and external support in favor of promoting CSA facilitates its establishment.

Before CAB was introduced at SASU, a series of fatal accidents led workers and managers to question their traditional approaches to safety. At about the same time, a large number of new workers joined the service unit. Lacking years of indoctrination in the traditional railroad culture, these workers were inclined to try nontraditional approaches. New union and management leadership united to provide direction towards alternative, cooperative approaches—including CSA.

CSA Successfully Implemented on Railroad

CSA can be successfully implemented within railroad departments despite the prevailing railroad culture and other factors (such as dispersed work settings) that pose special challenges for CSA. Implementation of the pilot demonstration was not entirely completed as planned, but it was sufficient to justify an evaluation of outcomes.

Evidence of successful implementation includes:

- Training of workers was completed on or ahead of schedule.
- Over 50 percent of workers were trained.

- On average, each road worker experienced a feedback session once every seven months and yard workers every 2–3 months. This rate was below the initial target rate of once per month but was considered sufficient since each employee was very likely to repeatedly experience feedback sessions.
- The steering committee analyzed observation data regularly and identified barriers.
- Managers supported CAB efforts by removing barriers and attending training sessions.

On the basis of SASU's experience with CAB, a successful implementation depends on:

- Overcoming worker resistance.
- Obtaining management commitment.
- Acquiring a sufficiently high participation rate in feedback sessions.
- Identifying and implementing visible, relevant corrective actions.

Overall, the CAB demonstration pilot achieved a fairly high level of implementation. Although there were both anticipated and unanticipated barriers, CAB largely overcame these through a combination of:

- Training.
- Worker control over and anonymity of the feedback sessions.
- Management's willingness to provide the necessary resources even when budgets were tight.
- Program modifications based on evaluation feedback.

CSA Improves Safety and Safety Culture

Process metrics, surveys, interviews, and corporate safety data all point toward safety-related improvements from implementing CSA:

- Worker practices targeted by the CSA process improved with the program's introduction. Process metrics indicated that the commission of at-risk behaviors decreased approximately 80 percent.
- Management promoted the CSA safety process and improved their coaching of workers in field operations tests.
- Systemic safety conditions improved when management and workers removed barriers to safety.
- The rate of engineer decertification gradually declined by 79 percent after the introduction of CSA. Decertification rates at comparisons sites did not decline.
- The rate of derailments and other incidents decreased by 81 percent at a yard with a strong CSA implementation. This improvement was more than that observed at the location with a moderate implementation, which showed a 32-percent decrease, and the yards with a weak implementation, which showed no significant change.

- Labor-management relations improved at the demonstration site, with greater trust and cooperation between workers and managers.
- General safety awareness, personal responsibility for safety, and safety dialogue improved among demonstration site personnel.

In summary, CAB has demonstrated the potential for CSA processes to improve safety and safety culture in the railroad industry.

CSA Appears Sustainable on the Railroad

The ability of CSA to sustain itself for the longer term is a result of genuine cooperation between labor and management. Even though, in the short term, labor appears to be driving the process, the long-term view is that CSA cannot succeed without management's support. This requires ongoing cooperation between labor and management to effectively resolve sensitive issues related to CSA implementation, by both protecting the integrity of the worker's process and addressing management concerns.

As a result of the CAB pilot, labor and management stakeholders have enhanced sustainability through:

- Negotiating a budget and work hours plan that balanced value with cost.
- Establishing facilitator and member rotation schedules.
- Integrating elements of other safety programs.

CSA Can Transform the Railroad

The outcome of the CAB demonstration pilot suggests that CSA can be effective in promoting a transformation in the broader organization toward more proactive, nondisciplinary approaches to safety. Transformation resulted from labor and management cooperating to make CAB effective, and communicating its successes to important decision-makers.

Promoting CSA in the Railroad Industry

Evaluation results show that CSA is capable of improving safety culture and safety performance within transportation departments. Furthermore, the railroad industry and railroad workers would benefit from CSA processes at sites nationwide.

FRA, the rail industry, and the unions can promote complete, sustainable, and transforming implementations of CSA.

- FRA can promote CSA by sponsoring implementations and publicizing the results. It could also form a national working group, made up of representatives from FRA, labor, management, and other entities, whose purpose would be expanding the use of CSA in transportation departments, and introducing it to other departments and trades in the rail industry.
- Industry leaders can promote CSA within their respective railroads by implementing the process initially at small selective sites, then by expanding it to more challenging sites. At each site, an industry leader should recruit managers,

union representatives, and workers, ensure there are sufficient resources and data protection, and publicize progress.

- National and local union leadership can promote CSA by identifying and preparing workers to cooperate with management on CSA, and by contacting railroad union leaders at other CSA implementations.

The evaluation of the CSA process at SASU shows that, although implementing CSA is not without its challenges and costs, widespread adoption of CSA can substantially reduce the rate of accidents, injuries and fatalities, improve the railroad safety culture, and advance the level of safety within the railroad industry.

1. Prologue

The following fatal accident, as detailed by the NTSB (2006), was the seminal event that spurred a pilot safety process at Union Pacific that this report documents, a process that reduced the risk of such accidents and ultimately transformed safety culture.

In the early morning of June 28, 2004, the 74-car Union Pacific (UP) Train MHOTU-23, rolling at 44 mph, passed a red signal just east of the small community of Macdona, TX. Immediately ahead, the 123 cars of BNSF Railway Train MEAP-TUL-126-D were pulling off the main line into a siding. The UP train, its throttle remaining open at the Number 3 notch, struck the midpoint of the BNSF train.

The impact derailed all four 200-ton UP locomotives. Nineteen additional UP cars piled up against the locomotives, some of them breaking open and spilling their contents. Among them was a tank car that sustained an 11-inch rip from which escaped over 50 tons of liquid chlorine that instantly vaporized forming a cloud of poisonous gas 1,400 feet across that engulfed the UP locomotives and two homes near the tracks.

Lying under debris but not seriously hurt, the crew of the UP train smelled the chlorine inundating their cab and extricated themselves from the wrecked locomotive to escape the gas. As they attempted to walk toward clean air, the conductor's breathing became increasingly difficult and he could not continue. The engineer, himself suffering from respiratory distress, found that he could not carry him. The conductor's body was found a few hours later by firefighters. Two people in one of the homes nearby were also found dead. At least 30 others were injured (NTSB, 2006).

2. Motivation

2.1 Context: The Problem of Railroad Safety and Culture

Despite continued efforts by the Federal Railroad Administration (FRA), industry, and unions, accident rates have remained relatively flat over two decades (FRA, 2001, 2008). Accidents, injuries, and fatalities continue to occur unabated. As shown in Figure 1, even though the industry made significant progress in reducing accidents between 1978 and 1986, since 1987 accident rates have essentially flattened, not significantly worsening or improving.

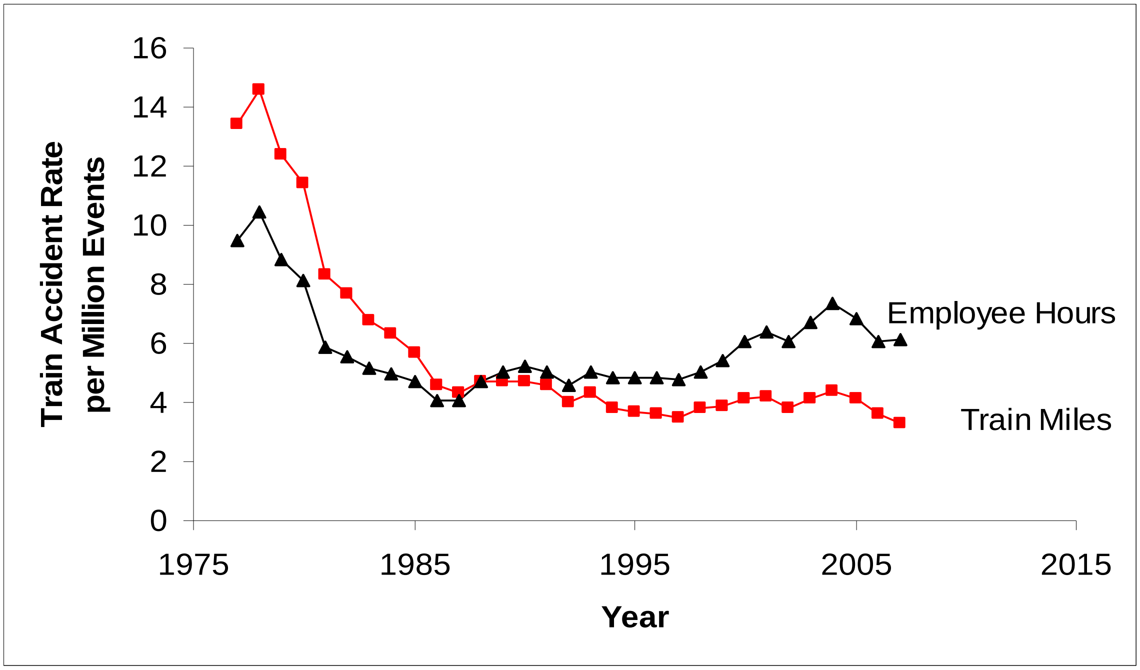


Figure 1. Train Accident Rate by Train Miles and Employee Hours (from FRA (2008))

This stagnation suggests that further advances in safety will require approaches other than those traditionally pursued by the railroad industry. Although these approaches can be credited with the railroad industry's current high degree of safety (Gamst, 1982), further analysis suggests that these same traditional safety approaches, and the overall railroad safety culture in which they operate, may be limiting further advances (Ranney and Nelson, 2003). Safety improvements must therefore not only include alternative approaches to safety but also introduce changes to the safety culture itself. Ideally, a new safety process can be the vehicle to initiate change in the culture.

Before a new safety process can be established and effective, it must first contend with the following characteristics of railroad safety and culture:

- A command-and-control management style.
- A tendency toward reactive operations.

- Dated safety analysis orientation.
- Poor labor–management relations.

2.1.1 Command-and-Control Management Style

Considered one of the first large-scale businesses of the industrial revolution, the railroad industry modeled its organizational structure on the only other large organization of the time—the military. At some railroads, such as Union Pacific, top managers were often former army officers (Gamst, 1982; Ambrose, 2000). This 19th-century military model promoted a top-down, authoritarian, command-and-control management style that has persisted into the 21st century. Typical of this management style, managers tend not to elicit input from workers, but to instead issue orders, not always with explanations, that workers are expected to follow. Compliance is achieved primarily through the threat of discipline (Gamst, 1982). Although a discipline system is perhaps essential for maintaining safety, a complementary process to encourage safe behavior beyond the rules would promote even greater safety.

2.1.2 Reactive Operations

The operation and management of railroads tends to be reactive. Normal operations are well established, given the nearly 200 years of railroad history, but the railroads themselves are tightly coupled systems that have long been susceptible to unpredictable events (Armstrong, 1982). Capacity is constrained by the high cost of laying new track, which may be especially expensive for accessing industries and intermodal facilities in urban areas. Congestion and bottlenecks are often the result. Weather changes, errors in scheduling, and mechanical failures (perhaps as simple as a broken knuckle on a car) can trigger ripples through the system and profoundly affect productivity and profitability.

These conditions tend to encourage reactive operations. Planning for the future is discouraged when occurrences in the present make the future uncertain. Effective railroad administration requires quick and decisive action. The approach is often to make do with whatever is available to get the job done in response to the latest unpredicted event. The day (or night) of a railroad manager is characterized by battling one crisis after another, charging into each situation, taking control, and issuing commands before moving to the next crisis.

Workers likewise operate under uncertainty, responding to the reactive decisions of management. Road workers, who make up the majority of the workforce, experience particularly high levels of unpredictability (Gamst, 1982). In a typical railroad company, road workers have no fixed schedule and instead are allocated to a rotating pool of workers, and assigned to trains as they become available. Because train schedules are unpredictable, so is the road crew's schedule. It is common for crews to have only a few hours' notice before reporting for duty.

On the road, they contend with unpredictable orders from dispatch, which is itself trying to manage congestion. The train may be diverted into sidings at any time to wait for another train passing in the opposite direction. Consistent with a command-and-control management style, crews are generally not informed of dispatch's intent (e.g., how long to expect to remain in a siding). At the end of the trip, a train may arrive within sight of the terminal only to be held for hours just outside while tracks are cleared to accommodate it.

Such an environment encourages workers to also be reactive: they are predisposed to confront sudden changes with quick action, but also disinclined to anticipate what those changes may be.

2.1.3 Limiting Safety Analysis Orientation

Railroad culture remains attached to two older orientations toward safety analysis in which human error is the primary cause of accidents and injuries rather than a symptom induced by other underlying factors (Hale, 2000). According to both these views, human workers on the front line contribute variability to the work operations, resulting in accidents and injuries. However, the two orientations differ in their approach to reducing the effects of human variability on operations.

The first safety analysis orientation is *technological*. It provides safety through design of the work environment and attempts to insulate human variability from dangerous conditions using devices and structures (e.g., personal protection equipment, air brakes, signal systems, positive train control). Human error is assumed to be unavoidable and technology is often seen as replacing or protecting the variable, error-prone tendencies of humans with the deterministic, non-error-prone tendencies of technology.

The second orientation is *procedural*. It provides safety through design of work practices and attempts to reduce human variability by creating rules of behavior that workers consistently follow. Human error, including failure to conform to rules, is assumed to be controllable by sufficiently motivating the workers. Consistent with the railroad command-and-control management style, discipline is used to enforce conformance with the rules.

Neither of these orientations is consistent with a more contemporary *systems* safety analysis orientation in which safety is regarded as a product of the organizational system. Accidents and injuries result not only from front-line employee behaviors and technology, but also from latent "upstream" organizational processes that manage the employees and introduce the technology. Safety is improved not necessarily through additional technology or rules, but through continuous learning and adaptation by the organization.

Adopting this orientation has resulted in safety improvements in other industries such as aviation (Reason, 1997). To be effective, the systemic orientation requires open communication between managers and workers so they can share the information necessary for performing root-cause analyses that identify the sometimes subtle organizational causes of injuries and accidents.

2.1.4 Adversarial Labor–Management Relations

In the railroad industry, labor unions, made up of highly trained workers who cannot be easily replaced, have emerged as a powerful, well-organized national force. Therefore, a union's threat of a strike carries substantial weight. This leverage evolved from the bitter and sometimes violent confrontations between management and labor dating back to the 19th century. The fundamental conflict that persists in railroad labor–management relations to this day involves management efforts to extract the greatest productivity for the least cost, pitted against union attempts to acquire the greatest employee benefits irrespective of cost. Among these benefits are safety improvements to working conditions.

With such competing agendas, labor and management tend to select opposing orientations towards safety. Management is more inclined to take a procedural orientation, since enforcing rules or creating new ones is faster and cheaper than procuring new technology or rebuilding the physical work environment. Labor selects more of a technology orientation to promote better working conditions and protect workers from discipline. The net effect is a "blame culture" of safety with each side holding the other responsible for accidents and injuries, and neither side working to acquire a deeper system-based understanding of safety.

The conflicting orientations on safety are aggravated by the litigious environment surrounding injuries in the railroad industry. The 100-year-old Federal Employer's Liability Act (FELA) allows railroad employees to sue their employer for damages if they are injured. This creates a strong financial incentive for each side to directly blame the other to protect their immediate legal interests, potentially at the expense of improving safety. Workers are likely to regard accidents as the result of company policies and workplace conditions, while downplaying the role of their own decisions and behaviors. Management is likely to do the opposite, and may use its disciplinary power to stress the apparent role of worker behavior in safety (e.g., by creating safety rules that are so numerous and general that nearly any incident can be attributed to a rules violation).

Labor and management each hold an incomplete perspective on the causes of accidents. This interferes with information exchange and with cooperation on safety improvements. Cooperation on new safety processes can be threatened by the adversarial relations between labor and management when one side uses the process to gain an advantage over the other.

2.1.5 Rotating Management

The lifetime of a new safety process can be truncated by the rotation of managers. In the railroad industry high-level managers at a service unit or region traditionally are rotated approximately every 18 months. New managers may have different agendas than previous managers. The outcomes of safety initiatives may take years to be detected because the random nature of accidents and injuries requires a large sample of time to identify reliable statistical changes.

The effect of rotating management can result in discontinued safety processes at a location before their effects can be known. Sometimes the colloquial expression, "flavor of the month," is used to refer to these safety programs after the ice cream store, Baskin-Robbins, which used to showcase one of their 31 flavors each month. Not only is this pattern likely to kill potentially effective safety processes, it also breeds cynicism among lower managers and workers about the permanence of a safety program. They are unlikely to dedicate themselves to a new process that seems unlikely to last.

2.1.6 Net Effect of Railroad Industry Culture on Safety

Combine a command-and-control management style, reactive tendencies, rotating management, outdated orientations of safety split along labor/management lines and motivated by the legal environment, and the result is a safety treadmill that may account for a static safety record. Consistent with a reactive management style, little attempt is being made to proactively identify features or trends that may become factors in future accidents or injuries.

Labor and management react to each injury or accident when it happens. Management reacts by promoting rules violations as the cause, whereas labor reacts by promoting workplace deficiencies as the cause. Potential or realized legal battles stifle the information sharing necessary to identify potential upstream organizational causes. Any introduced safety processes tend to be discontinued when new management rotates in and before they can have an appreciable effect.

Labor assumes that safety is achieved by engineering away dangers by adding layers of technology and maintenance to address each risk. Management assumes that safety is achieved by incrementally coercing the worker to be safer. If the cause of an accident is judged to be technological (e.g., a physical failure), the physical environment is repaired. If the cause is judged to be procedural, the workers are punished for breaking the rules. If there are no technology failures and no violation of written rules, then new rules are created to address the situation and discipline prepared to enforce them (Gamst, 1982).

Attempting to engineer away danger has its limits, especially if it disregards the larger operational and organizational context. Complex technological safety systems, deployed separately without consideration of each other, can interact creating unanticipated chains of events and producing accidents that otherwise would not occur (Perrow, 1999).

Attempting to coerce workers to be safer likewise has had a limited impact on actual safety (Elling, 1991). Workers may avoid safety disclosures to management because they fear disciplinary action (Baram, 1997), even though their knowledge of the job best equips them to recognize safety risks. As a result workers are dissuaded from sending information to the organization relating to safety deficiencies that could prompt safety improvements.

To improve safety in the railroad industry, a new process should have the following attributes to avoid the limitations imposed by current industry safety culture:

- *Nondisciplinary*.
- *Proactive*, including the collection of upstream data for predicting safety.
- *Systems safety analysis orientation*, objectively analyzing data to derive risk-reducing corrective actions.
- *Cooperative*, engaging stakeholders within both management and labor.
- *Sustainable*, including mechanisms for long-term sustainment in the face of rotating management.

In addition, the new safety process should be established within the current railroad safety culture. Given the overall effectiveness of the industry's traditional safety approach, any new process should *improve* it, rather than *replace* it, raising safety to a new level.

However, some aspects of the traditional railroad culture are incompatible with the attributes required of a new safety process. The safety culture itself must be transformed to share the same attributes as the new safety process while still preserving existing procedures known to be effective.

2.2 CSA Concept

One of the goals of the FRA Office of Railroad Policy and Development is to improve safety throughout the railroad industry by removing barriers or limits associated with the traditional railroad safety culture. This can be done by demonstrating and evaluating alternative systems-oriented approaches to risk reduction that result in a cultural change in tandem with safety performance improvements (FRA, 2008). If such innovative approaches demonstrate success, FRA will encourage their use and adoption throughout the industry. One current example is the CSA process.

2.2.1 CSA Components

CSA integrates three approaches that have been applied previously to proactively improve safety: Peer-to-peer Feedback (Krause, 1995), Continuous Improvement (Juran, 1964; Harrington, 1987; Krause, 1995), and Safety Leadership Development (Krause, Seymour, and Sloat, 1999).

- *Peer-to-peer Feedback (PPF).* PPF uses behavioral analysis methods to identify and address at-risk behaviors before they cause injuries by using peer-to-peer observations of, and two-way feedback about work behavior, conditions, and organizational factors.
- *Continuous Improvement (CI).* In a safety context, CI establishes a means of gathering data to identify systemic causes of observed at-risk behaviors and conditions, and then implements corrective actions to address the causes. Potential systemic causes include organizational policy, training, tool design, environmental conditions, procedures, and cultural aspects. Data gathering continues after a corrective action is taken, allowing time to evaluate its effectiveness.
- *Safety Leadership Development (SLD).* SLD provides management training to promote proactive safety practices, such as PPF and CI.

When used alone, PPF-type approaches have often placed too little emphasis on the influence of upstream managers, systems, and behaviors on at-risk behavior and conditions. As a result, several unions have given early PPF approaches negative reviews (Spigener and Hodson, 1997; Howe, 1999; Frederick and Lessin, 2000). Thus, recent variants of PPF have integrated CI, utilizing PPF's feedback sessions as opportunities to collect the data needed by CI.

Although PPF and CI tend to emphasize the impacts of behaviors and conditions immediately present at the potential accident site, SLD targets latent factors in accidents that are further back in the chain of causation, such as safety climate and culture

(Reason, 1997). SLD trains organizational leadership to eliminate these causes since it has the resources and authority to alter the direction of the organization (Boyd, 2008). SLD can therefore accelerate changes initiated by PPF and CI and make them a lasting feature of the organization's safety culture.

By combining PPF, CI, and SLD, the responsibility for safety is distributed among workers and managers. PPF activities are predominantly or fully the workers' responsibility since peers make observations and provide feedback. SLD changes organizational leadership and is fully the managers' responsibility. CI is a joint responsibility, but since management is accountable for ensuring the corrective action process is effective, they have more responsibility than workers (Walton 1986; Deming 2000).

2.2.2 CSA Implementation Process

Figure 2 illustrates the theory of action for CSA. Detailed CSA activities are in the Implementation box on the left and its theoretical outcomes are on the right. The arrows indicate the effect that prior activities or outcomes have on subsequent activities and outcomes, with influence moving primarily in a left to right direction. Outcomes are grouped into columns representing proximal and distal outcomes. Proximal outcomes result directly from the implementation activities, whereas distal outcomes are mediated by proximal outcomes. Within the Implementation box, activities are grouped according to their primary association with PPF, CI, or SLD.

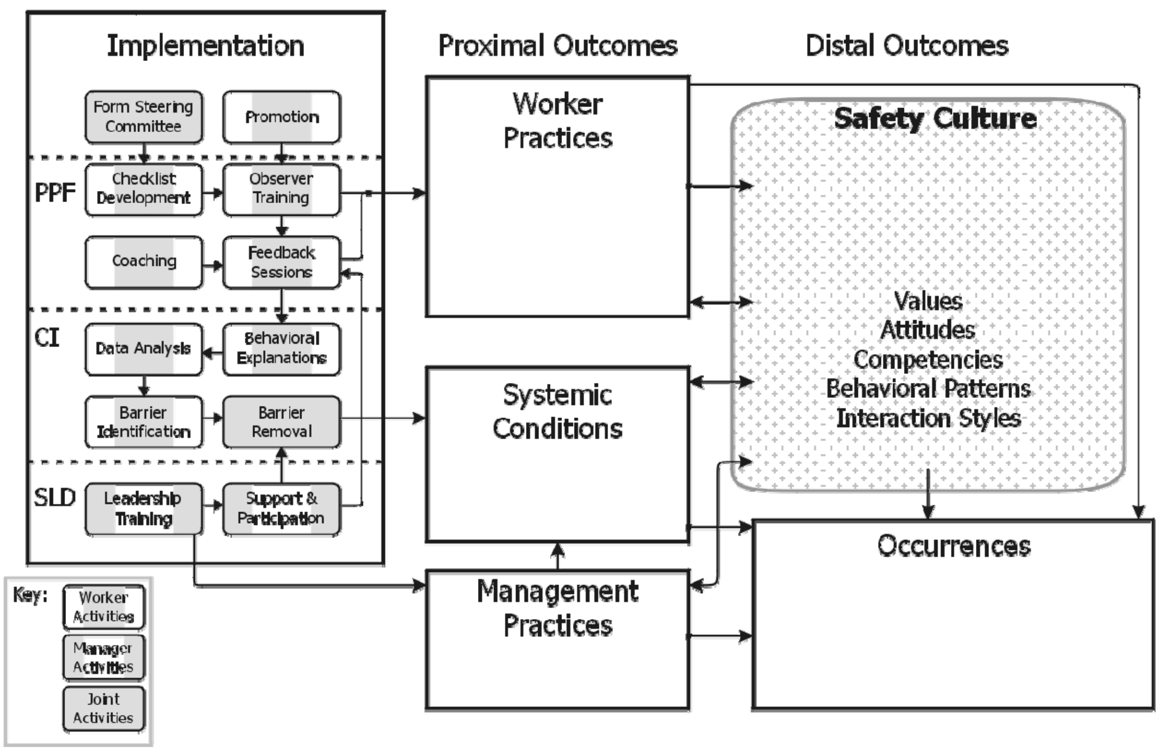


Figure 2. Theory of Action for CSA, Showing Activities and Theoretical Outcomes

To initiate CSA, management must first make a commitment to provide the necessary resources to establish the process. This step is consistent with the organizational performance literature that identifies managers as safety climate engineers (Simard and Marchand, 1994, 1997). This step is shown in the upper left of the Implementation box, where management contacts labor and forms a steering committee made up of workers (and sometimes several managers) to develop and direct the process. Employee involvement is thought to be essential in generating buy-in and ensuring that the process is self-sustaining.

Within the PPF component, the steering committee develops a checklist of safe and at-risk worker behaviors and working conditions based on analyses of accident reports and other sources of safety information (Krause, 1997). The steering committee recruits, trains, and coaches workers in conducting nonconfrontational feedback sessions of their coworkers. The trained workers then use the checklist to observe their peers' safety behavior, confidentially provide coaching feedback to the peers on safety, and gather feedback from the peers for improving conditions.

Within the CI component, workers are trained to interview their coworkers during the feedback sessions about any observed at-risk behaviors or conditions. The observed behaviors and conditions and coworker reasoning for the at-risk behaviors or conditions are documented, aggregated, and analyzed by the steering committee via root-cause problem-solving to identify barriers to enhancing safety. The steering committee executes corrective actions on barriers it can remove (e.g., through feedback to workers during observations). If the barriers require actions beyond the authority of the steering committee (e.g., purchasing new equipment or changing procedures), a joint labor–management Barrier Removal Team then reviews the barriers, prioritizes them, and develops corrective actions, which management then executes.

Within the SLD component, managers are trained to use effective, nondisciplinary, proactive techniques to help personnel work safely, including, but not limited to, supporting safety-related activities such as feedback sessions and barrier removal. These SLD processes are conducted along with existing disciplinary processes. SLD activities are not a substitute for addressing rules violations.

In summary, CSA has the following attributes that should allow it to avoid the limitations of the railroad industry safety culture:

- *Nondisciplinary.* Observations of worker behavior are nonpunitive, anonymous, and conducted by nonthreatening peers.
- *Proactive.* Data is collected on at-risk behaviors and conditions to prevent associated accidents or injuries before they occur.
- *Systems safety analysis oriented.* Objective data analysis identifies underlying organizational factors in safety.
- *Cooperative.* Management and labor each have responsibilities in the process and work together on selecting and deploying corrective actions.
- *Sustainable.* Further data collection assesses the effectiveness of the corrective actions, providing early and objective documentation of process successes, which may encourage support from successive management leaders.

2.2.3 Theoretical Impacts on Safety Performance and Culture

The Theory of Action for CSA in Figure 2 specifies improvements in the following outcomes, represented by boxes to the right of the Implementation box:

- *Worker Practices*. The at-risk and safe behaviors identified in the PPF checklist and targeted by the feedback sessions. Examples include walking with eyes on the path, use of personal protective equipment, and following crew safety communication procedures.
- *Systemic Conditions.* Aspects of the physical and organizational environment that impact safety, such as facilities, tools, training, and disciplinary policies. Examples include switches that are in good operating condition, availability of proper tools and personal protective equipment, and training on safe worker practices, such as the proper way to align a coupler.
- *Management Practices:* The on-the-job manager behaviors that take the form of proactive nondisciplinary promotion of safe behaviors, conditions, and processes. These practices are in addition to implementation activities listed above. Examples include coaching employees on safe behavior and encouraging participation in safety processes.
- *Occurrences.* The safety events associated with injuries, fatalities, damage to equipment (e.g., from derailments), and related close calls.
- *Safety Culture.* A broad amalgamation of factors that represents an organization's vigilance, reliability, and effectiveness for ensuring safety, including safety-related norms, competencies, beliefs, values, attitudes, relations, practices, and interaction patterns (Reason, 1997).

As depicted by the arrows in Figure 2, the implementation directly promotes the targeted safety practices of both managers and workers through its SLD and PPF components respectively. Workers' practices change from both the feedback sessions and the training for performing feedback sessions (Ranney, Zuschlag, Coplen, and Nelson, in preparation). The CI process directly improves systemic conditions.

There are reciprocal effects between the specific practices and safety culture. Changes in safety attitudes encourage workers and managers to change their customary practices. Likewise, changing practices may change attitudes (Krause, 1995). Safer practices also improve the culture by building better labor–manager relations. As workers encourage each other to work more safely, management trusts workers to perform tasks safely and perceives less need to discipline them. Concurrently, as managers engage in more proactive nondisciplinary approaches to safety, workers are assured of management's commitment to worker safety. The result is that as unfair disciplinary actions become less frequent and management commitment to safety becomes more apparent, workers trust managers more.

Improved systemic conditions also improve labor–management relations within the safety culture by demonstrating management's commitment to safety. An improved safety culture can reciprocally improve systemic conditions by fostering labor–management and cross-trade cooperation to improve conditions.

Safer practices and improved systemic conditions reduce occurrences, as does the enhanced safety culture through increased dialogue about safety. Collectively these improve worker and manager safety practices beyond those specifically targeted by the implementation.

In summary, CSA improves safety culture simultaneously with safety performance through a chain of feedback loops.

2.2.4 CSA Effectiveness

2.2.4.1 Other Industries

Previous research provides evidence that observations and feedback on the safety of behavior can be very effective at reducing injury rates and improving productivity. For examples, see Sulzer-Azaroff and Austin (2000), Guastello (1993); Kopelman (1986), and Krause, Seymour, and Sloat (1999).

Although apparently there is no documented evidence on the effectiveness of CI alone, Krause, Seymour, and Sloat (1999) provide a meta-analysis of 73 sites showing positive safety effects of CI combined with PPF and SLD. There is also indirect evidence from the meta-analysis of Guastello (1993), which examined a number of interventions that were similar to the CI activities in the CSA process. For instance, the comprehensive ergonomic interventions studied included data collection and trend analysis, followed by corrective actions with management support, process improvement, and a framework that suggested that supervisors were ultimately responsible for safety. These interventions were associated with declines in injury rates. Similarly, quality circle interventions appear to share important elements with the CSA process. They were also associated with injury rate declines, although not as consistently as the comprehensive ergonomic interventions.

2.2.4.2 Rail

Studies of CSA in the railroad industry are rare. Ricci (2003) found a 17 percent increase in the prevalence of safe behavior and a 57 percent reduction in injury frequency in a CSA-type implementation at a major freight carrier. He provided some evidence on the effectiveness in the railroad industry of observations without feedback but with continuous improvement. Unfortunately, a new technology was implemented with the departments in the study during the CSA treatment period, suggesting alternative explanations to his results.

The Station Services Department at Chicago's Amtrak terminal participated in a CSA demonstration pilot over two separate periods (Ranney et al., in preparation). An evaluation of the pilot found a decrease in injury rates following employee training on CSA methods. Furthermore, during both periods, injury rates significantly decreased as the total number of feedback sessions increased. These results were statistically detected despite an interrupted implementation and the lack of a SLD component, suggesting that there may be even stronger effects with an improved implementation.

Neither of these prior studies evaluated a CSA implementation for transportation operations, which makes up the bulk of the work on the railroad.

2.3 An Evaluation of a Pilot Demonstration

This report documents the first comprehensive field evaluation of a CSA demonstration pilot in freight operations on a service unit of a Class I U.S. railroad. The purpose of the evaluation was to assess the potential viability and effectiveness of the CSA process for transportation departments in the railroad industry.

3. Demonstration Pilot: CAB

In 2005, FRA identified a candidate site for evaluating the potential of CSA for railroad transportation departments: the San Antonio Service Unit (SASU) of Union Pacific Railroad (UP). A chain of serious road and switching accidents at this service unit, including the 2004 fatality at Macdona that the NTSB (2006) had attributed to insufficient road-crew attention, spurred UP interest in alternative approaches to safety, particularly for the SASU transportation department. Subsequently, FRA's Office of Railroad Development offered to sponsor a CSA demonstration at SASU, which received UP's wholehearted endorsement (see Initial Conditions, page 61). Local stakeholders named this CSA demonstration pilot Changing At-risk Behavior (CAB).

As a demonstration pilot, the SASU implementation is an experiment to test the viability of CSA in transportation departments. If successful, it can be an implementation model applied across the carrier operations and throughout the railroad industry.

3.1 Local Characteristics

3.1.1 Site

SASU is in the Southern Region of UP. At the start of the evaluation period in September 2005, the Southern Region also comprised the Houston, Fort Worth, and Livonia Service Units (the latter in Louisiana). During this period, as part of the UP reorganization, two more service units were added to the Southern Region. For the purpose of this evaluation, these were not included.

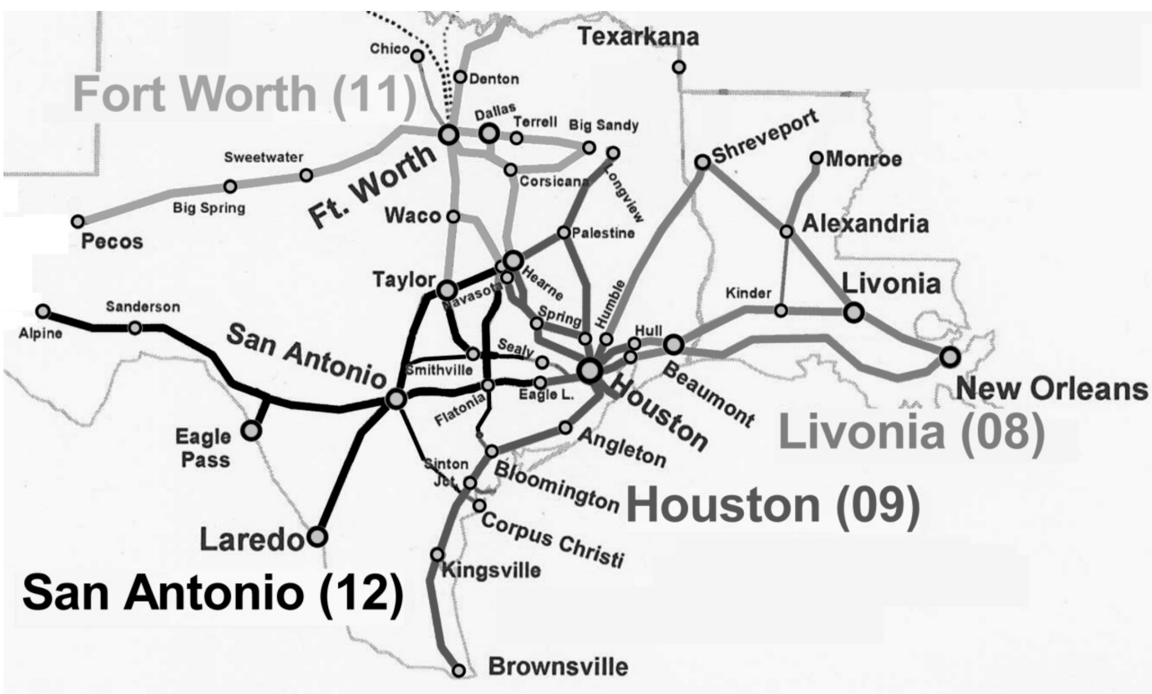


Figure 3. SASU (Black Lines) and the Three Other Service Units of the Original Southern Region of Union Pacific Railroad (modified figure from UP)

SASU is headquartered in San Antonio, where approximately 60 percent of the workforce reports. SASU comprised nearly 1,200 miles of main track, radiating from San Antonio to the cities of Corpus Christi to the east, Laredo and Eagle Pass to the south, Del Rio to the west, and Hearne to the north, with the latter including small terminals at Taylor, Round Rock, Smithville, and New Braunfels. To the north and east, SASU linked with tracks of the Fort Worth and Houston Service Units. Most track was under Centralized Traffic Control (CTC). The largest exception was the subdivision to Corpus Christi, which operated under track warrants.

There are three major yards in San Antonio: Kirby, East Yard, and South San Antonio (SOSAN). This report refers to these yards collectively as the San Antonio Complex. Important yards were also located at Laredo (two yards) and Eagle Pass along the Rio Grande. These three locations comprise the largest yards in SASU, accounting for 70 percent of all switching work as SASU (including switching work at minor stations such as sidings) and 78 percent of all switching-related derailments and other accidents prior to the implementation of CAB. All other yards at SASU are each a fraction the size of any yard at these three locations.

At the time of the pilot, approximately 50 managers and 1,100 workers were employed in the transportation department at SASU. Most of the work consisted of road operations. More than 80 percent of the workers were through-freight engineers, conductors, and brakemen.

3.1.2 Focuses and Phases

For the purpose of this evaluation, the pilot demonstration naturally broke into three phases corresponding to the type of work, as summarized in Table 1.

Table 1. Focuses and Phases of CAB, the CSA Implementation at SASU

Focus	Phase		
	Baseline (Jan. 2001–Aug. 2005)	**First (Sept. 2005–Sept. 2006)**	**Second Oct. 2006–Nov. 2007)**
Implementation	No CAB	CAB-CRZ only	CAB-CRZ and CAB-Switching
Work targeted by CAB		Road only	Road and yard

3.1.2.1 Baseline

The baseline phase represents the period before the feedback sessions began. As discussed in Step Comparison and Gradual Trends on page 46, archival data from this period provides a basis of comparison for the effects of CAB.

The evaluation period (consisting of times when the most data were collected for evaluation) covers the first and second phases.

3.1.2.2 First Phase

During the first phase, CAB focused on road work, specifically worker practices related to Cab Red Zone (CRZ) procedures. CRZ procedures primarily relate to road rather than switching crews. These were relatively new procedures and safety rules to improve work-crew attention during CRZ conditions—that is, times when the crew was working on multiple tasks. CRZ conditions include:

- Operating under signals such as Approach or Restricted, which imply an imminent need to stop the train.
- Approaching a temporary restriction, where a speed reduction or stopping may be required.
- Approaching the end of a train's authority, where stopping may be required.
- Copying mandatory directives radioed to the train crew, specifying locations within which a train must remain.

In an effort to improve crew attention and situational awareness, CRZ procedures included verbal cross-checking between the engineer and the conductor, confirming their perceptions of the CRZ condition, for example, which signal is showing, the current and maximum allowed speed, and the location of the next signal. The intent was to make the locomotive crews act more as a team, with crewmembers reinforcing each other and ensuring that everyone was awake, alert, and aware of the situation.

Stakeholders expected that the introduction of such CRZ procedures would face a challenge from the competitive relations between the two unions representing the transportation workers. Railroad engineers are generally represented by one union, Brotherhood of Locomotive Engineers and Trainmen (BLET), whereas the remaining transportation workers, predominantly conductors and switchmen, are represented by the United Transportation Union (UTU). These unions at times have competing interests (see Competition between Unions, page 30) that can interfere with communication and cooperation among engineers and conductors. Stakeholders hoped that CAB would provide the vehicle to encourage crewmember cooperation and communication that would otherwise not naturally occur.

Focusing on CRZ, the steering committee created a checklist for recording data from feedback sessions. The checklist included behaviors related to CRZ procedures (for example, crewmembers call out a signal name when it is more restrictive than Advanced Approach) and alertness under CRZ conditions (for example, crewmembers do not use cell phones).

3.1.2.3 Second Phase

In the second phase, CAB scope expanded to include observation and feedback of switching practices in the yards. The steering committee created a separate feedback checklist and training curriculum and materials for switching. The checklist included behaviors specifically associated with safety in the yard (for example, workers check switch points for alignment after throwing a switch) and those associated with safety under general industrial circumstances (for example, workers keep their eyes on their direction of motion).

Classes and data analysis for switching were conducted separately from those for CRZ, making CAB essentially two CSA implementations run by the same steering committee. (For this report, the implementations are distinguished as CAB–CRZ and CAB–Switching.)

Because road operations were largely centralized at San Antonio, most of the road employees across the service unit were part of the CAB–CRZ group from the beginning of the first phase. The CAB–Switching group was more localized. The steering committee established CAB–Switching at Eagle Pass yard first, then at the yards within the city of San Antonio, and later at other yards.

Coincidentally, the second phase corresponded to CSA or CSA-like implementations underway for switching practices at two of the other service units of the Southern Region: Houston and Livonia (see Comparison Data, page 47).

The evaluation team planned for an approximately 2-year evaluation period because this appeared to be the minimum time necessary to statistically detect improvements based on previous accident and injury rates and the expected rate of the CSA implementation. However, this meant that CAB–Switching had an evaluation period of only one year. As a result, the ability to detect the effects of CAB–Switching was somewhat compromised.

The second phase ended in December 2007. At that time, in order to enhance the sustainability of CAB, the CAB steering committee combined the CAB–CRZ and CAB–Switching into a simplified and consolidated process, with a single training curriculum, checklist, and database for the feedback data. CAB thus continued well past the end of the second phase, but, for evaluation purposes, quantitative process and feedback data after the second phase could not be meaningfully compared to the process and feedback data from before. This effectively marked the end of the evaluation.

3.1.3 External Consultation

Because a meaningful evaluation of CSA potential requires a complete and highly functioning implementation, FRA contracted with Behavioral Science Technology, Inc. (BST) to provide analysis and direction during the CSA pilot program for SASU. BST has implemented CSA-like programs in a broad range of industries.

As an outside expert consultant, BST was able to establish credibility with both labor and management, serving as a bridge between the normally adversarial groups to address the concerns of each, facilitate cooperation, and encourage support for the demonstration pilot.

A BST consultant conducted on-site training and data collection during approximately 10 visits to SASU. BST also provided training materials and software for tabulating and analyzing data collected by the process. The consultant tailored the proprietary BST process to the railroad industry. The consultant also worked closely with the pilot demonstration steering committee to anticipate obstacles to the implementation and to develop resolution strategies. Overall, the consultant instructed the steering committee on methods to solve problems rather than directly suggesting solutions. This helped ensure that the implementation was designed in a way that would work well for SASU, to foster ownership of the process, and to transfer knowledge to overcome future challenges after the consultant's role was phased out.

The consultant also conducted focus groups and a proprietary forced-choice survey[1] before and after the implementation to assess the corporate culture. The results of this data collection were reported in a timely manner to the steering committee, management, and union leaders to help design the implementation and sustainment of the process.

Finally, the consultant conducted the pilot demonstration's SLD program described on page 25. This involved working with top site managers to plan the program, and directly assessing and training all site managers.

3.2 CAB Process

This section summarizes the CAB process. For more details about the initiation and evolution of the CSA implementation in the context of SASU and UP refer to Field Notes and Project Records, page 61.

3.2.1 Organization

UP financed worker time and other resources, such as office supplies, computers, and working space. A Design Team, comprising local labor leaders of both UTU and BLET and top local management, appointed a CAB steering committee of UTU and BLET members to create and run the CAB process. All committee members worked on CAB part-time. Two additional workers were CAB *facilitators* who had full-time responsibility for leading and running CAB day-to-day. The Design Team selected workers by reviewing the qualifications of respondents to a position posting. Mindful of a certain level of political competition between UTU and BLET (see Competition between Unions, page 30, for details), one facilitator was an engineer with BLET and the other was a conductor with UTU. The steering committee's first task was to create the checklist of safe and at-risk behaviors for conducting feedback sessions as part of the PPF component of the CSA process. The committee also prepared training materials, such as media examples of safe and at-risk practices. A top SASU manager, appointed as the local management sponsor for CAB, acted part-time as the chief liaison between the steering committee and management.

3.2.2 Peer-to-peer Feedback and Continuous Improvement

Figure 4 illustrates the CAB-specific process for PPF and CI.

---

[1] At the request of the evaluation team, additional nonproprietary scales were included in the survey to allow an independent assessment of the safety culture (see Forced-Choice Survey, page 48).

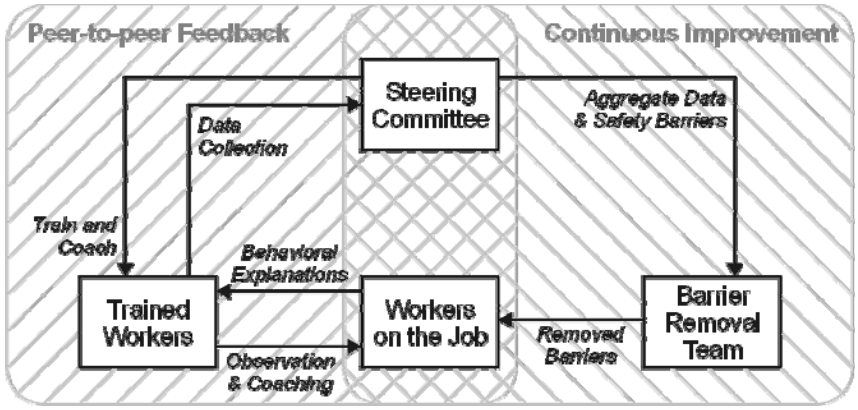


Figure 4. CAB Agents and Information Flow (adapted from figure from Krause, 1997)

Following development of the checklist for safe and at-risk behaviors, the primary task for the steering committee in the PPF component was to train and coach workers to conduct the feedback sessions—to teach them to become "samplers," as the steering committee called them. The samplers regularly observed workers in the field, openly and with their permission. Samplers recorded safe and at-risk behaviors on the checklist and reviewed them with observed workers, providing them with nonconfrontational feedback and documenting their explanations or reasons for any at-risk practices. The observed workers remain strictly anonymous throughout.

There were two forms of feedback sessions:

- *Third-party.* A sampler conducts a session with a worker performing regular work activities in a locomotive cab or a yard. The sampler's only duty is to conduct feedback sessions.
- *Cross-crewmember.* In the course of normal, regular work activities, a sampler conducts a session with a fellow crewmember. For example, an engineer may observe a conductor operating the locomotive on the road.

For the CI component of CAB, checklist data on safe and at-risk behaviors, along with the reported reasons for the latter, were compiled, aggregated, and analyzed by the steering committee to identify systemic barriers to safety—that is, repeated instances of at-risk behavior that can be attributed to training, tools, environmental conditions, procedures, management style, or other organizational issues.

The steering committee divided barriers to safety into two categories: those in which at-risk behaviors were considered voluntary or at the workers' discretion, and those in which employees were considered to have no choice but to perform at-risk behaviors due to conditions. The committee addressed the former type of barrier through continued observation and feedback or other communications with workers, such as articles in the CAB newsletter. Information on the latter type of barrier, along with supporting data, was submitted to a joint labor-management Barrier Removal Team, which had the task and authority to choose and execute the means to eliminate barriers.

3.2.3 Safety Leadership Development

The SLD component of CSA ran simultaneously with the PPF and CI components, but without the involvement of the steering committee. Based on the results of focus groups and surveys of management,[2] BST trained managers to more effectively coach workers on safety. The consultant supported top SASU management in developing action plans for encouraging more proactive and nonpunitive safety activities among all managers.

[2] These surveys were separate from those used to answer the evaluation questions on outcomes.

4. Evaluation

The evaluation of CSA generally followed the Context Input Process and Product (CIPP) model for program evaluations (Stufflebeam, 1983):

- **Context**. The goals for change were derived from the understanding of the context of the railroad industry (see Context: The Problem of Railroad Safety and Culture, page 8).
- **Input**. CSA was selected as the safety process for the pilot demonstration based on a review of the literature on risk-reduction methods (see CSA Concept, page 13).
- **Process and Product**. As described in this section, the pilot demonstration was evaluated both for its implementation (process) and outcomes (product), along with its sustainability and transformational capacity.

The evaluation was both formative and summative. It assessed a CSA demonstration pilot within two types of work environment, semidispersed (yard work) and dispersed (road). In the absence of strict experimental control, the evaluation used a multimeasure, mixed-method (qualitative and quantitative), quasi-experimental design to assess the implementation, outcomes, and impacts of the pilot demonstration.

4.1 Evaluation Questions

To assess CSA's potential viability and effectiveness for railroads, this evaluation addressed the following questions:

- **Implementation.** What characteristics of a CSA implementation allow its activities to be completed as planned in the railroad industry? How should the CSA process be improved to more effectively establish it in railroad settings?
- **Outcomes.** What are the effects of the CSA process on safety and safety culture?
- **Sustainability.** What characteristics of the implementation and organization (e.g., economic and political) are necessary for CSA to be sustained once it has been successfully implemented?
- **Transformation.** How does the CSA process influence the broader organizational environment of the pilot demonstration, if at all? Specifically, does the process make the stakeholder organizations more likely to adopt proactive, nondisciplinary approaches to safety, thereby encouraging the spread of processes similar to CSA to other parts of the organization or industry?

The question relating to outcomes addresses the effectiveness of CSA (Rossi, Freeman, and Lipsey, 1999) and as such constitutes the bottom line of the evaluation.

The questions relating to implementation and sustainability are the core of the formative aspect of the evaluation, addressing two issues. First, by determining the characteristics that allow a complete and sustainable implementation, the evaluators can provide stakeholders with practical advice on the demonstration pilot. A complete implementation is one in which all the implementation activities in the theory of action (see Figure 2, page 14) are carried out at the planned levels. The formative evaluation may provide information to help adjust the implementation for the realities of the site, and the railroad industry in general.

Second, the evaluation also documents the characteristics of a complete and sustainable implementation. They can become lessons learned for stakeholders considering initiating a CSA process at subsequently identified railroad sites.

The implementation and sustainability questions also address two issues related to the summative evaluation. First, the questions imply an assessment of *if* the implementation was complete and sustainable. The successful piloting of a complete and sustainable CSA implementation would address the viability of CSA in the railroad environment and show that implementation is possible. Second, establishing the presence of a complete CSA implementation is necessary to justify the evaluation of outcomes. Like most program evaluations, it is unsound to judge the outcomes of CSA if they cannot be associated with a sufficiently complete implementation (Patton, 1997).

Implementing a complete and sustainable CSA pilot demonstration is not trivial. As previously discussed, even though CSA may improve the safety culture within the railroad, this existing culture may interfere with establishing and maintaining CSA in the first place. Like other new approaches to railroad safety (Ranney and Nelson, 2004), CSA promotes cooperation and trust, but it also requires a minimal level of cooperation and trust in labor–management relations to get started. To implement CSA, managers must trust workers to monitor and correct their own behavior, and workers must trust managers not to subvert the process for disciplinary purposes. The trust from labor is particularly difficult to obtain if they are suspicious of the PPF component, as some have been in the past (Spigener and Hodson, 1997; Howe, 1999; Frederick and Lessin, 2000). In addition, managers and workers must cooperate on the barrier-removal process.

The solution to the apparent paradox that CSA initially needs labor and management to trust each other to be able to later promote such trust lies in the initial context and implementation. Theoretically, an effective implementation, with the right outside support and initiated at the right time and place, can establish an "island" of trust within a "sea" of distrust characteristic of the railroad culture. It can then expand on that trust until the entire culture at a railroad site has changed.

Therefore, it is important to evaluate and document the implementation characteristics and external factors that both promote and impede the completion and sustainability of an effective CSA process. This provides guidance and practical examples for any future attempts to implement CSA in the railroad industry.

The question about transformation addresses the usefulness of CSA in helping to achieve the larger FRA mission of improving the railroad industry's safety culture. Success at a single site is desirable, but spreading this success to other sites is proportionally more beneficial. If CSA can influence larger organizations, this could be very promising in moving towards achieving widespread cultural improvements.

4.2 Evaluation Scope

The evaluation focuses on railroad transportation (road and switching) operations. Transportation represents the bulk of railroad operations and includes the following features relevant to testing the CSA process:

- Potential high magnitude of safety outcomes.
- Substantial role of human factors in transportation safety.
- Dispersed and semidispersed work settings.
- Competition between unions.

4.2.1 Potential High Magnitude of Safety Outcomes

Switching operations are appropriate for testing CSA because historically they have been associated with the majority of injuries and accidents on railroads, particularly derailments (FRA, 2006). Switching improvements should therefore result in the greatest reduction in injuries for the industry. Furthermore, the rates of fatalities and serious injuries from switching have remained unchanged for 20 years (FRA, 1999, 2001), suggesting that new approaches to safety are necessary to improve these outcomes.

Finally, there is reason to believe that a safety process with a behavioral component would be particularly effective for switching operations. In one analysis of switching fatalities, all recommended interventions were related to new procedures for improving worker communications (FRA, 1999). Since such procedures are manifested by worker behavior, a process designed to change this behavior would be the most suitable.

Road accidents are rarer than switching accidents, but they are more likely to be catastrophic, since they include high-speed collisions that can endanger not only railroad workers but the general public (see, e.g., NTSB 2005a, 2006). PPF interventions have typically been applied to work in which the primary concern was personal injuries, such as those associated with falls, pinch points, or impacts of small objects. For these kinds of accidents, the contributing personal behavior and conditions are relatively easy to observe. Accidents involving road operations, in contrast, typically are associated with internal personal states, such as boredom, fatigue, level of attention, and situation awareness, which can lead to behavioral errors, such as missing signal indicators. These internal states are not readily visible in a feedback session. Road operations therefore provide an opportunity to test a PPF-related intervention in a challenging work context.

4.2.2 Human Factors in Transportation Safety

Injuries and accidents in transportation operations are usually attributed to human factors. For yard operations, the most severe accidents tend to result from inadequate inter-worker communication (FRA, 1999) or errors of omission (NTSB, 2005a). For road operations, accidents are more often associated with worker boredom, inattention, and fatigue (NTSB, 2006). These are products of road workers' unpredictable and irregular work schedule, which disrupts circadian rest cycles (Pilcher and Coplen, 2000) combined with a job featuring long stretches of relative inactivity interspersed with periods requiring sustained attention.

A safety improvement process must address the role of workers in accidents and injuries, as well as their behaviors and psychological states. From a systemic orientation, a full understanding of accidents and injuries within organizations recognizes that "human factors" is not an adequate explanation for accidents and injuries. Rather, these proximal human factors causes of accidents and injuries are themselves caused by latent organizational factors—systemic conditions related to the workplace environment, such as training, policy, and informal incentives (Reason, 1997). CSA represents a process to improve worker behavior through PPF and organizational context through CI and SLD, making it particularly suited for improving safety in transportation.

4.2.3 Dispersed and Semidispersed Work Settings

To thoroughly test the effectiveness of CSA for the railroad industry, it is important to assess whether it works regardless of the degree of worker dispersion in the work setting. For most railroad workers there are three types of work settings:

- *Nondispersed, or in-facility setting.* Workers are generally located within sight of each other and supervisors are close by. This is characteristic of mechanical shops or terminal materials-movement departments, among others.
- *Semidispersed setting.* Groups of workers may operate in teams that are spread out or individual workers may perform tasks separately. In either case, workers are within walking distance from one another and supervisors. This is characteristic of transportation-yard-switching and maintenance-of-way departments.
- *Dispersed setting.* Workers are not within sight of each other or their supervisors when they perform most of their activities. This is characteristic of transportation road and signal departments.

As work and workers become more dispersed or increasingly spread out, supervision and coordination mechanisms are adjusted to support the work efficiently and effectively. It is relatively easy to coordinate and supervise workers who are all located in one building on a single shift, but it becomes increasingly complex to coordinate and supervise workers who are not co-located. If increased worker dispersion requires special features from CSA to be effective, these features should be identified before the process is recommended to the industry. If CSA is not effective in dispersed settings, the industry should be aware to avoid wasting resources.

Of these three settings, dispersed settings are thought to be the most challenging for processes like CSA because of their coordination challenges (Krause, Seymour, and Sloat, 1999; Sulzer-Azaroff and Austin, 2000; Ranney et al., in preparation). PPF interventions have typically been implemented in nondispersed settings involving workers co-located in the same facility, such as those in manufacturing, with regular, predictable duty shifts and work assignments. This makes it relatively easy to conduct PPF training, coaching, and feedback sessions. A prior demonstration pilot found positive effects of CSA with in-facility workers (Ranney et al., in preparation).

Because most railroad work, such as road and switching operations, is in transportation, work settings tend to be dispersed or semidispersed. Road operations are dispersed, with engineer-conductor pairs typically running trains over hundreds of miles. Typically, management draws from a pool of workers to operate trains when crews are needed (subject to hours-of-service regulations). This results in erratic schedules for individual workers. The work of running a train is itself irregular, featuring times of both intense concentration and virtual inactivity. For example, when a train stops at a siding depends unpredictably on combinations of traffic congestion, track conditions, and terrain.

The dynamic nature of road operations complicates planning for training, coaching, and feedback sessions, making it difficult to conduct a high rate of sessions. Without enough observation, there may not be sufficient feedback to provide the PPF component of CSA. CI relies on feedback sessions for data on which to act, but data may be scarce if observation is not sufficient. Road operations thus represent the most challenging setting for CSA. It may be argued that if CSA can be made to work there, it can be made to work in most other railroad work settings.

Switching operations, in contrast, are less dispersed, typically featuring regular hours for workers and a steadier level of activity, although operations for a single crew may be spread across a yard. Two workers on the same crew may be separated by a mile or more as each protects opposite ends of a cut of cars. This means switching operations provide a more dispersed work setting than an in-facility work setting even though this is a less dispersed setting than road operations. The level of difficulty in planning for training, coaching, and feedback sessions is therefore intermediate in relation to nondispersed work-settings of prior research and dispersed work settings of road operations.

The demonstration pilot described in this report tests CSA in transportation under semidispersed and dispersed work settings.

4.2.4 Competition between Unions

Although labor–management relations are contentious in the railroad industry, there is also conflict between transportation labor organizations, BLET and UTU (e.g., Pierce, 2005), that may undermine labor's cooperation within itself. Although the two unions share many interests and perspectives, including favoring the technological safety orientation, the two unions have at times competed against each other on certain issues. For example, when remote-controlled locomotives were introduced for yard work, BLET, which represents predominantly engineers, lobbied that only certified engineers be allowed to control these vehicles, whereas UTU, which represents predominantly non-engineers, advocated for allowing switchmen to operate them (Hume, 2003). Both unions were protecting the jobs of their respective members. More recently, UTU and BLET have strained to achieve a common response to corporate efforts to reduce road train crews from a conductor and an engineer to one engineer (Railway Age, 2006).

Such competition has undermined trust between the unions, which may interfere with CSA since the process requires unions to cooperate with each other. At the very least, CSA must not to appear to favor the members of one union over the other.

4.3 Evaluation Design

4.3.1 Formative and Summative Evaluation

The evaluation of the demonstration pilot is both formative and summative. The purpose of the summative component is to provide scientifically credible conclusions relating to the viability and effectiveness of CSA for railroad transportation departments based on the assessment of the CSA process at the site. This report documents those summative conclusions.

Given the experimental nature of CSA, the demonstration pilot may fail to be viable or effective, not because of inherent limitations of CSA, but because of deficiencies in the specific implementation or operation at the site. Ideally, to assess the potential of CSA to improve safety, the most effective implementation and operation possible is needed.

To maximize both the likelihood of high-quality implementation and operation, as well as the overall usefulness of this evaluation (Patton, 1997), the evaluation includes a formative component (Rossi et al., 1999) in which the evaluation team continuously assesses whether the demonstration pilot is being implemented as planned. The evaluation team periodically and directly informs site stakeholders of its assessment of the demonstration pilot and provides recommendations for improvement. The formative evaluation must be completed before the summative evaluation to provide an opportunity to test and tailor the CSA process to the railroad industry.

The evaluation team provided two formal feedback sessions on the demonstration pilot's performance during the evaluation period and a third session after the evaluation ended. The team also coordinated with stakeholders to provide informal feedback throughout the evaluation period. CSA effectiveness may have been augmented by this systematic evaluation and feedback.[3] (The apparent value and specific impact of the evaluation team is discussed on Effects of the Evaluation, page 145.)

4.3.2 Mixed-Method Evaluation

It was not feasible to randomly assign UP personnel at the demonstration site to participate in the CSA process. Because workers operate in crews with ever-changing membership, it would be unlikely that a control group at the site would remain truly isolated from the effects of the intervention. Given the limited manpower, hours-of-service requirements, and pool of workers used for road work, it would be operationally necessary for workers who have received feedback as part of the CSA process to interact with workers who have not. One group may likely influence the other, reducing differences between the experimental and control groups and threatening the integrity of the experimental design (Rossi et al., 1999).

---

[3] This specific demonstration pilot used its own metrics for quantitatively evaluating and improving the implementation in accordance with a consultant's recommendation. Some of these metrics were used by the evaluation team for summative purposes, as described in CAB Process Metrics on page 40.

The evaluation team thought the prevalence of distrust among workers and between workers and managers in the railroad industry would result in workers viewing random assignments with suspicion that could ultimately undermine the process.[4] The team also thought worker suspicion would also be aroused by the collection of observation data for a baseline period to compare the effects of CSA in a pretest/posttest experimental design.

Instead of randomly assigning site personnel to experimental and control groups, it was necessary to involve all transportation personnel.

Consequently, in situations such as this, an experimental design with random assignment was not feasible or desirable for this evaluation. The most rigorous research design in this case uses both quantitative and qualitative data collection and analysis methods from multiple data sources (GAO, 2009). Following a concurrent triangulation strategy (Creswell, 2003), causation is inferred to the degree that the results of the diverse quantitative and qualitative methods and data sources are consistent with an effect of CSA, as described below.

4.3.2.1 Quantitative

Quantitative methods and data sources include data collected as part of the CSA process, (e.g., implementation-effectiveness metrics), attitudinal surveys, and safety-related corporate repositories of occurrences such as injury rates and derailments.

The team analyzed quantitative data in a quasi-experimental design (Shadish, Cook, and Campbell, 2002), in which performance is measured and evaluated following the application of CSA to the demonstration site. For safety-related occurrence data, performance is compared with contemporaneous performance without CSA. Such comparison occurrence data may be obtained from locations without a CSA process, or for occurrence types not associated with CSA. For example, derailment rates at yards with a CSA process may be compared with rates at yards without CSA during the same time period to help separate changes that may be due to regional and corporate factors rather than to CSA. (For further details about the use of comparison data refer to Comparison Data on page 47.)

The team analyzed data for trends during the CSA implementation and, if available, compared differences before and during the CSA process.

An effect of CSA on a variable is considered likely if all of the following are true (Shadish, Cook, and Campbell, 2002):

- Gradual or step changes in the variable are detected for the appropriate locations and occurrence types since the start of the CSA implementation.

[4] In this evaluation, a proposal for a limited random-assignment experiment on the effects of observer training was subsequently withdrawn due to stakeholders' concern that it would arouse damaging suspicion. Stakeholders felt that workers in the experimental group might resent being "singled out" as if they were especially unsafe and in need of special attention.

- Similar changes are not detected in comparison data from the same time as the CSA implementation.
- The changes are not a continuation of trends from before the start of the CSA implementation.
- Alternative explanations are not sufficiently plausible given the evidence.

Analysis Strategies for Corporate Safety Data on page 46 discusses specific techniques for assessing the above statements.

4.3.2.2 Qualitative

Qualitative methods and data sources include field notes and project records, as well as interviews. These are important for assessing the effects of uncontrolled factors, such as management change, which are unpredictable but inevitable in a field evaluation. The team evaluated the data using a case-study methodology (Yin, 2003).

CSA is considered to have an effect if the qualitative analysis demonstrates the following:

- Outcomes and impacts are linked, as shown in the Theory of Action (Figure 2, page 14, and described in Figure 7, page 38).
- Results from field notes and interviews are consistent with each other and with any relevant quantitative results.
- Alternative explanations are not sufficiently plausible given the evidence.

4.4 Evaluation Plan

To answer the evaluation questions, the evaluation augments the Theory of Action, Figure 2, to become a logic model[5] (W.K. Kellogg Foundation, 2004).

4.4.1 Implementation Evaluation

To answer the evaluation question relating to the success and adjustments of the implementation, the logic model includes the same implementation activities found in the Theory of Action (see Figure 2, page 14).

The presence of the implementation activities indicates the successful completion of the implementation. The evaluation records the presence of each activity. Completion of observer training and feedback sessions are represented quantitatively by the number of workers trained and the frequency of feedback sessions, respectively.

Events and stakeholder actions during the pilot demonstration that appear to be responsible for the completion of each implementation activity (or lack thereof) can also

[5] A logic model is a visual depiction of the activities of a process, such as CSA, and typically is used to guide a program evaluation. The model depicted in this report in various forms represents a consolidation of several logic models (in addition to the Theory of Action) created by the evaluation team for various purposes over the course of the evaluation.

be determined qualitatively. The result is a list of process characteristics that are important for a successful implementation.

4.4.2 Outcome Evaluation

To answer the evaluation question relating to safety outcomes attributable to the CSA process, the logic model subdivides each outcome in the Theory of Action into qualitatively or quantitatively measurable elements, following the approach used for the implementation evaluation. This subdivision of the outcomes into elements was not included in the original theory of change model so a new evaluation model is provided. These measurable outcome elements are represented as shaded, rounded boxes in the outcomes illustrated in Figure 5.

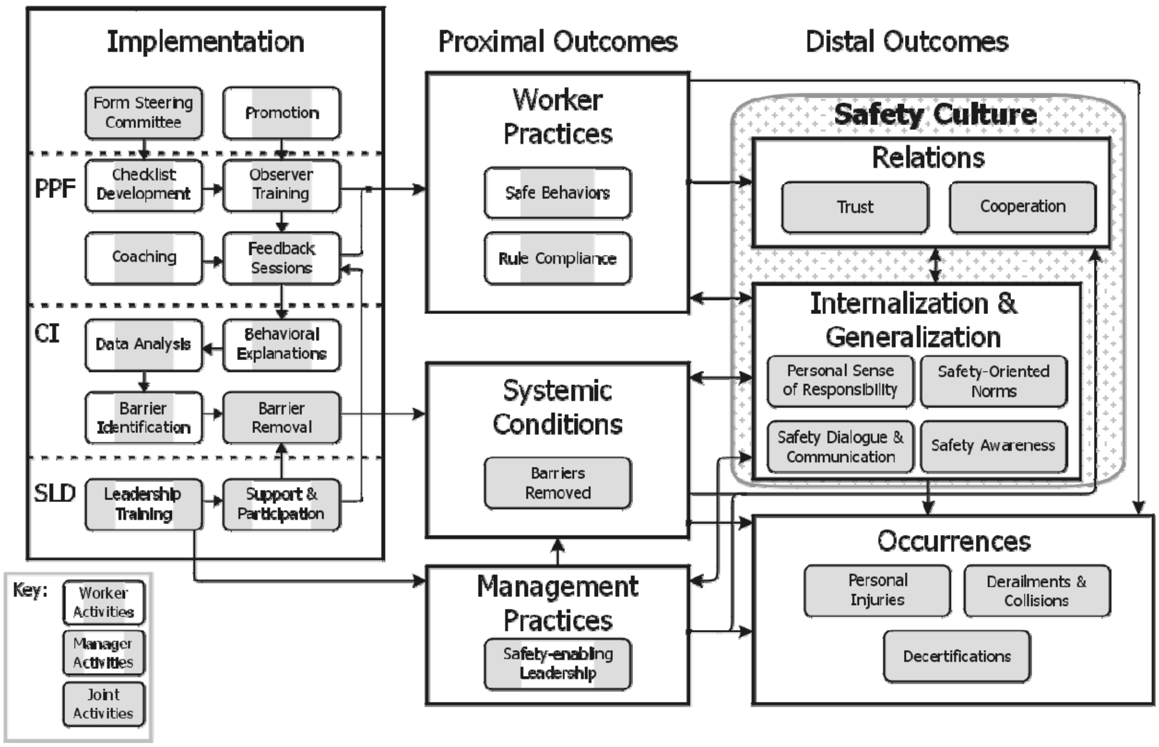


Figure 5. Logic Model, with Elements for Evaluating the Outcomes

Safety culture can be split into two separate outcomes:

- *Relations*. The organizational components of safety culture given by Reason (1997), especially relations among workers and managers relating to safety. They represent the degree to which labor-management relations are conducive to effective safety processes, which depend on the open exchange of information between labor and management that in turn requires cooperation and trust between the two entities.
- *Internalization and Generalization.* The personal components of safety culture given by Reason (1997), including individual worker and manager attitudes, values, norms, and ways of thinking related to safety. It also includes personal conversations on safety in which such attitudes, values, and norms are both

expressed and acculturated. These components represent the degree to which safety culture has been psychologically internalized by workers and managers so that it may influence behaviors beyond those specifically targeted by the implementation.

The evaluation assesses each outcome separately by measuring each element. Improvements in worker practices targeted by the PPF component are measured by quantitative changes in safe behaviors recorded by workers in their feedback sessions, provided these records meet minimum psychometric criteria. Rule compliance provides an indication of these worker practices, which can be quantitatively assessed using records of the railroad's field operations testing.

The evaluation qualitatively estimates the improvements in both worker and management practices targeted by the CSA process based on changes reported by workers and managers in interviews. A raw count of barriers removed provides a crude quantitative measure of improvements in systemic conditions, but qualitative information from interviews and field notes may give a better indication of the value of each barrier removed—for example, the number of people affected or the magnitude of the improvement in the work environment.

Internalization and Generalization are quantitatively measured by survey instruments for safety-oriented norms, and qualitatively measured by interviews for the presence of safety dialogue (including discussions between workers and between a worker and supervisor), safety awareness (for safety issues beyond those specifically targeted by the PPF component), and personal sense of responsibility within each individual.

For Relations, a survey instrument measures labor–management cooperation, whereas the presence of cooperation and mutual trust is qualitatively indicated by interviews.

Key occurrences of interest at the site include personal injuries and accidents, such as derailments and collisions. As a proxy measure for catastrophic road accidents, engineer decertifications are used, since they relate to rule violations that can easily result in catastrophic events.

4.4.3 Sustainability Evaluation

To answer evaluation questions relating to the sustainability of the CSA process and its capacity to transform the larger organizational context, the pilot demonstration must be viewed within the larger corporate and industry environment that it impacts, as shown at the bottom of the modified logic model in Figure 6. In that figure, implementation activities and outcome elements are hidden for clarity. The gray arrows represent influence links between the environment and the pilot.

Sustainability is largely represented by the financial and legitimization support for the implementation from external organizations, as shown by the Support arrow at the bottom left of Figure 6. Such organizations include:

- Workers affected by the process.
- Unions and union leadership at various levels.
- Management at the local, regional, and corporate levels.

- FRA.

It is assumed that the implementation achieves such support by three forms of stakeholder involvement:

- *Visibility.* External organizations must be aware of the implementation and be accurately informed of what it aims to achieve.
- *Interests.* These organizations must believe that the implementation is consistent with their own interests, which, in addition to improving safety, includes controlling costs for management, and maintaining or increasing control over working conditions for unions.
- *Engagement.* The organizations must have a role to play in the implementation so that they have an investment that they will want to realize.

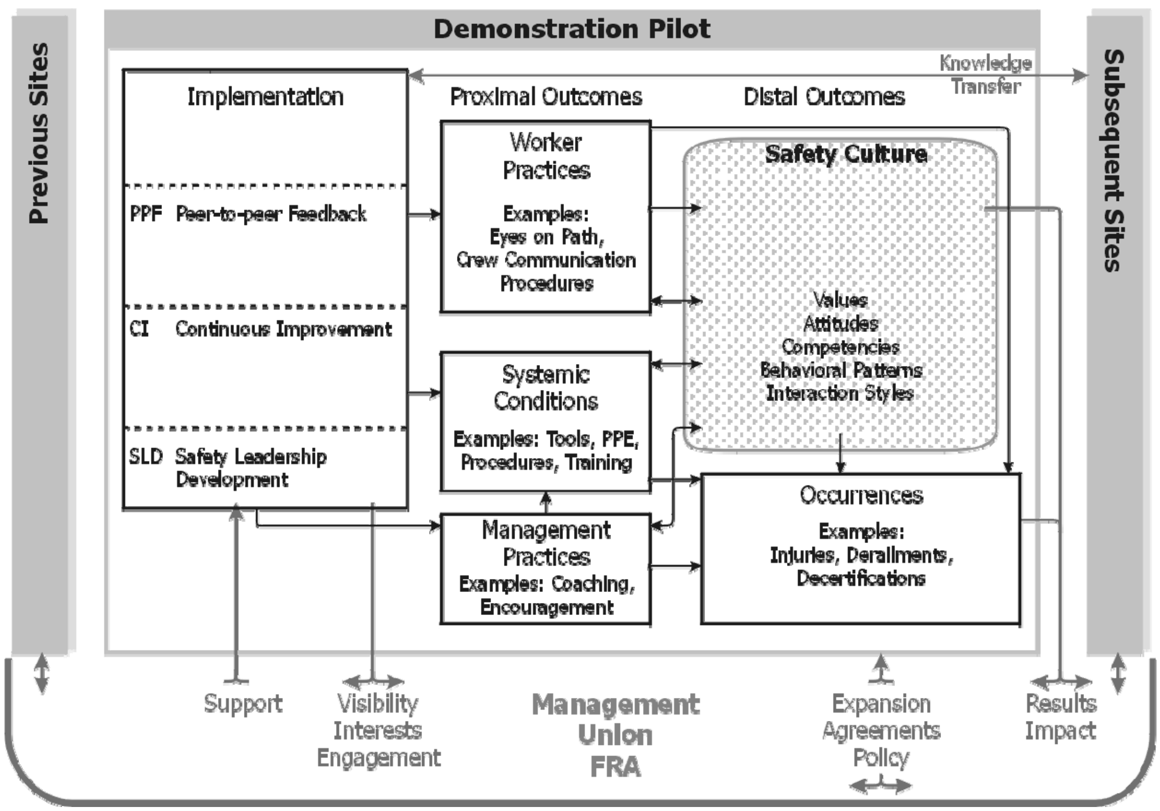


Figure 6. Logic Model with a Corporate/Industry Environment for Evaluating Sustainability and Transformation

4.4.4 Evaluation of Transformation

Transformation is an impact largely manifested by expanding the CSA process to other sites, along with the introduction of regional- and national-level agreements and policies consistent with CSA, as shown in Figure 6. Results hypothetically drive a transformation: the more corporate and labor organizations see CSA as positively improving outcomes such as occurrences and relations, the more they will be transformed. This will lead organizations to establish CSA or CSA-like processes at subsequent sites, where their own results will feed back into the transformation. Direct communication between implementation personnel at the initial site and those at subsequent sites can result in knowledge transfer, which in turn improves the chances of success at both sites.

Successes at any previous CSA-like sites likewise propel the transformation, as can significant difficulties in relations or occurrences at sites without CSA, which may motivate organizations to make changes.

Qualitative analysis of field notes identifies the presence of each of these forms of influence to assess sustainability and transformation.

4.5 Summary of Evaluation

This report documents a formative and summative evaluation of a CSA demonstration pilot. Evaluation questions are related to implementation success, outcome effectiveness, sustainability, and transformation of the organizational environment in which the pilot demonstration is embedded. The questions are answered by assessing the implementation activities, manager and worker practices, conditions, safety culture, occurrences, and transformational influence on the broader organizations. Each question is assessed with multiple qualitative and quantitative measurement elements. There are often multiple elements per outcome to provide triangulation toward underlying changes. When possible, comparison data are used to separate factors not associated with the CSA process. The fully detailed logic model for the evaluation, with all the elements discussed above, is illustrated in Figure 7.

If the evaluation finds that CSA is successfully implemented, has positive outcomes, and is sustainable and perhaps transformational, then CSA will become a scale-up candidate for transportation departments across the entire railroad industry. This is an ambitious move based on the evaluation of one demonstration pilot on a single service unit. It is justified by the inclusiveness of the demonstration pilot, which focuses on both yard and road operations, and by the thoroughness of the evaluation.

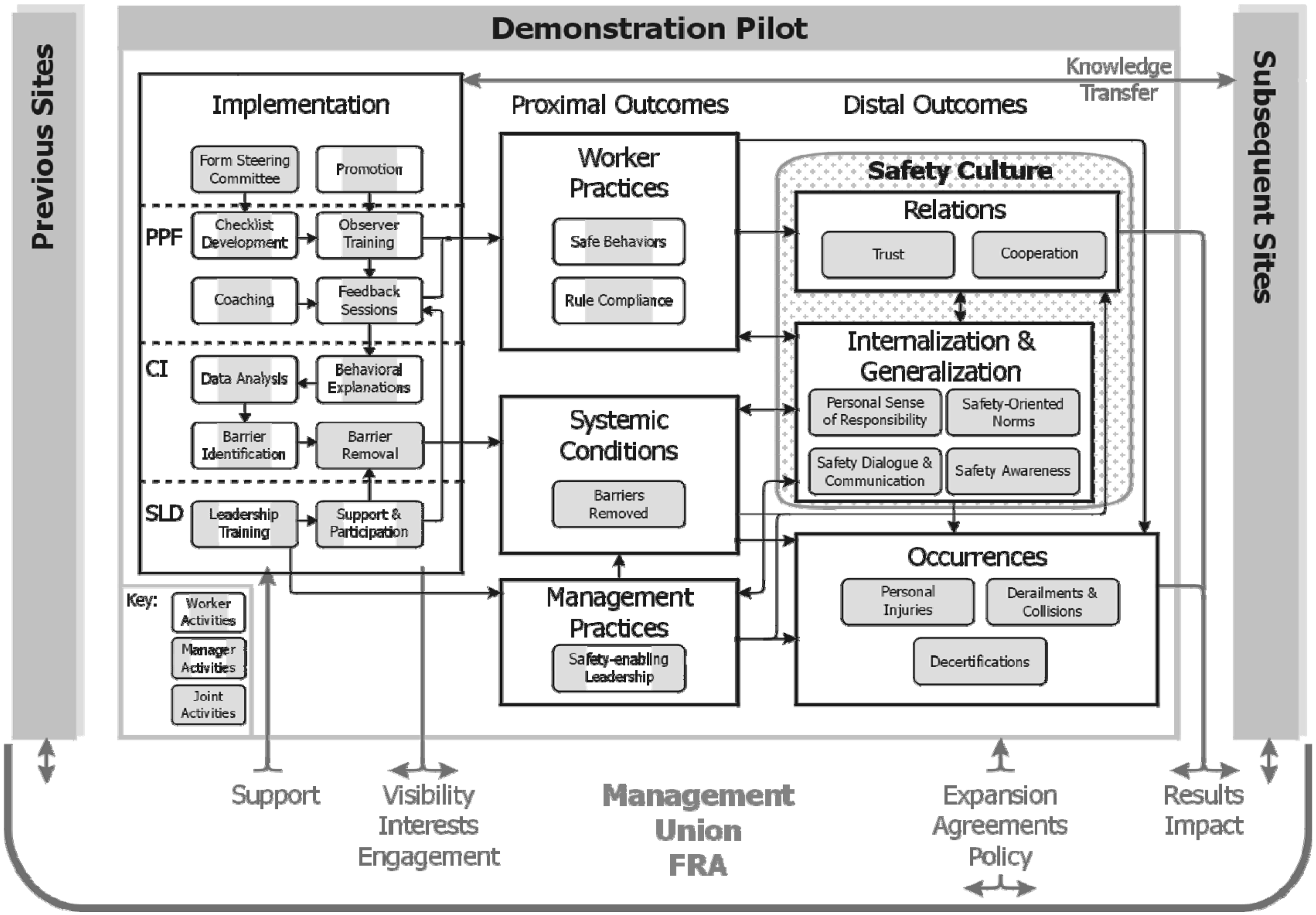


Figure 7. Full Logic Model for CSA Evaluation

5. Evaluation Methods

To conduct an independent evaluation of the demonstration pilot, FRA assembled a team from the John A. Volpe National Transportation Systems Center of the U.S. Department of Transportation, an entity separate from FRA and UP. The evaluation team was engaged over a 3-year period, concurrent with the establishment of the pilot demonstration and its first 2 years of operation, with the latter constituting the evaluation period.

The team evaluated CAB's implementation, outcomes, sustainability, and transformational effects using multiple qualitative and quantitative methods. Multiple measures were used for the implementation and each outcome (where possible) to provide redundancy.

Table 2. Methods of Evaluation and Their Relation to Logic Model Factors

Method	Type	Assessment
Field Notes and Project Records	Qualitative	Implementation Outcomes: Management Practices and Systemic Conditions Sustainability and transformation: Union, Management, and FRA Impacts
CAB Process Metrics	Quantitative	Implementation
Feedback Session Data	Quantitative (gradual trend)	Outcomes: Worker Practices
Corporate Safety Data	Quantitative (gradual trend and before/after)	Outcomes: Occurrences
Forced-Choice Survey	Quantitative (before/after)	Outcomes: Worker Practices; Internalization and Generalization; Relations
Interviews	Qualitative	Implementation; Outcomes: Worker Practices; Internalization and Generalization; Relations; Systemic Conditions

Table 2 summarizes methods of evaluation, type of evaluation, and the role of each method in answering the evaluation questions. The team collected data with each method repeatedly or continuously to assess changes associated with CAB's operation. For quantitative methods, the Type column in the table shows in parentheses the analyses for detecting change:

- *Before/after*, the means or rates before and after the start of the implementation are compared.
- *Gradual trend*, a correlation with time is calculated for the time span corresponding to the implementation.

Figure 8 illustrates the timeline for the methods of evaluation, superimposed on the phases discussed in Focuses and Phases on page 20.

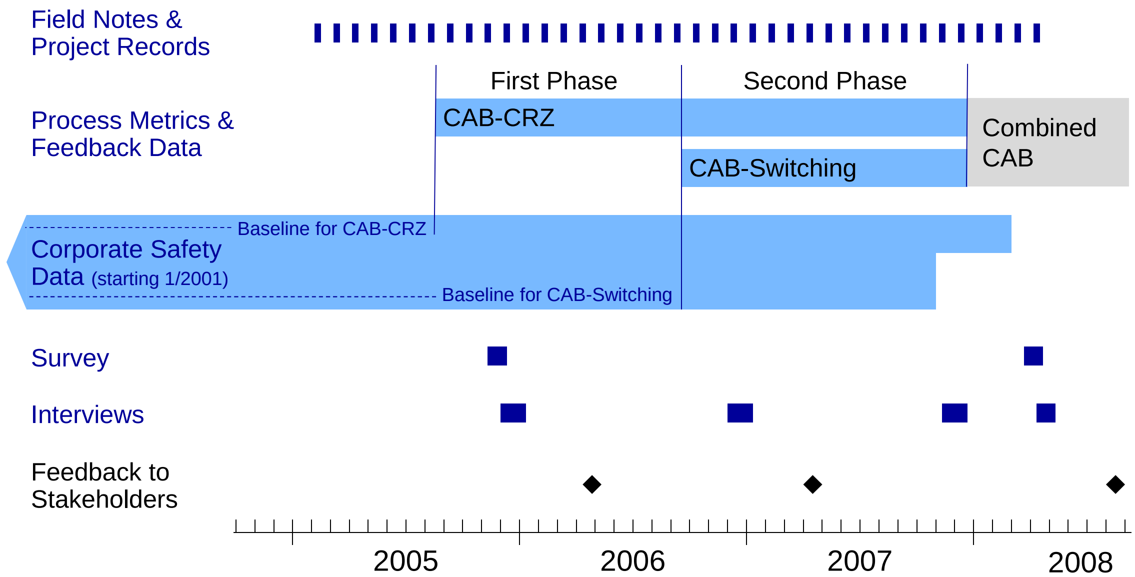


Figure 8. Timeline for Evaluation Methods

The evaluation team compiled field notes monthly through a roughly even split between site visits and telephone conferences (telcons) with stakeholders (see Acquiring Resources and Managers' Time a Challenge, page 66). The team used the forced-choice survey on two occasions and conducted interviews on four occasions with the fourth having a different scope than the previous three. The team also conducted two feedback sessions with stakeholders during the evaluation period as part of the formative evaluation, and a third feedback session after the evaluation period.

Details of the methods are described in subsequent subsections.

5.1 Field Notes and Project Records

Field notes and project records provide qualitative data on events, perceptions, and decisions involving CAB stakeholders, such as facilitators, the steering committee, and the management sponsor. Evaluation team members compiled these field notes as handwritten accounts of direct observations and interactions with stakeholders. Field notes were recorded from February 2005, when CAB was being designed, to April 2008, after quantitative analyses had been completed. Quotes were generally from stakeholder meetings and telcons witnessed by evaluation team members rather than from private conversations with team members. The quotes were typically responses to other stakeholders' comments or to the business at hand rather than replies to direct questions by evaluation team members (compare with Interviews, page 58).

Project records are documents provided by the stakeholders on the implementation. They include PPF checklists, business letters, reports, news releases, newsletters, Web site articles, and transcripts of testimonies. The evaluation team tracked the date of creation for each project record.

These data provide information on the effectiveness of the implementation and its impacts on the stakeholder organizations, particularly with regard to sustainability and transformation. They also serve as a means to document management practices and changes in systemic conditions.

5.2 CAB Process Metrics

In accordance with BST training and materials, CAB kept data on its process to allow the steering committee to monitor the status of the implementation based on comparisons with other BST implementations. For the purpose of evaluating the implementation and assessing its completeness, there were two key metrics: *training completed* and *contact rate.* Each metric provided a quantitative measure of the effectiveness of the implementation. Raw *number of barriers removed* is another quantitative-process metric often used in BST implementations. However, the barrier-removal process for CAB had features that precluded this as an accurate measure (see Steering Committee Selective in Identifying Barriers for Formal Removal, page 72). Useful process metric data were available from the start of the CAB–CRZ and CAB–Switching implementations until the two implementations were combined at the end of 2007.

5.2.1 Training Completed

Training completed refers to the percentage of the targeted workforce that completed training for conducting peer feedback sessions. For CAB–CRZ, the targeted workforce was the 900 road workers, whereas for CAB–Switching, it was the 200-yard workers. Training completed was an indication of the steering committee's productivity and of the degree that CAB had penetrated the workforce.

5.2.2 Contact Rate

Contact rate is the number of feedback sessions conducted per month, divided by the number of workers in the targeted workforce (900 for CAB–CRZ, 200 for CAB–Switching). Thus, a contact rate of 100 percent implies that on average each worker receives feedback once a month. If that face-to-face feedback and discussion of worker behavior promotes safe practices, higher contact rates should hypothetically lead to faster improvements in practices, all other factors being equal. Because the capacity to conduct feedback sessions depended on the number of trained peer observers available, it was expected that contact rate would increase gradually throughout the evaluation. For this site, with advice from the BST consultant, the CAB Design Team chose a target contact rate of 100 percent. This evaluation therefore used 100 percent to judge the strength of implementation at this site. However, BST has reported that contact effectiveness depends more on the quality of the peer interactions during the contacts than on the quantity of contacts, and the ideal contact rate for a specific site may be lower given necessary logistics and types of risk exposures (Bell, 2010). Indeed, a contact rate below 100 percent has been associated with detectable safety effects at another railroad site (Ranney et al., in preparation).

5.3 Feedback Session Data

In accordance with BST training and materials, the CAB steering committee collected the checklists for completed feedback sessions. Like the process metric data, feedback session data were available from the start of the implementations until CAB–CRZ and CAB–Switching were combined.

Software provided by BST scored the data on each checklist to record which behaviors were performed safely and which behaviors were at risk. A behavior was scored safe if it was observed at least once and if, for however many times it was observed, it was always conducted in a safe manner as defined by the steering committee. A behavior was scored at risk if it was observed at least once and if, at least once, it was conducted in an unsafe manner as defined by the steering committee. The steering committee tabulated scores from the checklists to calculate the percentage of at-risk/safe behaviors observed for a given checklist item, time period of observation, and/or location or other condition. It used these statistics to identify patterns of at-risk behavior that imply a systemic safety barrier, to prioritize barriers for removal, and to assess the effectiveness of barrier-removal efforts.

For purposes of evaluation, the overall percentage of at-risk behaviors for the service unit for a time period provides an indication of the safety of worker practices targeted by CAB–CRZ and CAB–Switching. If CAB is effective at improving CRZ or switching practices, then the percentage of at-risk behaviors should, on average, decrease over time at SASU. Because of concerns about the confidentiality of the data, the steering committee provided only monthly aggregated data to the evaluation team. Analyses for time periods less than a month were not possible.

The evaluation team coordinated with the steering committee to create a method to assess the psychometric soundness of CAB's feedback session data of safe and at-risk behaviors by analyzing the worker ratings of videotapes used in training and coaching. At the request of the evaluation team, CAB–CRZ trainers showed one of two videos at the end of training. Each scripted video portrayed a typical mix of safe and at-risk behaviors for five CRZ events. The steering committee specified the "standard observation" of safe and at-risk behaviors for each event. The same videos were used for the coaching of many samplers sometime after training, but during coaching each sampler saw a different video than that seen during training.

Having members of the same population rate the same videos yielded data that allowed the evaluation team to calculate measurements for consistency. Table 3 lists these measures.

Reliability is necessary for actual changes in practices to be statistically detectable with CAB feedback session data.

Bias itself is not a concern for the evaluation, as long as it is consistent to allow month-to-month comparisons.

Time drift represents a systematic tendency for the same practices to be rated as safer (positive drift) or more at risk (negative bias) in later than earlier training. It could result from changes in training or in the kinds of workers trained over time. As the implementation progresses, there are proportionally more workers with later than earlier training, so a positive time drift may produce an illusion of improvements in practices when CAB feedback session data are used. Similarly, negative time drift may mask improvements in actual practices.

Table 3. Measures of Consistency for CAB Feedback Session Data

Measurement	Calculation
Reliability	Percentage of agreement between a peer-observer and the standard observation for each CRZ event.
Bias	Difference in percentage of at-risk scores between peer-observers and the standard observation for each CRZ event.
Time drift	Changes in bias over time, as indicated by the correlation of bias with the date that the video was rated.
Retest drift	Changes in bias by comparison of ratings at the end of training with those during coaching.

Retest drift represents a tendency for the same practices to be rated as safer (positive) or more at risk (negative) relative to how long ago an individual worker was trained. It could result from a systematic tendency for workers to forget certain things learned in training or from influences associated with the experience of conducting feedback sessions. As the implementation progresses, the average time since training increases, so retest drift may produce an illusion of improvement or mask actual improvement, similar to time drift.

5.4 Corporate Safety Data

The CAB steering committee provided feedback session and process metric data, whereas UP provided corporate safety data from its own safety, training, and discipline tracking systems. Unless stated otherwise, all data extend from January 2001 (the beginning of the baseline phase) through November 2007 (the end of the second phase of the evaluation period).

5.4.1 Field Training Exercises

UP's implementation of FRA-mandated field operations training comprise field-training exercises (FTX), during which managers observe workers on the job and record compliance with operating safety rules as either pass or fail. The percentage of passes relative to failures is a measure of the degree to which worker practices conform to the safety rules. The FTX program was initiated at UP in March 2003, and data collection began from that point.

With input from the CAB steering committee, all of UP's operating safety rules (Union Pacific, 2005) were classified into three categories:

- *CRZ rules.* One long rule related to practices in CRZ conditions, (Rule 1.47).
- *Switching rules.* Fifty-nine rules related to the behaviors targeted by CAB–Switching.
- *Non-CAB (Other) rules.* Rules unrelated to any behaviors targeted by CAB.

If CAB–CRZ is effective, there should be a decrease in the percentage of failures of CRZ rules relative to comparison groups from the baseline phase to the first and second phases, with the latter corresponding to the duration of the CAB–CRZ implementation during the evaluation period. If CAB–Switching is effective, there should be a decrease in the percentage of failures of Switching rules relative to comparison groups from the baseline and first phases to the second phase, with the latter corresponding to the duration of the CAB–Switching implementation during the evaluation period.

Because the frequency of FTX tests in general has increased since the inception of the program in 2003, the percentage of failure of FTX tests was calculated for aggregates of a fixed number of tests rather than a fixed time interval, such as months, so that all data points represent aggregates of equal weight.

In a given FTX event, a manager may conduct multiple tests to evaluate multiple workers—generally an entire train or switching crew—on multiple rules. To calculate the percentage of FTX failures for a given set of FTX tests, FTX events were grouped, totaling approximately 6,000 tests per group. This figure represents a compromise between maintaining a large sample size with a relatively fine level of detail and minimizing cases of missing data, which happens if a group has had no tests for a category of FTX rules at a service unit. For each group of 6,000, the tests were divided into subgroups by service unit and test category. The percentage of failures was calculated for each subgroup. Each percentage constitutes a single data point for analysis.

This process makes any two data points from different service units independent. However, data points for different test categories within a service unit are not independent because a given FTX event at a given service unit may include tests from more than one category. Repeated-measures analyses are necessary for comparing the percentage of failures between test categories at a single service unit.

The number of tests for Switching and Other rules were an order of magnitude higher than those for CRZ rules, which represent a smaller number of total rules at UP. As a result, the number of tests aggregated in each subgroup of CRZ rules was smaller than that for Switching or Other rules, causing it to have substantially higher variance.

5.4.2 Injuries

Injury data provided by UP for this evaluation include only FRA-reportable injuries, that is, those requiring medical attention beyond first aid. Far more injuries are associated with switching than with CRZ operations, so if CAB–Switching is effective in reducing occurrences, injury rates associated with switching should decrease in the second phase after the initiation of CAB–Switching.

Following railroad practice, the evaluation used worker-hours to control for differences in the exposure to the risk of injuries at different locations. For example, injury rates for were calculated as the number of injuries per 200,000 worker-hours at a location. Because worker-hours were not available for locations smaller than a service unit, injury rates could not be calculated and compared for individual yards within SASU.

5.4.3 Incidents

Incidents are events involving moving train cars and railroad-property damage. These events include collisions, fires, and derailments, with the latter accounting for the large majority. Data provided for this evaluation by UP include both FRA-reportable and nonreportable incidents, with the difference being the cost of the incident. UP provided data only on incidents attributed in a UP investigation to an error by a transportation worker. Incidents attributed to mechanical faults, such as a rail breaking, were not included in the data.

The vast majority of incidents are associated with switching rather than with CRZ operations, so if CAB–Switching is effective in reducing occurrences, incidents should decrease in the second phase after the initiation of CAB–Switching.

Following railroad practice, the evaluation used car-moves to control for differences in the exposure to the risk of incidents at different locations. For example, incident rates were calculated as the number of incidents per 100,000 car-moves at a location. Car-moves are the number of cars moved through a location. UP provided monthly data for cars moved for each station, allowing comparisons among yards within SASU as well as between SASU and other service units.

5.4.4 Engineer Decertifications

A locomotive engineer can be decertified or lose FRA authorization to run trains due to a serious safety violation. Table 4 lists possible violations.

Table 4. Safety Violations Resulting in Engineer Decertification

Violation	Description	CRZ-Related?
Stop	Passing a stop signal.	Yes
Main track authority	Moving outside an authorized stretch of track, as in when operating under track warrants.	Yes
Speed	Exceeding train or track speed limits by more than 10 mph.	Yes
Brake	Failing to conduct an air-brake test prior to departure.	No
Tampering	Interfering with the operation of a locomotive's alerter mechanism.	No
Drug and alcohol	Testing positive for alcohol or prohibited drugs while on duty.	No

As shown in Table 4, three kinds of decertification are associated with CRZ practices because all three often result from a loss of crew attention under CRZ conditions. If CAB–CRZ is effective in reducing occurrences, CRZ-related decertifications should decrease beginning in the first phase after the initiation of CAB–CRZ.

UP provided decertification data for all violations through February 2008. Like injury data, worker-hours were used to control for differences in exposure to the risk of decertification at different locations.

5.5 Analysis Strategies for Corporate Safety Data

5.5.1 Step Comparison and Gradual Trends

In theory, a safety process such as CAB has an impact on safety shown by the solid-gray test-data line in Figure 9.

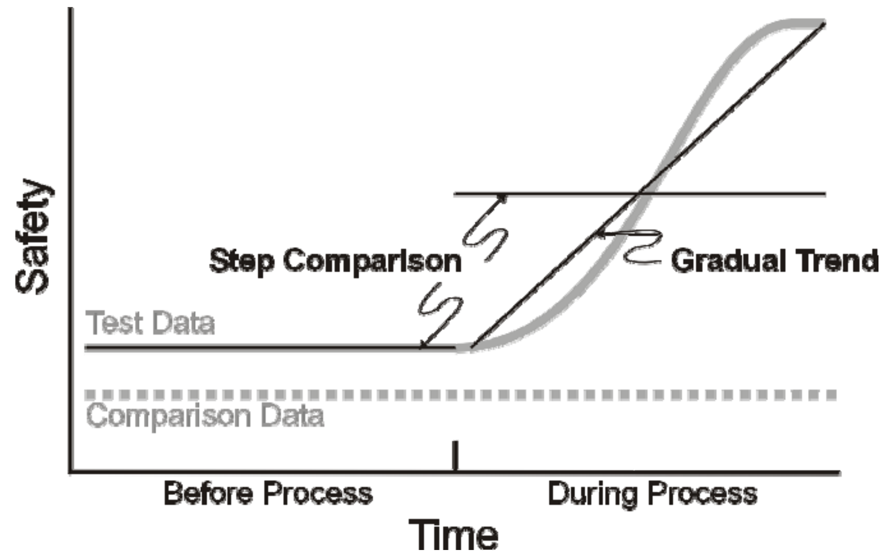


Figure 9. Schematic Representation of Safety Process Outcome and Analyses

Safety is stable during baseline, prior to initiating the process. After initiation, safety measures follow an S-curve on average, with greater outcomes realized as the process matures (e.g., personnel are trained), at which point the process has an asymptotic effect on safety. The shape of the S-curve is unknown, but a change in safety associated with the safety process can be inferred by two analyses illustrated by the thin black lines in Figure 9:

- *Step-comparison,* in which the safety before initiation is compared with the safety after initiation. Following the convention for time-series analyses, this evaluation acquired data from four years prior to initiation for most data sets. (The baseline phase begins in January 2001.)
- *Gradual trend,* in which the relation of safety with date is determined since the initiation of the process.

Significant results for either analysis are consistent with the safety process affecting safety. If both analyses are significant, the argument for the safety process affecting safety is strengthened. However, depending on the exact shape of the S-curve combined with the small sample size and noisy distributions typical of safety data, such as occurrence data, it is certainly possible for one analysis to have significant results while the other does not, when the process does, in fact, have an effect. In this report, both analyses are performed for each source of corporate safety data.

Because of the phased implementation of CAB, different dates are used to represent the initiation of the process. CAB–CRZ existed through both the first and second phases, so if CAB–CRZ is effective in improving CRZ safety, it should be apparent from data for the first and second phases. CAB–Switching began with the second phase, so if CAB–Switching is effectively improving switching safety, it should be apparent in the second-phase data.

5.5.2 Comparison Data

Because this is a field study, strict experimental control is not possible, and there is always the possibility that confounds, such as other safety programs introduced at the same time as the CAB process could be the true cause of changes in safety. However, by comparing the CAB pilot data with another non-CAB-related data in the same region and under the same management, we can observe whether the effect is unique to the CAB site. In theory, we should observe an effect in the CAB data but not in comparison data (Rossi et al., 1999; Shadish, Cook, and Campbell, 2002). The ideal performance of such comparison data is shown by the gray-dashed line in Figure 9, where safety remains constant through implementation. (Comparison data may be offset above or below the test data, as shown in Figure 9, because of systematic differences in the data that are irrelevant to the evaluation.)

There are two forms of comparison data:

- *Different locations,* such as other service units or other yards, where the safety process is not implemented.
- *Different categories* of data that the safety process does not target, such as safety rules or decertification violations not associated with the process.

With comparison data available, the chief statistical analysis is the relative performance of the test and comparison data, that is, the presence of a statistical interaction, whether step changes or in gradual trends. If the changes in safety are different for the test and comparison data, the implication is that some factor associated with only the test data, such as the safety process, is specifically affecting the test data. In contrast, if there are no differences in safety changes, there is a strong possibility that a single factor other than the safety process is responsible for changes in both the test and the comparison data.

The pilot demonstration at SASU provides several opportunities for comparison data. Data for FTX tests and decertifications relevant to CAB can be compared with data not relevant to CAB. For CAB–CRZ, data at SASU can be compared with data at other service units in the Southern Region. For CAB–Switching, comparison with the other three service units is problematic because two of these units had similar safety programs coinciding with SASU's program. However, comparisons within SASU can be made among yards that varied in their implementation strength of CAB–Switching during the evaluation period.

5.5.3 Testing for Trend Continuation

As a field study without experimental control, there may be uncontrolled pre-existing long-term factors that improve safety at a given site. For example, a service unit may have a unique safety process established prior to CAB that has been gradually improving safety. If such factors exist, then an improvement in safety at SASU since the start of

CAB, whether it a gradual or step improvement, may not necessarily be due to CAB, but instead may be due to a continuation of the trend created by pre-existing factors.

To test for the presence of such pre-existing factors, the baseline data for all corporate safety data was analyzed for gradual trends. The presence of a significant trend in the baseline data suggests an underlying factor in safety that may continue to operate during the CAB implementation, casting doubt on any impact CAB may appear to be having.

5.5.4 Occurrence Data Analysis

5.5.4.1 Gap Times

The occurrence data in this evaluation were analyzed as *gap times* (Cook and Lawless, 2010) to determine the occurrence rate changes associated with CAB. With gap times (also known as *inter-arrival times*), each data point represents the time between each pair of adjacent occurrences. "Time," in this case, is expressed in units that represent the site's exposure to the risk of an occurrence. For example, for incidents, the number of car-moves completed between each pair of incidents is calculated whereas for decertifications, the calculation is for the elapsed worker-hours completed between each pair of locomotive engineer decertifications. Figure 10 shows the gap times for four hypothetical decertifications. For instance, 4,800 worker-hours of exposure elapsed at a site for the decertification occurrence on April 23.

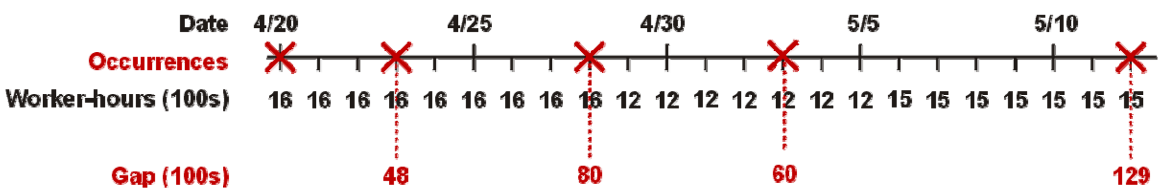


Figure 10. Example of Gap Times

Gap times are the mathematical inverse of a rate. For example, if 4,800 worker-hours were completed between two decertifications, then there was 1 decertification per 4,800 worker-hours for that time period. In terms of a decertification rate per 200,000 worker-hours, it is 200,000 × 1 / 4,800 or 41.7 decertifications per 200,000 worker-hours. Thus, the more work completed between occurrences, the lower the rate of occurrences per unit of work, and vice versa. For instance, from May 4 through May 12 in Figure 10, there was 1 decertification per 12,900 worker hours, for a rate of 15.5 decertifications per 200,000 worker-hours. As the gap times increase on average in Figure 10, the rate of decertifications on average decreases. Exposure data provided by UP are aggregated by month, whereas occurrences are known to the day, so linear interpolation is used to estimate the gap times for each pair of occurrences.

Because the gap times are calculated as elapsed exposure rather than calendar time, the data are corrected for different levels of exposure (e.g., worker-hours) at different places and times. For example, in Figure 10 the decertifications on April 28 and May 3 each occurred 5 calendar days after the previous decertification. However, because the intensity of work was higher before April 28 than May 3 (1,600 versus 1,200 worker-hours per day), the April 28 gap time (8,000 worker-hours) is greater than the May 3 gap time (6,000 worker-hours). The difference in the gap times thus reflects the fact that there was greater risk exposure before the April 28 decertification than before the May 3 decertification.

Statistical analysis of gap times and other analyses of elapsed time until an event are routine in medical research (Lee and Wang, 2003; Clark, Bradburn, Love, and Altman, 2003) and reliability engineering (Department of Defense, 1996). For example, mean time between failures is a familiar statistic from reliability engineering. However, analysis of railroad safety data does not traditionally use gap times. Instead, data points are typically the event rates for a convenient block of time, such as a month or year. For example, FRA (2006) uses such "block rates" in the form of the number of occurrences for each month, divided by the exposure (e.g., worker-hours), and multiplied by a constant (e.g., 200,000).

For any reasonable data set, block rates and gap times will descriptively show the same changes. Furthermore, for large numbers of occurrences (e.g., dozens per month), analysis of block rates is generally adequate. However, block rate analysis has the following disadvantages for sparse data sets encountered in a single railroad service unit:

- *Lower statistical power.* Aggregating sparse occurrences into blocks of time converts them to a few possible values, which introduces noise into the data distribution. This reduces the information in the data, thus resulting in less accurate estimates and larger standard errors that lower statistical power.
- *Violations of parametric assumptions.* Block rates for sparse data tend to have non-normal distributions with means correlated with variances. This may invalidate assumptions of conventional parametric statistics, such as analysis of variance and least-squares regression, which are typically employed on block rates (e.g., Hale, 2008; O'Toole, 2002).
- *Sensitivity to block parameters.* Results including p-values are affected by arbitrary block parameters such as block size and starting point.
- *Imbalanced weighing of data points.* Block rate statistics weigh each time block equally rather than by exposure (e.g., worker-hours) in each block as it should.

Proper analysis of gap times has none of these disadvantages. Appendix B: Analysis Block Rates versus Gap Times on page 183 provides a full explanation of the problems with block-rate data for sparse data sets.

5.5.4.2 Normalized Cumulative Incidence Plots

In reliability engineering, time until occurrences, including gap times, are typically graphically represented with cumulative incidence plots (Nelson, 2003; Cook and Lawless, 2010). In such plots, the cumulative number of occurrences is plotted against the cumulative amount of exposure. This provides a means to visually assess trends which are difficult to perceive in the raw data. Figure 11 shows the cumulative incident plot for the occurrences in Figure 10.

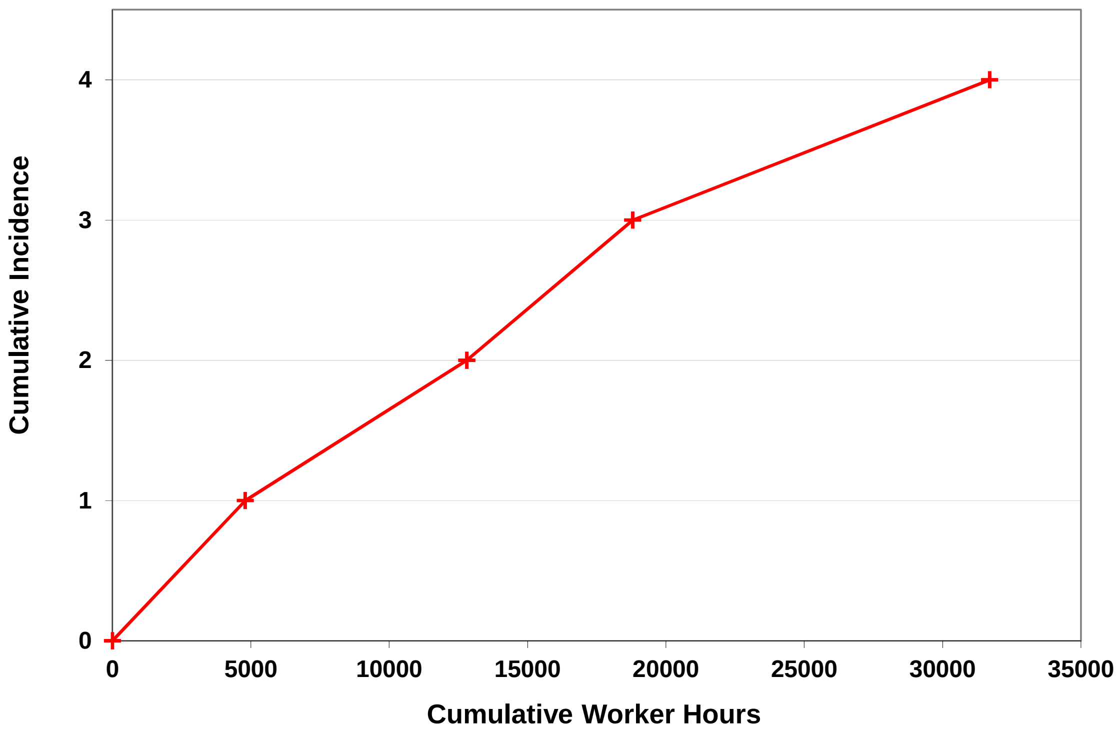
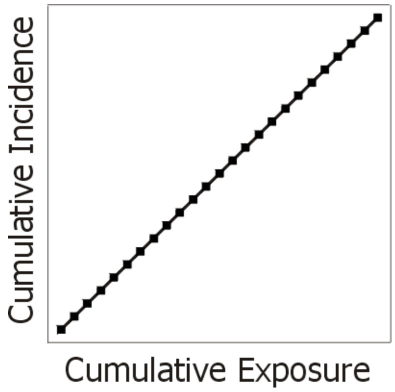


Figure 11. Cumulative Incidence Plot for Hypothetical Data in Figure 10

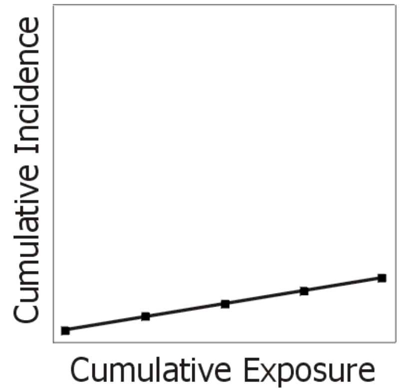
Variations of the cumulative incidence plot include plotting the proportion, rather than the number, of occurrences, and plotting the cumulative incidence on log-log paper (i.e., a Duane plot, see Department of Defense, 1996). Such plots provide detailed descriptive information about the trends of occurrences. This information may be used to plan inferential statistical tests and interpret their results.

By plotting the *cumulative* incidence of occurrences, cumulative incidence plots always have an upward trend. However, the *steepness* or geometric slope of the cumulative incidence plot is equal to the occurrence rate at a given point in time or cumulative exposure. For example, the cumulative incidence plot in Figure 12(a) represents an occurrence rate six times greater than that in Figure 12(b). A zero occurrence rate would appear as a flat line. (Unlike the notional depictions in Figure 12, cumulative incidence plots of actual occurrences show random deviations from the general trend.)

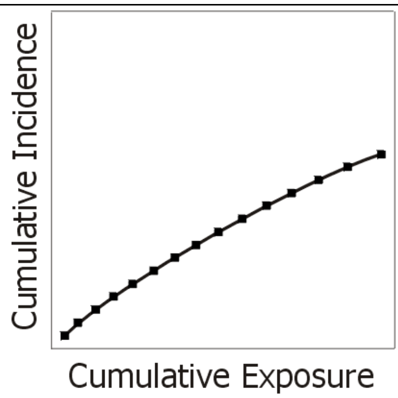
A generally straight line trend in the plot, as shown in Figure 12(a) and (b), indicates a constant occurrence rate, and, by implication, there was no impact of a safety process introduced amidst the occurrences. In contrast, a cumulative incidence plot that curves downward indicates gradually decreasing rates. For example, Figure 12(c) represents a rate decreasing by a factor of three, or a 67-percent decrease in the rate (Figure 11 also represents a decreasing trend on average). If the plot curves upwards, such as in Figure 12(d), rates are gradually increasing over time. In either case, the data may be quantitatively tested for a gradual trend in the gap times. An “elbow” in the plot indicates a step change, prompting a step comparison. For example, Figure 12(e) represents a step decrease of 67 percent (or one third as high), while Figure 12(f) represents a step increase of 200 percent (or three times higher).



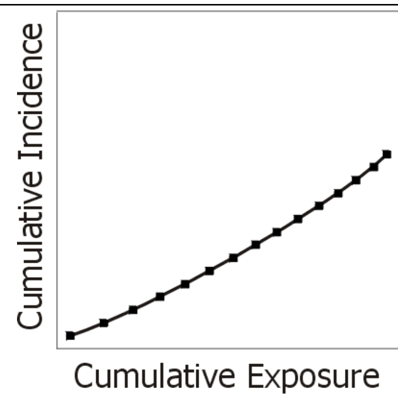
(a) Constant High Rate



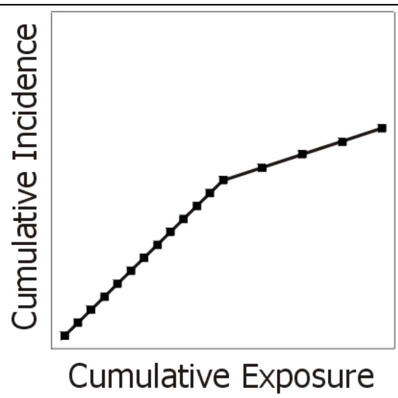
(b) Constant Low Rate



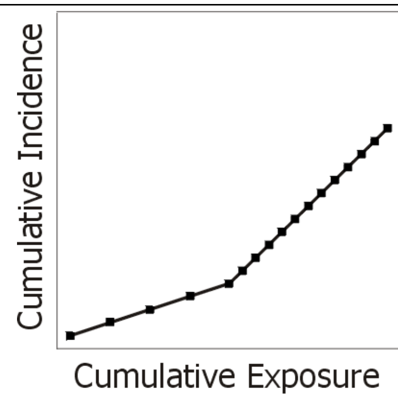
(c) Gradually Slowing Rate



(d) Gradually Accelerating Rate



(e) Step Slowed Rate



(f) Step Accelerated Rate

Figure 12. Interpretation of Cumulative Incidence Plots

For comparing across railroad locations, such as service units, conventional cumulative incidence plots (e.g., against worker-hours) are not ideal because different locations may accumulate exposure (e.g., worker-hours) at different amounts per day. Locations should be compared on given calendar days rather than on given cumulative exposures because corporate or industry-wide confounds should be tied to the calendar, not to exposure.

To compare two cumulative incidence plots for the same calendar time, the x axis of the plot must be the date. To account for differences in exposure across locations and time, the cumulative incidence is normalized by the work per day for each gap time. That is, instead of the cumulative incidence value on the y axis being incremented by 1 for each occurrence, it is incremented by the following:

$$i_j = K\, d_j \,/\, w_j$$

Where i_j is the increment, d_j is the elapsed days in the time gap, w_j is the elapsed work completed in the time gap, and K is a constant to convert the increment to convenient units (e.g., 200,000 for worker-hours). This creates a *normalized* cumulative incidence plot. Figure 13 shows the normalized cumulative incidence plot for the data in Figure 10.

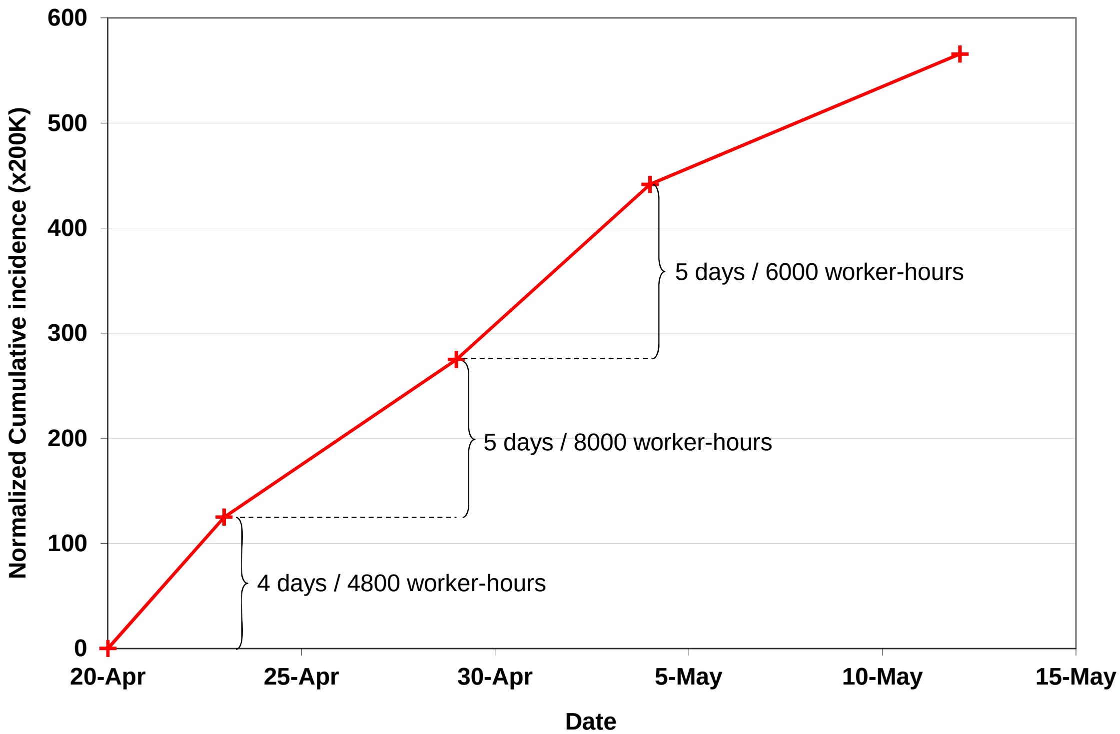


Figure 13. Normalized Cumulative Incidence Plot for Data in Figure 10

By normalizing the cumulative incidents by the work density, the slope of the normalized cumulative incidence plot remains equal to the rate of occurrences. The occurrences and rates are now plotted against dates allowing comparisons across sites with different amounts of work per day. Changes in the slope of the plot are interpreted the same as with conventional cumulative incidence plots.

5.5.4.3 Independence of Gap Times

Most statistical analyses, including those employed in this evaluation, assume statistically independent data points. For the occurrence data, independence implies that the length of one gap time is not related to the length of any other gap time. This may be a reasonable expectation in this evaluation. For example, that one derailment occurred shortly after another generally does not change the underlying conditions that cause derailments (e.g., track conditions, worker training) any more than would one derailment occurring a long time after another. Thus, the gap time of a subsequent derailment is independent of gap times of the previous derailments.

However, if gap times *were* to lack independence, most likely adjacent or nearly adjacent gap times should be most strongly related. For example, one can imagine a site's series of closely spaced derailments resulting in extra caution that extends the gap time for the next derailment. Alternatively, some underlying transitory variable that affects derailments, such as weather, can result in strings of closely spaced derailments (e.g., during a month or two of poor weather) followed by strings of widely spaced derailments (a month or two of good weather).

Thus, to test for independence of gap times in this evaluation, lag-1 and lag-2 autocorrelations (Cohen, Cohen, West, and Aiken, 2002) are calculated for the gap times of each occurrence variable for each site. Because gap times tend to be approximately exponentially distributed (as discussed in Survival Analysis of Occurrence Data on page 53), the correlations are performed on logarithmic transform of the gap times. A logarithm transform yields a relatively symmetrical distribution more suitable for conventional correlations (Winer, Brown, and Michels, 1991). If these autocorrelations are all low ($| r | < 0.2$) and on average not significantly different from zero, then gap times appear to be adequately independent for statistical analysis for gradual trends and step comparisons.

Many statistical analyses, including those employed in this evaluation, also assume the observations are identically distributed, where the all gap times are treated as samples from the same-shaped distribution. It is reasonable to assume identically distributed gap times for the same reason for assuming independent gap times. Except for known factors that are accounted for in the analysis (e.g., the safety process), the underlying factors that cause safety occurrences do not change systematically, and thus the distributions apparently remain the same for all gap times. The assumption of identically distributed gap times cannot be statistically tested in a single site (Cook and Lawless, 2010). However, Weibull probability plots (Nelson, 1982) may be visually inspected for curves, which may result from a mixture of distributions in the data.

5.5.4.4 Survival Analysis of Occurrence Data

Provided good evidence for independence of the gap times, a given series of safety occurrences at a site may be adequately represented as a nonhomogeneous Poisson process, where occurrences accumulate independently within any arbitrary time period (Nelson, 2003). Indeed, one may expect safety occurrences at a site to be consistent with a homogeneous Poisson process, where the probability of an occurrence is constant within each gap between occurrences (Maguire, Pearson, and Wynn, 1952). In this way, safety occurrences are treated no differently than failures among a population of replaceable components or phone calls coming into an exchange (Nelson, 2003).

In a simple Poisson process, the times between occurrences have an exponential distribution, which have a positive skew of 2 and a mean equal to the standard deviation

(Devore, 1991). These characteristics challenge the assumptions of conventional parametric statistics such as analysis of variance and regression (Hays, 1981; Pedhazur, 1982). Certain features in the gap times may cause distributions to deviate from the exponential, most notably transitory spikes or holes of occurrences associated with unknown factors that temporarily change the underlying conditions that produce occurrences. These appear as "waves" or "wiggles" in cumulative incidence plots. However, even when such features are present, the data distributions remain inappropriate for conventional parametric analyses.

Rather than conventional parametric analysis, this evaluation employs survival analysis. Survival analysis comprises a suite of statistical techniques to analyze the effects of variables on the times until an occurrence. These include gap times resulting from a nonhomogeneous or homogeneous Poisson process. Survival analysis techniques are suitable for the analysis of gap times if the gap times are statistically independent and identically distributed (Cook and Lawless, 2010).

Survival analysis is preferred over conventional least squares techniques, such as analysis of variance, because survival analysis:

- Is more capable of addressing the approximately exponential distributions typical of gap times.
- Produces more accurate parameter estimates due the use of maximum likelihood estimation.
- Allows inclusion of a "right-censored" gap time, being the first occurrence that occurs *after* the end of the evaluation period.

This evaluation employs two forms of survival analysis to ascertain the relationship between the safety process and occurrences gap times: Cox regression and Weibull regression (Nelson, 1982; Lee and Wang, 2003). Both regression types estimate the *hazard function* of the gap times, which is the probability function for the chance of the occurrence over the gap given it has not occurred earlier in the gap.

In analysis of gap times, a Cox regression uses the following model for the hazard function of the gap time distribution:

$$h(g) = h_0(g) \exp(\Sigma\ b_i x_i)$$

A Weibull regression uses the following model:

$$h(g) = g^{\alpha} \exp(\Sigma\ b_i x_i)$$

Where:

g = The gap times

$h(g)$ = Hazard function for the gap times

$h_0(g)$= Base hazard function derived from the gap time distribution

α = Shape parameter for the Weibull distribution, determined by fitting the data

b_i = Regression coefficients, determined by fitting the data

x_i = Independent variables (e.g., representing the safety intervention) for predicting the changes in the gaps

For purposes of this evaluation, the difference between the two regressions is that the Cox regression derives a base hazard function from the data distribution, while the Weibull regression fits a Weibull hazard function with shape parameter α. The Weibull regression is suitable for safety occurrence data because the exponential distributions is a special case of the Weibull distribution, specifically, one where $\alpha = 1$. The α will be less than 1 if there are transitory spikes and holes in the incidence rates that may appear due to unknown factors. However, if α is not significantly different from 1, then the analysis is re-run as an exponential regression with α fixed at 1. The suitability of the Weibull regression was assessed by Weibull plots (Nelson, 1982).

The Cox regression is more flexible than the Weibull regression in that it does not fit a specific parametric distribution to the data, but rather nonparametrically estimates the hazard function from the ranked gap times. However, it assumes that the impact of the independent variables is multiplicative – that the hazard functions for different values of the dependent variable are proportional to each other. This assumption can be tested by correlating the Schoenfeld residuals of the regression with the gap times (Bradburn, et al., 2003). If the correlations are low and not significant, then hazard functions across the independent variable values appear to be adequately proportional to each other.

Like any regression, the independent variables in either a Cox or Weibull regression may be continuous or categorical, provided the latter are coded as dummy variables. Interaction effects of two independent variables are tested by adding an independent variable that is the product of two or more other variables.

Also like any regression, the coefficients from Cox and Weibull regressions, b_i, represented the magnitude of the effect. Specifically, for Cox regression $\exp(b_i)$ is the *hazard ratio* for a unit change in the associate predictor variable. That is, it represents the ratio of the relative incidence rates for each integer increment of the predictor variable (Bradburn et al., 2003). Thus, to calculate the percent change Δr in relative incidence rates associated with two values, x_{i1} and x_{i2} of the predictor variable, one calculates the following:

$$\Delta r \text{ (Cox)} = [\ \exp(b_i\, x_{i1}) / \exp(b_i\, x_{i2})\ ] - 1$$

The resulting value is then expressed as a percent. Weibull regression coefficients represent the effect of the predictor variable on the gap time rather than the hazard, but the hazard ratio can be calculated as $\exp(-\alpha\ b_i)$ (Box-Steffenmeier and Jones, 1997). Thus, the change in incidence rates is the following:

$$\Delta r \text{ (Weibull)} = [\ \exp(-\alpha\, b_i\, x_{i1}) / \exp(-\alpha\, b_i\, x_{i2})\ ] - 1$$

For either Cox or Weibull regression models, a negative Δr value indicates the percent reduction in the relative incidence rate, and therefore the improvement of x_{i1} over x_{i2}. A positive Δr indicates the percent increase in the relative incidence rate. For gradual trends, we use the start and end dates of the relevant evaluation phases for the xs to gauge the improvement from the begining to the end of the time period of interest. For step comparisons, the xs represent each step as dummy codes (0 for one step, 1 for the other).

5.5.5 t-Tests with Separate Variance Estimates

Conventional practice uses a 2×2 analysis of variance (ANOVA) to evaluate mean differences in a variable for test and comparison data across two time periods, in which a significant interaction implies different changes for the test and comparison data. However, corporate safety data tend to exhibit heterogeneity of variance across groups, making ANOVA problematic—a condition aggravated by groups often having unequal sample sizes, as this is not under the control of the evaluators in this field study. The heterogeneity of variance may be due to floor effects in the data. FTX failures, for example, usually have very low frequencies, as they are single-digit percentages before CAB is initiated. Any mean improvement presses the distribution closer to the limit of 0 percent, resulting in lower variance.

Nonparametric statistical tests are generally unattractive because the key statistical tests are for an interaction with comparison data, which are not supported in nonparametric tests such as Kruskal-Wallis. Many nonparametric tests also represent a reduction in statistical power, which is problematic given the small sample sizes for some of the occurrence data.

The solution is to use t-tests for separate variance estimates. To evaluate for interactions, the evaluation team performed t-tests on linear combinations of four means, corresponding to the four cells of a 2×2 ANOVA. Table 5 lists coefficients for the linear combination.

Table 5. Coefficients for a Linear Combination to Compare Changes over Time

Time	**Data Source**	
	Test	Comparison
Before process	–1	1
During process	1	–1

With these coefficients, the null hypothesis is equivalent to "the difference in the changes of the means for test and comparison data is zero." The variance for the sampling distribution of the linear combination is the sum of the variance of the sampling distributions of the means, assuming they are normally distributed.

Analogous coefficients, using t-tests with separate variance estimates for main effects of data source and time, are also available.

T-Tests of this type were used for FTX data analyses comparing means in this evaluation.

5.6 Forced-Choice Survey

A forced-choice survey included measures of various aspects of practices and safety culture, specifically self-reported work practices, Internalization and Generalization, and Labor-Management Relations. Table 6 describes the scales, along with their sources and supporting research, and the outcome in the logic model to which each is related.

Table 6. Forced-Choice Survey Scales and Their Relation to the Logic Model

Scale	Description	Number of Items	Item Source	Outcome
Unsafe Behaviors–CRZ	Self-reported extent to which employees follow CRZ rules and procedures*	6	UP Code of Operating Rules, checked by UP subject-matter experts	CRZ practices
Unsafe Behaviors–Switching	Self-reported extent to which employees follow key Switching rules and procedures*	6	UP Code of Operating Rules, checked by UP subject-matter experts	Switching practices
Propensity to Safe Behavior	Workgroups' propensity to comply with safety rules	3	Simard and Marchand (1997) modified to include worker version along with manager version	Internalization and Generalization
Coworker Safety	Norms of workers around compliance and safety	4	Mueller, DaSilva, Townsend and Tetrick (1999)	Internalization and Generalization
Labor–Management Relations	Perceived cooperation between labor and management	6	Dastmalchian, Blyton & Adamson (1989)	Culture/ relations

*** Conceptually based on Hofmann and Stetzer (1996).**

For scales on practices and Propensity to Safe Behavior, managers reported on their subordinates, whereas workers reported for themselves. For the Coworker Safety scale, workers reported on coworkers.

The scales are derived from those associated with safety-related outcomes in prior implementations and research. For example, the Coworker Safety scale is associated with safer work practices (Hofmann and Stetzer, 1996) and fewer accidents (Hayes, Perander, Smecko, and Trask, 1998).

These five scales were added to 11 proprietary scales in an organizational climate survey used by the BST consultant for evaluating its implementation sites. Although the scales were based on existing scales in the organizational safety literature, each was evaluated for inter-item reliability using Cronbach's alpha because this is the first time that such scales were used in the railroad industry.

SASU management distributed the survey by mail to workers and managers on two occasions: the start of the first phase (in December 2005) and the end of the second phase (in April 2008). Table 7 lists numbers of respondents and return rates.

Table 7. Numbers of Respondents for the CRZ Scale for Each Phase and Respondent Type

Phase	Respondent Type*	
	Worker	Manager
First	179 (19%)	26 (52%)
Second	86 (9%)	26 (52%)

***Return rates are in parentheses.**

The evaluation team compared distributions of the self-reported years of service and ages of the respondents with actual respective distributions from data provided by SASU Human Resources. There were no significant differences, implying that respondents were representative of the workers and managers at SASU.

If CAB has an effect as predicted, improvements in average scale scores should be seen from the first to the second phase.

The Unsafe Behaviors–Switching scale was developed after giving the first survey. For the second survey, respondents were asked to answer questions 1) about their current practices and 2) retrospectively to "what degree did you engage in the following practices before the CAB process?" Changes in practices were indicated by a difference between what respondents reported currently and retrospectively. Respondents were asked to skip the Switching Practices scale if they had not worked in a yard that used the CAB process. Similarly, yard employees were asked to skip the Unsafe Behaviors–CRZ.

5.7 Interviews

Periodic interviews (Patton, 2002) provided qualitative data concerning perceptions of the implementation and outcomes. The interviews were confidential and were conducted one respondent at a time. Respondents were assured that their identities would not be released. Interviews were conducted three times, in the winters of 2005–2006 (initial interviews), 2006–2007 (midterm), and 2007–2008 (final).

The interview questions were tailored to obtain perspectives on the state of CAB's implementation and support, safety, and labor-management relations. Respondents were asked about any outcomes attributed to CAB and were probed on the role of CAB with regard to any reported changes to safety and relations. The interviews were open-ended and semistructured, in which the questions allowed respondents to reply however they wanted. In addition, the questions changed to follow the each respondent's unique thoughts. Thus the respondents were not all asked the same questions.

To obtain diverse viewpoints, the evaluation team interviewed workers and managers, BLET and UTU members, yard and road workers, and workers at various levels of involvement in CAB, including steering committee members and workers both trained and not trained in CAB. Participants included respondents who were both supportive and skeptical of CAB. Table 8 lists the types and numbers of respondents.

Facilitators and the local management sponsor were interviewed on all three occasions. CAB facilitators and top management helped identify prospective respondents by selecting people whom they regarded as respected, credible, and prosafety.

Table 8. Respondents to Periodic Interviews

Respondent	Number of Respondents		
	Initial (2005–2006)	Midterm (2006–2007)	Final (2007–2008)
CAB Steering Committee	4	4	4
Trained CAB samplers	5	4	3
Workers untrained in CAB	5	3	3
Management			
Frontline/middle	3	3	4
Upper	2	2	1
Corporate/regional	0	2	1
Total	19	18	16

The evaluation team analyzed interview data by breaking the responses into comments, then identifying and sorting the comments into themes representing the most frequent kinds of comments related to overall safety issues, CAB, and labor-management relations (Miles and Huberman, 1994; Patton, 2002). Two or three team members sorted comments separately into themes and then reviewed each other's work to arrive at a consensus on the themes.

In May 2008, an additional set of interviews was conducted to investigate the process of personal change because of CAB.[6] For these interviews, engineers, conductors, and managers at SASU were asked about their personal experience with safety and how that might have changed over the last few years. Especially important was whether or not they had made specific changes in behavior, and if they could relate those changes to CAB. Discussions with managers also focused on decisions that they make in their job role and how safety affected those decisions.

Respondents for the personal change interviews were identified by purposive and snowball sampling (Crano and Brewer, 2002). CAB facilitators in SASU purposefully identified workers and managers whom they recognized as having experienced a change related to safety over the last few years. During interviews, the evaluation team also asked participants if they could identify someone else who had experienced change (snowball sampling); these individuals were subsequently contacted and interviewed. An effort was made to ensure that even numbers of engineers and conductors and of senior (15+ years) and junior (5–10 years) workers were selected. Table 9 lists types and numbers of respondents by years on the job.

[6] The evaluation team also conducted bimonthly, confidential, unstructured interviews of key stakeholders to remain abreast of events at SASU. These interviews were valuable for formative purposes but were not systematically analyzed because summative findings were captured by other methods.

Table 9. Respondents to Personal Change Interviews

Respondents	Number	Years in Job
Engineers	8	4 to 38
Conductors	6	2 to 40
Managers	2	10 to 15

For additional information on these interviews refer to Appendix A: Changing At-Risk Behavior (CAB) SASU: Individual Stories of Change, on page 162.

6. Results

6.1 Field Notes and Project Records

6.1.1 Initial Conditions

6.1.1.1 Traditional Railroad Culture before Accidents and CAB

Initially, the culture at SASU was much like that found in the railroad industry in general, comprising a reactive and disciplinary management practicing command-and-control leadership that exacerbated the already adversarial labor-management relations.

"Management tends to react, not plan ahead more than a quarter," claimed one local union officer. FTX was a particularly contentious issue between labor and management. "FTX is supposed to be training, but has become punitive," said one worker prior to the initiation of CAB. Under UP policy, whereas a worker's failure on an FTX test would not necessarily lead to formal discipline, it would result in points being deducted from the worker's Employee Development Review score representing the company's general evaluation of that worker. As another worker observed around the same time, "There's a giant wall between management and labor. I've given up trying to break it down."

Distrust ran deep on both sides. Management tended to see workers as "thieves." One union leader gave an example of a supervisor assuming that a worker's reported problem with an engine was a lie told to avoid work. Another union leader said workers, for their part, were taught that management was "the enemy." There was little faith in the value of cooperative safety programs among workers, owing to previously attempted programs falling away soon after initiation. "I've seen many safety programs come here before. It's a feeling of frustration from being burned by prior programs in which management doesn't follow up," said one union official. "Management has done so much that I don't trust them to do anything good. There's bad baggage from previous safety programs," said a worker, after the initiation of CAB.

Each side saw the other as being responsible for safety, with management blaming workers for unsafe behavior and workers blaming management for unsafe working conditions. Meanwhile, worker at-risk practices were the norm, aided by management's willingness to ignore at-risk practices in the name of productivity. "It's the path of least resistance," explained a union leader. Another added, "It's traditional―the way it has always been done."

6.1.1.2 Accidents Motivate Change

Rooted in traditions going back more than a century (Gamst, 1980; Ambrose, 2000), the culture at SASU was suddenly disrupted by traumatic events. On December 7, 2003, a foreman was operating coupled remote-controlled locomotives at East Yard in San Antonio on his own. After activating a switch to send the locomotives down one track, the foreman turned his back on the locomotives and walked across another track. However, the switch failed to change, and the locomotives rolled down the track the foreman was on, running him over and killing him (NTSB, 2004). Then on June 28, 2004, only 11 miles from the main offices of SASU, the fatal accident at Macdona occurred, claiming the lives of an SASU conductor and two others (see Prologue, page 7). On November 10 of the same year, there was another fatal accident in San Antonio, this time at the Crystal Cold Storage Warehouse, which occurred when the engineer lost radio contact with the conductor who was outside directing the engineer to back up 29 cars on an industrial track. A car was pushed off the tracks and into a maintenance building, killing a non-UP worker (NTSB, 2005b).

The deaths sent shockwaves through the service unit and beyond. The service unit constituted a closely knit social group (e.g., the conductor killed at Macdona was the son of a SASU engineer), so one tragedy emotionally impacted many individuals. According to one union officer, these fatalities were among a string of six catastrophic switching and road work accidents at SASU that alarmed everyone involved with the service unit. By December 2004, the FRA Office of Railroad Safety had increased the number of inspections at SASU, concluding "that crew compliance with railroad operating rules is not satisfactory, and that sustained effort will be required to ensure both that the FTX Program is functioning as intended and that employees are responding appropriately." In an effort to improve safety, FRA placed SASU and two other service units in the Union Pacific Southern Region under a compliance agreement, which required the service units to overhaul their FTX programs. According to one local union official, UP received substantial negative publicity, and some members of the public were pressuring UP to move railroad tracks out of San Antonio. "UP is a joke around town, to this day" (Spring 2005), said the union official. "Working for UP used to be a source of pride." Another union official announced, "We've got to do something. We can't have this anymore."

Management likewise saw the need for change. The operating safety rule governing the duties of road crews (Rule 1.47) was expanded to define CRZ conditions and responsibilities. On November 18, 2004, UP corporate management named a new superintendent to head SASU and charged him with improving safety performance. There was a general recognition that safety rules were not consistently followed. A high-level SASU manager brought in by the new superintendent stated, "I heard when I first got here: 'I know what the rule says, but I don't believe in it.'" The new superintendent reported, "The same behaviors that caused an accident were overlooked by supervisors since no one got hurt before. This produced a culture of risk, unspoken rules.… Omission equals permission…. But this is changing now." Union leaders also saw improvements. "Past management had a 'get it done' attitude, but that has changed with the new management team," said one leader.

This common goal of management and labor to urgently improve safety led to a new willingness to cooperate. Two days after the Crystal Cold Storage accident, SASU opened a Safety Command Center to provide safety briefings to workers before they started work. A joint, labor-management Crew Resource Task Team, which included union leaders from both BLET and UTU, was formed to review safety at SASU and identify ways to prevent accidents. Among the approaches considered were the following, most of which UP has since enacted to various degrees:

- Crews to call signals on radio.
- Dispatch to brief crews on what to expect.

- More prompt removal of workers from trains when the workers reach their hours-of-service limits.
- More advanced warning for call-ins for road work.
- Implementation of a close-call reporting system.

The Crew Resource Task Team also developed a checklist for road crews to use in the CRZ. This checklist was based on the revised Rule 1.47 of intra-crew communications and was intended to improve attention to signals and other requirements. The Crew Resource Task Team proposed that each road worker use the checklist to verify his or her safety practices.

The new SASU superintendent recognized that the accidents did much to raise consciousness and spur a change at SASU. "Just by chance, UP got these accidents for doing what all railroads were doing," he noted. A regional manager concluded, "We've had some significant emotional events that helps promote change at SASU."

6.1.1.3 Previous Successes with PPF Known to Key Players

Although accidents motivated workers and managers to cooperate to improve safety, individuals who would become key players in the pilot demonstration were also favorably disposed to CSA as a means to improve safety. The lead NTSB investigator for Macdona was impressed with the labor-management cooperation behind the establishment of the SASU Safety Command Center and contacted the FRA Office of Railroad Development in the summer of 2004 to ask if FRA could encourage more activities in this area.

It happened that the Human Factors Program Manager in the Office of Railroad Development was concluding an FRA-sponsored demonstration of CSA with Amtrak in the Station Service department in Chicago (see Ranney, et al., in preparation, for details). He had also been trying unsuccessfully for two years to find a site to pilot CSA in freight operations, but this had been thwarted by labor and management's distrust of each other and of FRA. The positive findings from Amtrak's demonstration encouraged FRA to pilot CSA at the SASU transportation department, where management and labor had expressed willingness to try new cooperative approaches to safety. FRA Railroad Development's promotion of CSA gave the demonstration process credibility with workers and managers that it would not have had if it had been a strictly UP initiative. Because of FRA's backing, it appeared less likely that management would stop the process prematurely.

In contrast to the traditional adversarial railroad mentality described in Traditional Railroad Culture before Accidents and CAB, on page 61, the new SASU management team was predisposed towards cooperating with workers, recognizing that more discipline did not necessarily result in more safety. Formerly, an injured worker would automatically be charged with a rule violation. That practice became less common when the new superintendent arrived, according to a union leader, although this escaped the immediate notice of some workers.

Some of the management team had direct positive experiences with CSA-type processes. The new superintendent had worked in the mechanical department on the North Platte service unit of UP, where PPF and CI had been implemented in a process known to UP as Total Safety Culture (TSC).

This experience convinced him that CSA would be effective and sustainable. “Every place I’ve seen [PPF] implemented, it has worked, and every place I’ve seen it implemented, they’re still doing it. How many programs can you say that about?” he said. Nonetheless, trying out PPF on the transportation department at SASU was daring given the prevailing confrontational culture and other challenges. A colleague of the superintendent related, “He was willing to bet the farm on this new CSA approach.”

Even before the development of CAB’s checklists for use in feedback sessions, one of the facilitators noted that top-level SASU managers were very committed to CSA. These high expectations about CSA spread to the workers. One union officer said, “It has worked in so many places. There’s no reason we can’t make it work here.”

6.1.1.4 New Workers More Amenable to Change

Demand for abundant clean coal from the Powder River Basin of Wyoming brought a surge in business for U.S. railroads (McPhee, 2005) at about the same time that the long-time workers (the “old heads”) were starting to retire. The result was a workforce with an increasing proportion of young workers. At the start of CAB at SASU, the superintendent reported that 30 percent of the workforce had less than 5 years’ experience.

Such a large number of inexperienced workers posed a safety concern among those at SASU, but it also introduced new attitudes and expectations about work. “Haven’t been around for a change like this in 20 years,” reported one union official. These newer and younger workers were more inclined to cooperate with managers on new initiatives since they had not been indoctrinated in the confrontational railroad culture, were not jaded from a history of discontinued safety programs, and were not prepared to accept the traditional harsh approach of managers. “The old heads may be slow to change,” said the union official, "but the influx of new ones—the greenhorns—are more ready.” This included union officials themselves, as well as the rank and file. “There are new people in union leadership, which makes such participation in CAB possible,” said one union official. Shortly after the creation of the CAB-CRZ checklist, the BST consultant noted that the SASU local unions at the site exhibited more commitment to CSA than those at most other sites he had worked.

Cooperation appeared to be spreading among workers as well as between the workers and their managers. There were also improvements in conductor-engineer relations. “It used to be we never talk to each other except when necessary,” reported one CAB facilitator. “We used to not stay in the same hotel. Now it’s just jokes about each other.”

6.1.2 CAB Design and Strategy for Implementation

With an established history of working together on safety, members of the labor-management Crew Resource Task Team reconstituted themselves as the Design Team for CAB. Because the transportation workers were represented by two different unions with a contentious history, the Design Team appointed two facilitators (an engineer from BLET and a conductor from UTU), in contrast to the usual BST practice of having a single facilitator. Likewise, steering committee members were selected with an eye toward balancing the number from each union. Given the level of distrust between management and labor, all members of the steering committee were workers. Some members of the Design Team also became steering committee members, allowing them to carry over their experience and relationships from the Crew Resource Task Team.

6.1.2.1 CRZ Selected as Initial Focus, with Switching to Come in Later

Although top management recognized that switching was a significant safety concern, they were adamant that initially the process would focus on promoting safe CRZ practices, since the Macdona accident had raised that issue. "Because CRZ had a recent accident, it's a likely place to get acceptance" from both new and old workers, a union leader agreed. Furthermore, workers needed to be educated about the new CRZ procedures. In the distrustful world of the railroad, focusing on CRZ alone carried some risk of resentment. Road workers might feel singled out for unjustified scrutiny, and yard workers might feel excluded from a process that would improve their job safety. Ultimately, the Design Team agreed that CAB would start with CRZ, but expand to include switching once workers and managers had begun to accept CAB.

The CRZ focus represented a challenge to the usual BST process. Normally, the safe behavior checklist used for feedback sessions with workers would be developed by the steering committee studying local injury and incident reports and identifying behaviors that would prevent the accident. However, although train collisions associated with CRZ operations can be catastrophic, as the Macdona incident demonstrated, they are also infrequent. An engineer estimated that only six incidents in recent years were associated with CRZ problems. Fortunately, the Crew Resource Task Team had already identified behaviors expected to improve safety under CRZ conditions. The steering committee used their knowledge of the Crew Resource Task Team checklists to make the CSA checklist used in feedback sessions by peer observers in CAB–CRZ.

6.1.2.2 Proaction against Worker Resistance Planned

Top managers and the steering committee recognized that many workers would resist CSA, given the history and culture of the railroad. For this reason, CAB training was voluntary throughout the evaluation period rather than mandated by management. The observation of workers for at-risk behavior that is necessary for PPF was inconsistent with the prevailing employee position that working conditions were the primary contributory factor in accidents and injuries. One facilitator admitted, "I would've been very closed-minded once I heard the word 'behavior.'" The stakeholders anticipated that educating workers on the CI component of CSA would counteract such resistance. "We clean our house, and they clean their house," the facilitator suggested.

Similarities to the discredited FTX program were also expected to be a significant source of worker resistance. "Everyone is concerned CAB will turn out like that, i.e., become punitive. Everyone is gun-shy" from the experience with FTX, one steering committee member said. Other steering committee members added, "The main concern is how to convince and reassure the guys out there that it won't come back to them—that there'll be no retaliation." Management likewise recognized the threat. "The key is to be nonpunitive," said one top manager.

Delineating CAB from FTX was attempted by several actions. The word *observation* was associated with FTX, therefore the steering committee selected the word *sampling* to refer to the peer observation and feedback process; those trained to perform feedback sessions were to be *samplers*.

Early in the design process, the BST consultant successfully moved the steering committee away from the traditional railroad view that rule-compliance equals safety and toward the notion that identifying and preventing exposure to risk, irrespective of the rules, would be more successful. This action separated CAB from the disciplinary overtones acquired by FTX, which tests worker compliance with specific rules. The BST consultant advised the steering committee to define at-risk behavior based not the violation of a rule but on the degree to which a behavior increased the likelihood of an injury or accident. “We’re risk involved,” rather than enforcing rules, concluded one steering committee member. A facilitator emphasized that samplers approach peers with an attitude that suggests, “I’m asking you to do this because I don’t want you to be hit by a train.” The BST consultant urged that samplers present themselves as a peers rather than supervisors and refer only to the CAB checklist and its underlying definitions rather than the rules. The stakeholders anticipated that this approach would remove the confrontational aspect of being critiqued. “It makes the person feel like part of the process,” predicted a steering committee member.

6.1.2.3 Acquiring Resources and Managers’ Time a Challenge

Managers were challenged to supply sufficient resources and time during the formative stages of CAB. The traditional reaction-driven nature of the railroad manager’s job made attendance at key meetings difficult because of constant cell phone interruptions and various other demands. A lack of management attention caused the steering committee, BST, and FRA to doubt management’s commitment to CSA. BST representatives expressed concern when no managers attended the meeting to discuss checklist development. It was later learned that a top manager visited the conference room after hours and studied the whiteboard diagrams to acquaint himself with the progress of the steering committee. Local managers did what they usually do, fitting necessary work in wherever they could between daily events. This issue was never fully resolved in the implementation.

Resources were also scarce when setting up CAB. Since there was no available office space in which to conduct CAB business and training, UP had to bring in a trailer office, which took nearly four months to arrive. In part, problems in resource supply arose because local management was not fully aware early on of the scale of resources necessary. “We've learned that one needs to plan for the resources, such as equipment and software, that are needed up front,” a top manager conceded in retrospect. Other problems resulted from the difference between the railroad environment and the factory floors, a more typical BST environment. Workers taking time off—to participate in BST meetings, focus groups, and training—proved problematic, owing to the collective bargaining agreements with the unions. By such agreements, road workers were paid per trip. If a meeting or training caused a worker to miss a trip, the company was obligated to reimburse the worker for all lost pay. Attending a one-hour meeting could cost the company 2 days of pay.

Ultimately, BST coordinated and negotiated with local management to schedule meetings and trainings with minimal impact to schedules. For sampler training, BST scheduled two days, a Thursday and a Friday, for which the company paid workers for eight hours per day. The number of hours was somewhat less than would ordinarily have been granted under the collective bargaining agreement, but management sweetened the deal by guaranteeing workers the following Saturday off work, consistent with other terms of the collective bargaining agreement. The prospect of being able to plan for a weekend day off proved to be a strong motivator and encouraged workers to volunteer for sampler training.

Meanwhile, FRA contacted corporate management and worked out a written commitment plan for supporting CAB for the duration of the CSA evaluation. FRA also established regular Oversight Committee telcons—attended by top SASU and corporate managers, CAB facilitators, union leaders, and the BST consultant—to raise CAB visibility (and thus priority) and to provide a forum for resolving problems. These telcons[7] began in October 2005, about a month after feedback sessions began, and were held every four to six weeks throughout the evaluation period. In spring 2006, the Oversight Committee was expanded to include representatives from a second UP CSA demonstration pilot, this one on the Livonia Service Unit (see Regional and National Level Embraces CAB, page 78). However, by summer 2006, superintendents at SASU and Livonia, neither being the original superintendents from the start of the respective CSA implementations, no longer attended the telcons.

6.1.3 CAB Evolves and Addresses External Factors

CAB evolved during the evaluation period, adjusting to challenges and expanding its scope. During this period, strategies were developed to handle worker resistance, boost contact rates, and address barriers to safety. CAB also resolved cost issues with management and expanded to include outlying areas and switching practices. The role of management became more defined and active.

6.1.3.1 Worker Resistance Overcome through Education

As anticipated by the steering committee, worker resistance was substantial, but the BST training and preparations conducted during CAB implementation effectively managed this. Educating workers about CAB was the key to removing resistance, so steering committee members had one-on-one discussions with skeptical workers. Because there was distrust among workers as well as between workers and managers, face-to-face communication was deemed essential so that the listener could gauge the communicator's sincerity.

Sampler training was also a major form of education, and CAB was never short of volunteers to take the class. In spring 2006, a facilitator and the local management sponsor reported that workers were "knocking each other over" to sign up. A bulletin

---

[7] In addition to helping the implementation, these telcons proved to be a key source of field notes for this report.

board posting of open classes was filled within three hours. In fall 2006, the waiting list numbered over 100 workers, approximately 10 percent of the entire workforce. Some workers came from outlying areas at their own expense to take the training. To meet the demand, class frequency was doubled during seasonal lulls in train traffic, when the competition for manpower was less.

It is not known how many workers volunteered for training out of curiosity about CAB and how many merely wanted to plan a Saturday off from work. However, what is known, based on management and labor leader feedback, is that once training was over positive attitudes prevailed. One trainer related, “Initially, most trainees are skeptical, but by lunchtime on the first day of training, outlook is already shifting to be more positive, with trainees talking about how they can change their own behavior.” One upper manager said, “The watch phrase is ‘come to class then see if you have any complaints,’” knowing that, once workers had completed the class, their concerns would be largely addressed.

Initially the facilitators were concerned that the new, more open-minded workers would emulate the well-respected old heads, and resistance would “spread like a fungus,” as one facilitator phrased it. However, a year later the facilitator recognized that “old heads are now retiring, which allows us to capture the new guys before they are influenced” by the old culture. Although there were incidents of anti-CAB rumors, along with name-calling and ridicule of samplers in the first year, steering committee members reported that, in general, samplers were allowed on trains to conduct third-party feedback sessions.

The most serious worker opposition to CAB occurred a few months after feedback sessions started, when one facilitator was threatened with official sanction by a few active members of his union for running CAB. In a race to educate these members on CAB before they could move the union to act, the facilitator persuaded them to attend training. When training ended, all had changed their positions to support CAB, except for one individual who left before lunch (“but even he quieted down,” according to the facilitator). Within a couple of months, the union trouble was “old history,” according to the facilitator.

Noting the significance of union support, the steering committee endeavored to bring all local union leaders together through CAB training. One local union chairman was highly impressed and, at the request of facilitators, wrote a testimonial letter that was published in the CAB newsletter (produced by the steering committee). After meeting with facilitators of other BST sites in spring 2006, one CAB facilitator concluded that overall, union support was better at SASU than at other sites using BST’s process.

6.1.3.2 CAB Scope Expanded Beyond CRZ and the San Antonio Terminals

The steering committee expanded CAB to outlying terminals. In January 2006, CAB–CRZ training began in Laredo, and in January 2007, classes began at Round Rock. Classes began at Del Rio in June 2007 after its terminal manager heard of the benefits of CAB and requested CAB classes. Managers at Taylor, Smithville, and New Braunfels, northeast of San Antonio, likewise asked for classes at their locations, although employee resistance at those locations proved protracted, which the steering committee attributed to a low number (but not rate) of incidents. “It’s hard if they’re not hungry for change,” observed one facilitator. To handle the expansion of territory, the facilitators recruited a sampler for each location who would provide expert advice on CAB and encourage feedback sessions. To encourage local ownership, Laredo was given more autonomy towards the end of the evaluation period by being able to allocate its part of the CAB manpower budget on its own.

Coinciding with the extension of CAB–CRZ to outlying locations was the expansion of CAB to cover switching operations. By July 2006, the steering committee developed the CAB–Switching feedback session checklist, and yard workers replaced some steering committee members who were cycling off. Management was pleased with the results. "They know the job better than anyone," said a local upper manager. A visiting regional manager said, "I think you've done a great job here," after receiving a thorough briefing on CAB–Switching.

Sampler training for CAB–Switching began in San Antonio in October 2006, but the first CAB–Switching class was almost a month earlier, at the small outlying terminal of Eagle Pass, south of San Antonio, on the Mexican border. According to one of the facilitators, the local manager, who was a personal friend, had "harassed me into coming down there to give a CAB class." Eagle Pass was described as a tight-knit community of about 20 workers and the entire workforce was trained in about two months (see Training, page 80). The facilitator viewed this location as a good test case for the impact of 100 percent training combined with a manager thoroughly committed to the process. A top San Antonio manager noted that the heavy switching conducted at Eagle Pass (despite its small size), and the frequent human factors derailments, provided an opportunity for CAB–Switching to show its potential.

In July 2007, CAB–Switching classes began at the Laredo yards.

6.1.3.3 Multiple Efforts Attempt to Raise Contact Rate

As expected, the dispersed work settings proved a challenge for conducting third-party feedback sessions at SASU. These sessions typically required a sampler to:

- Drive out and meet a train.
- Ride the train until a CRZ was traversed.
- Wait for an opportunity to provide feedback, such as when the train stopped at a siding.
- Arrange for transport off the train.

Only three or four feedback sessions per sampler per day could be conducted, given this process. If all CAB resources were dedicated to third-party sessions, allowing none for training or other necessary CAB activities, the contact rate would be between 33 and 58 percent. The Design Team chose a target contact rate of 100 percent, where the number of sessions per month equals the population of the workforce.

The stakeholders anticipated that cross-crewmember feedback sessions would supplement the third-party sessions to achieve the target contact rate. Train crews frequently have idle time in sidings. Stakeholders expected crews to use this downtime for reviewing practices in a previous CRZ and completing a CAB–CRZ checklist. However, the frequency of cross-crewmember sessions was disappointingly low, and the overall contact rate remained below the target throughout the evaluation period (see Contact Rate, page 82).

By spring 2006, recognizing that the overall contact rate was too low, the CAB trainers began to urge more cross-crewmember sessions. Meanwhile, the steering committee increased the proportion of third-party sessions. In summer 2006, the contact rate still remained "our big nemesis," according to one facilitator. Local and regional management also regarded the contact rate as "our number number-one area of concern" and met with the steering committee to discuss it. Even the evaluation team expressed concern to the facilitators that the current contact rate may not make the evaluation valid. Despite self-imposed and outside pressure to improve the contact rate, improvements remained elusive.

In winter 2006–2007, the BST consultant warned that the contact rate was falling below the tenth percentile of all sites implementing BST's process. Management continued to express concern, but also emphasized that the quality of the feedback sessions must not be sacrificed to meet the targeted numbers.

Resistance was a more significant factor for cross-crewmember sessions than for third-party sessions. Ironically, although workers were reluctant to conduct cross-crewmember sessions, they were generally enthusiastic about conducting third-party sessions. Samplers were "banging down the door" according to one facilitator, even though it could mean a slight decrease in pay compared to ordinary road work. Samplers who had conducted third-party sessions wanted to conduct them again at the earliest opportunity. According to the facilitators, the attractions of third-party sessions were that they were a welcome change from ordinary work and provided a means to directly contribute to safety.

Reasons for the aversion to cross-crewmember sessions were never clear. One stakeholder wondered if providing a Saturday off for CAB training was too strong a motivator and that workers were never particularly interested in CAB. However, this was inconsistent with the strong interest in third-party sessions. Trained workers remained highly supportive of CAB, yet did not conduct cross-crewmember sessions. When one manager boarded a train, "The first thing the crew said was 'CAB helped me, but I haven't done any observations.' They didn't say why."

Anticipated resistance from a crewmember may have discouraged cross-crewmember sessions. Consistent with this explanation was the fact that a class held at Taylor (a terminal known for its high resistance to CAB) was looking forward to cross-crewmember sessions but was not willing to observe anyone unfamiliar with CAB.

Although a third-party sampler can simply leave a train after encountering resistance, a cross-crewmember sampler must remain with the observed coworker on the train for many hours, possibly enduring an uncomfortable work environment. This would imply that the rate of cross-crewmember sessions would spontaneously accelerate once a substantial fraction of the workforce was trained on CAB, and that support for CAB would be the norm, but this was not observed during the evaluation period.

Another possible explanation is that the simultaneous initiation of CAB and CRZ procedures caused some unfortunate equating of one with the other. One facilitator reported that a crew told him that they “do it” (the new CRZ procedure), but don’t document it (complete a feedback session checklist). Workers may have felt they were doing their part for CAB simply by engaging in safe CRZ practices, but this does not explain the interest in third-party sessions.

Finally, cross-crewmember sessions may have represented excess work for the crews. One facilitator suggested that cross-crew sessions cut into a crew’s rest time when waiting on a siding. “I guess with all the work and paperwork for running a train, they just don’t feel like doing a sample,” suggested one steering committee member. In contrast, when acting as a third-party sampler, a worker’s task is limited to CAB observation and feedback, and they can rest otherwise. Furthermore, by conducting a day of third-party sessions, workers can count on having two subsequent nights at home according to the collective bargaining agreement, which makes third-party sessions an additional attraction.

In its efforts to increase cross-crewmember sampling, the steering committee set expectations more clearly during CAB training and used one-on-one coaching opportunities to encourage more feedback sessions. Training and coaching emphasized the importance of documenting changes in practices that workers claimed they were doing. A sampler on light-duty work was recruited to hand out checklists in a break room and urge outgoing crews to conduct cross-crewmember sampling. The steering committee scheduled a burst of third-party sessions for a month, anticipating that the experience would overcome concerns about resistance among the samplers and make them more willing to subsequently conduct cross-crewmember sessions. All these efforts had modest or little effect.

Ultimately, the process came to depend on third-party sessions more than was hoped. Following up on a suggestion from the BST consultant, the steering committee instituted a system to use the interest in third-party sessions to encourage cross-crewmember sessions. Samplers could enter their names in a rotating pool of samplers to conduct third-party sessions (not unlike a pool used for trip assignments). To remain in the pool, each sampler had to conduct cross-crewmember sessions. Limited resources prevented heavier use of third-party sessions in this form until spring 2007 (see CAB Budget, page 71). Although this system nearly doubled the overall contact rate without reducing observation or feedback quality (according to BST measures), it remained below the targeted levels for the duration of the evaluation (see Contact Rate, page 82).

6.1.3.4 CAB Budget Increase Gives Process Needed Boost

During the first year of CAB, local management fixed and administered resource allocation. In summer 2006, at the request of the facilitators, management provided the facilitators with a monthly budget of 100 worker-days to allocate (e.g., to use creative efforts to improve the contact rate). Management was pleased with the way in which the facilitators allocated the budget. “We’re moving forward,” said one top manager.

However, the size of the budget became increasingly limiting according to the steering committee, as CAB’s scope expanded and more locations were added. Facilitators recognized that refresher training for samplers could not occur for years under the current budget. Because of the budget limitations, the rollout of CAB–Switching was slower than desired. Said one facilitator, "We really have our hands tied in getting the yard going in San Antonio because of the budget. Only Eagle Pass is really going. I don't want to water down CAB–CRZ for the yard CAB–Switching effort." By spring 2007, CAB resources were stretched thin to include yard and road, and outlying locations. The budget remained constant although there was a larger steering committee (that included representatives from yards and outlying locations), increased sampler coaching, and additional third-party feedback sessions. Anticipating that it would soon be impossible to complete their planned activities under the current budget, some steering committee members considered drastically reducing CAB efforts, such as eliminating certain locations, abandoning the yard implementation, or reducing the steering committee size.

In winter 2007, the facilitators asked for a larger budget. However, at the same time, SASU suffered a slump in business and corporate UP pushed to cut costs under a productivity initiative called Project 75. Management furloughed workers and left open management positions unfilled. Local management maintained the CAB budget from the year before, but could not increase it. By spring, the facilitators raised their requests to the regional level. The regional vice president made an offer: for every worker-day reduced in CAB administration cost, the budget would be increased one-and-a-half days. "It was quite generous," noted one facilitator. "By having every CAB class be taught by no more than one steering committee member [the other instructor being a facilitator], we've achieved significant savings in our budget." In total, the steering committee cut 20 worker-days of administrative time per month, earning a budget increase to 130 worker-days. Available non-administrative worker-days nearly doubled, from 60 days to 110.

The new budget substantially improved the ability to expand CAB to outlying regions, turning CAB from a local San Antonio initiative into a service-unit-wide process. The additional workload imposed on facilitators (to conduct more training) proved manageable. Local management was very pleased with the increased budget and the performance of the steering committee. One top manager said, "The increase in funding is a big plus. I couldn't be happier with the number of samplers being trained. The guys have taken them up on that offer—reduce their costs somewhat and be able to do more training."

6.1.3.5 Steering Committee Selective in Identifying Barriers for Formal Removal

A few months after the initiation of CAB–CRZ feedback sessions, the steering committee soon recognized that unsafe CRZ practices were less likely to be associated with barriers that require capital expenditure to remove than with confusion about within-cab communication practices. A facilitator reported that the most common explanation for a safe practice that workers rarely performed was, "I didn't know I was supposed to do it this way."

In many BST implementations, a problem such as this would be raised to the Barrier Removal Team, which would likely increase and alter training to solve the problem. However, the steering committee was reluctant to release the associated feedback session results to managers, including those on the Barrier Removal Team, because of its concern that management would be compelled to react in the traditional, punitive manner (e.g., by intensifying FTX on the associated rule).[8] Said a facilitator in December 2006, “We probably can’t do that [share data with management] now.” A steering committee member agreed, “We need more trust first.”

Instead, the steering committee chose to handle such barriers internally, using their own resources to quietly educate the workers through one-on-one contacts and some employee-only printed materials. Over the course of the evaluation period, four items from the CAB–CRZ checklist were targeted for this treatment. The steering committee regarded such efforts to be generally successful. In fact, a facilitator reported that a manager unknowingly commented that he noticed more workers performing a particular practice more safely.

The managers and the BST consultant anticipated more work for managers on barrier removal with CAB–Switching, and in practice, they resolved problems with tools, lighting, and tripping hazards for the San Antonio yards. “We were able to use the data to make a case to [get the manpower] to get those guys out there cleaning up a yard,” said a facilitator.

Formal barrier removal nonetheless was uncommon compared with other BST implementations. One reason was the presence of alternative processes for addressing safety concerns at SASU. Small physical problems, such as a single switch needing repair, were handled informally by sending a note to the manager in charge of the area of concern. Larger problems at the San Antonio yards were handled by the Kirby, East Yard, and SOSAN (KEYS) safety committee. Formed about a year after CAB started, KEYS was a joint labor-management, cross-trade committee that met monthly to review problems and develop solutions. The CAB steering committee and Barrier Removal Team members were heavily involved with KEYS (see CAB Makes Horizontal Connections, page 77), attending meetings and alerting KEYS to barriers identified in feedback sessions (e.g., a missing light in a yard). This meant problems identified by the steering committee were naturally resolved through KEYS, making it a surrogate Barrier Removal Team in some cases.

With informal communications handling small problems and KEYS handling large problems, this left only exceptionally large problems to be handled by the Barrier Removal Team, according to one steering committee member. Two problems resolved during the evaluation period fell into this category. The first problem was re-ballasting the Eagle Pass yards to remove tripping hazards. Initial estimates by management were $15,000, but much more work was necessary, and the final cost totaled $65,000. A facilitator reported that conditions afterward were “a whole lot better than when we started. It counts as one barrier removed, but it was a big barrier.”

Another exceptionally large barrier removed through the CAB process was the use of lead locomotives without air conditioning. Under a hot Texas sun, the interior of

---

[8] This concern was so great that the steering committee would not release this information to even the evaluation team.

locomotives reportedly could reach 120 degrees. The availability of air conditioning in locomotive cabs was documented during feedback sessions, and the steering committee identified an association between at-risk behavior and the lack of air conditioning along the hottest locations in the service unit.

Upper-level local managers were impressed with the data presented by the Steering committee in January 2007 and agreed that trains assembled for the hottest locations would have an air-conditioned lead locomotive whenever possible. Crews would have the authorization to switch a locomotive with air conditioning to the lead when feasible. This was a local remedial solution to the problem, as trains may be made up in distant locations in UP's network and would not be under SASU's control. The steering committee took the issue to the corporate level and, in summer 2008, announced that corporate policy had been changed so that all trains in summer would have an air-conditioned locomotive in the lead.

6.1.3.6 Managers' Role Slowly Becomes More Active

While the steering committee was working through worker resistance, scope expansion, contact rate, budgeting, and barrier handling, the managers were developing their own role in CAB. Initially, their role was primarily not to interfere with CAB, to give employees ownership of the process, and to avoid any punitive associations. Top management instructed front-line managers to stay away from feedback sessions while the sessions were under way. This deliberate distancing of management from CAB gave CAB credibility and support among workers.

There were cases in which a manager arrived to conduct an FTX but left immediately once he realized CAB samplers were present. The workers "get a big kick out of that," said a sampler. Consistent with BST's advice, the steering committee and management agreed that management support should eventually be more active but that this had to be adjusted carefully. One steering committee member recommended that, "Management should enter slowly, as a natural progression. It shouldn't look like UP is taking over." A regional manager observed, "With an us-versus-them mentality, sometimes it's better not to say even positive things about CAB."

By summer 2006, front-line managers were more actively promoting CAB, encouraging training and feedback sessions. "It has to be the right manager," said one facilitator, "so it doesn't come across as 'you should've known better,'" which could be implied if CAB were suggested in the context of a failed FTX test. The facilitators worked with top local management to ease the front-line managers into this role. "We need to tell the superintendent not to push too much. One manager wanted to look at a checklist form. That was a fire we had to put out," said one facilitator. To familiarize the managers with the CAB process, managers attended one-day CAB training sessions, beginning in fall 2006.

SLD training was conducted for managers through summer 2007, the second year of the evaluation period. "To a person, all the reviews from other managers I've had say that these were useful modules," reported one top manager. The most immediate application of the training was on FTX testing procedures, which local management was currently reworking. The quality of FTX feedback was improving, which top SASU managers attributed to the SLD training. One manager reported that an FRA inspector described the FTX interactions as "the best we've seen at any location." A facilitator acknowledged that worker relations with managers were generally improving.

Consistent with BST recommendations, efforts were made for managers to more directly participate in CAB, but by the end of the evaluation period these efforts had exceeded the tolerance of the prevailing culture. As early as spring 2006, top managers suggested that managers could conduct their observations without any paperwork, and local union leaders endorsed the idea, however, higher-level union leaders quashed it.

In fall 2007, a middle manager asked if he could accompany samplers to feedback sessions in a San Antonio yard. Not only did samplers agree, but there was "No blowback about confidentiality from workers whatsoever," according to a top manager. Despite these positive events, when a manager conducted some experimental feedback sessions two months later, the steering committee concluded that it was still too early in the culture's evolution for this degree of management participation.

6.1.4 CAB Institutionalization in SASU and Beyond

Through the latter half of the evaluation period, the CAB process in particular and a cooperative approach to safety in general, showed signs of institutionalization in the culture, not only at SASU but also beyond.

6.1.4.1 CAB Withstands External Pressures

The CAB process continued to grow in the face of challenges and changes to the surrounding organizational environment. Rotation of managers to other locations is common in the railroad industry. Over the course of the evaluation period, the superintendent and two top managers at SASU plus the regional vice president, all strong proponents of CAB, moved on to other jobs in UP. New managers met with the facilitators, attended CAB training, and supported the process. Following BST's sustainability recommendations, several workers rotated off the steering committee during the evaluation period and were replaced; this was regarded as a positive event, bringing in an influx of fresh ideas and contacts.

CAB also withstood outside pressures. In November 2006, the discipline policy was modified at the corporate level, substantially increasing the severity for punishments for infractions of certain critical rules. Across the UP network, unions threatened to withdraw from joint labor-management safety programs in retaliation. This included the mechanical department in SASU, which was practicing the TSC process that originated at North Platte.[9] The facilitators of CAB, however, contacted regional union leadership and successfully argued against withdrawing from CAB. Key to their argument was that the Barrier Removal process provided leverage with which to improve physical work conditions.

---

[9] The mechanical union would eventually make good on its threat, withdrawing from TSC in retaliation for what appeared to be a misunderstanding by SASU management.

6.1.4.2 CAB Principles Internalized

At an individual level, there were reports of the workers internalizing responsibility for safety, and that this increased worker empowerment. CAB trainers reported this occurred as early as during the first hours of training. It often began with an exercise in which workers depict examples of themselves engaging in safe and at-risk behaviors, both at work and at home. As one trainer reported, in studying their own at-risk behaviors, workers recognized that little had prevented them from choosing a safer course of action.

The presentation of videos of their coworkers engaging in at-risk practices also had an impact.[10] “They laugh at the videos, but we say, ‘You do that every day,’” said one trainer. Another trainer observed, “Workers start recognizing they actually have a choice with much of their behavior, such as whether to walk around train ends with 20 feet of clearance or not. This sense of control is a major change that’s pretty tremendous.” The internal changes were reflected in practices in the field. A steering committee member, who worked at two terminals, one with and one without CAB observation feedback sessions, saw the difference in practices between the locations. One worker observed that others were “doing things differently; they’re changes for the better…. People like to do their jobs safely.” Once some do it, others follow. “It’s becoming more and more of a routine,” stated a steering committee member.

“CAB class is mostly about seeing safety value not learning rules,” said a trainer. With this awareness came a change in priorities among workers, who placed safety ahead of even protecting their jobs. Top managers reported that incident investigations were going better than before the start of CAB, with more people taking responsibility. In spring 2006, an engineer was decertified. He had not attended CAB training but had noticed that those conductors who had been trained engaged in more professional work practices. He told investigators that perhaps he could have avoided the decertification if he had attended CAB class. “This sort of comment would have been unimaginable one to one-and-a-half years ago,” said a top manager.

In summer 2006, a train collision resulted in serious injuries. The engineer, who had been through CAB training, took full responsibility for the accident (he had been using a cell phone). A top manager reported, “Not once did I hear anything other than the crew being at fault… not the signals… nothing about what the company could’ve done.”

Safety committee meetings, such as KEYS, acquired a different character. “Workers are no longer saying ‘it’s the company’s fault.’ Now it’s ‘we’ll talk to other crewmembers.’ They’re talking about training,” said one top manager. Formerly, workers would bring “a laundry list of complaints” for management to address. “Now they work with management more as a team, often arriving at the meetings with solutions of their own.” Another top manager reported, “There’s been a move from safety meetings with employees with a list of things for the company to fix to ‘how are *we* [emphasis in original tone of voice] going to fix these things.’ There’s a little bit of ownership.” “Even

---

[10] These videos were staged under controlled conditions by steering committee members, following a script they had prepared.

in the locker room," another manager added, "there isn't the 'we should have this or that' anymore." When SASU was having a problem with power switches, one union conducted its own training to reduce derailments, a reaction that one manager attributed to the CAB approach "rubbing off" on other workers. Some workers tried to help others to be safer, but if that did not work they brought the problem to the attention of a manager. Workers, especially new hires "are telling you more about other things" holding other employees accountable, rather than covering it up, said a top manager. "This trainer who trains all the time says to me, 'This guy—I've told him five times. He's not getting it.'"

6.1.4.3 CAB Makes Horizontal Connections

CAB and proactive safety were making inroads at SASU beyond CRZ and switching. At the request of SASU management, in December 2005 and 2006 CAB facilitators participated in a program to educate the public on the dangers of railroad crossings. Management also asked steering committee members to attend KEYS meetings, not just to identify safety barriers needing removal (as discussed on page 72) but also to provide an avenue for solutions. In one meeting, a manager suggested using the CAB newsletter to educate workers on a track condition as part of an intervention. When KEYS developed a new mentoring program, a steering committee member used CAB audiovisual equipment to create a video to promote it.

"There is a flow of information between safety committee meetings and CAB classes," noted a facilitator. A worker in the mechanical department read in the CAB newsletter that many locomotives do not have radio handsets on the conductor's side of the locomotive cab, making it difficult for the conductor to handle radio-communications and thus allow the engineer to focus on train control when under CRZ conditions. To address this issue the worker began carrying a few extra radio handsets into the field. The worker was the facilitator-equivalent for TSC in the mechanical department and also a participant in KEYS, where he met the CAB facilitators face-to-face. All agreed to attend each other's classes. According to both a top manager and a facilitator, the safety committees, CAB, and general training have become "all one big interface" for handling safety problems.

As CAB worked directly with other safety groups and management, union officials became increasingly isolated. Since their inception the function of unions has been to bargain with management on behalf of workers interests, including safety. The unions now found themselves in an unfamiliar culture. The tactics they had used to deter management, such as withdrawing from safety programs, were being undermined by the cooperation fostered by CAB and other cooperative safety efforts. Workers experienced the value of CAB-like processes and were liable to support those processes, irrespective of the pronouncements of the union officials. In the SASU mechanical department, some workers reportedly continued to practice TSC in defiance of the union's decision to pull out of it.

Some local union leaders were able to resolve the resultant conflicting roles, but others were challenged by it. One CAB steering committee member resigned, reportedly because it interfered with his ability to confront management as a union leader. "He painted himself into a corner," said another steering committee member. This resigning member had promised things to the rank-and-file, intending to use the threat of withdrawing from CAB to obtain them, but as a steering committee member he could not promote withdrawal from the process he was charged to run. Another union leader who was a steering committee member likewise struggled with other committee members about sharing feedback session information with management. A witnessing steering committee member said the union leader was concerned about the power this could give management over workers. The union leader also resigned from the steering committee following a serious challenge to his union position.

6.1.4.4 Regional and National Level Embraces CAB

Beyond SASU, regional and national leadership came to embrace CAB and nonpunitive proactive safety in general, and began to engineer its spread to transportation departments at other service units.

In fall 2005, regional union representatives visited SASU, and facilitators briefed them on CAB. One representative was sufficiently impressed to give a supportive talk at a subsequent union division meeting, which ultimately led the union to officially support CAB rather than remain neutral (like the other union).

By August 2008, union support for CAB spread to the national level. Signaling their positive regard for CAB, UTU and BLET national leaders, along with corporate UP management, accepted an FRA Distinguished Public Service Award for cooperating on the implementation of CAB.

Railroad management likewise embraced CAB and CAB-like processes. In March 2006, a senior manager for UP revealed to labor and local managers at Livonia Service Unit his ambition to change the company, from a reactive command-and-control safety approach towards a proactive systems orientation. He gave this cultural change a name: Transformation. Said one regional manager in summer 2006, "I'm pro-[PPF], since over 30 years I've seen what doesn't work. If you terrorize the worker to comply, that goes away when the manager is gone. We need to build safety culture as a base." The plan, he further elaborated, was to have CAB-like processes at all service units in the company. In testimony before the U.S. House of Representatives' Committee on Transportation and Infrastructure on October 25, 2007, a top corporate manager made public the plans for corporation-wide CSA as part of a suite of proactive nonpunitive safety initiatives. "The barriers to this program are sometimes difficult to overcome—for both the employee and the manager. However, we are committed to overcoming these barriers through education and example," he testified.

The expansion began at the start of 2006, in the Southern Region. The Livonia Service Unit (LVSU) was selected for a second FRA-sponsored CSA site with BST support. They started observation feedback sessions in fall 2006. In summer 2006, service units in the Western Region expressed interest in CSA, according to the BST consultant. Simultaneously, corporate leadership began working on the management of Houston Service Unit (HOSU) in the Southern Region to prepare it for CSA. "HOSU lacks the atmosphere of change, but it's starting now," reported a regional manager. By the end of 2006, HOSU was receiving FRA-sponsored SLD through BST and had started a home-grown PPF process for its yards.

According to FRA, early signs of success for CAB encouraged corporate leadership to select the Fort Worth Service Unit (FWSU) transportation department to receive a CSA-like implementation under UP's own TSC process, which UP originally developed for mechanical departments. Corporate management planned for TSC to be eventually implemented for the transportation departments of all service units.

CAB facilitators participated in the expansion of CSA to other service units. As early as December 2005, they made a presentation on CAB to HOSU. In spring 2007, one facilitator returned to exchange ideas on HOSU's and SASU's respective processes. CAB facilitators were concerned that HOSU's homegrown process might lack key features necessary for worker acceptance, which would reflect badly on other CSA processes such as CAB. LVSU managers joined the CAB Oversight Committee telcons in April 2006, and then were joined by LVSU CSA facilitators when they were selected.

A delegation from LVSU visited SASU to see CAB firsthand in fall 2006. A regional manager reported that some at FWSU used SASU as an example of how crewmembers should communicate in the cab. The CAB facilitators traveled to FWSU to give the superintendent an overview of CAB, raising his interest in CSA. The facilitators also attended TSC training, prepared by UP, to familiarize themselves. The CAB steering committee viewed the expansion of TSC into the Southern Region's transportation department with some apprehension. If workers were to see TSC replacing CAB, even only in name, it could be viewed as another example of management failing to follow through with a safety program. Working with local and regional managers, it was agreed that CAB would remain CAB in both name and process, but resources for CAB would be provided under the TSC program.

The nonpunitive proactive safety approach that characterizes CAB began making its way throughout much of the UP corporate hierarchy. One manager, strongly involved in CAB at SASU, until his promotion to the regional level, wrote a positive article about CAB for the UP intranet, where it could be read by all workers and managers.

The superintendent of SASU during the implementation of CAB served on a corporate team that modified FTX so that workers could earn points on their Employee Development Review scores for positive FTX results. "Some workers are now asking to be tested to earn points," said a union official at SASU. In 2008, UP relaxed the increase in punishments it had put in place in November 2006 for certain rules violations. This partial reversal of policy may have been abetted by union leaders pointing out that such harsh discipline was inconsistent with CAB-like approaches to safety. High-level managers who promoted CAB at SASU had since moved up the management ladder, where they lobbied to make face-to-face verbal coaching a required aspect of FTX testing, according to a regional manager.

6.1.5 Summary

Tragic events, leadership changes, and shifting demographics at SASU made the site more conducive to alternative cooperative approaches to safety, such as CSA, than some other sites. Nonetheless, traditional aspects of the railroad culture required CSA to be adapted. Stakeholders selected CRZ for the initial focus for CAB, planned to minimize worker resistance, and adjusted training to address cost concerns related to railroad collective bargaining agreements on work hours.

Worker training on conducting feedback sessions proved to be a major educational tool for overcoming resistance. Lowered resistance allowed CSA to expand into outlying locations, as well as into switching operations. The expansion ultimately required cuts in CAB administrative overhead and an increased budget for it to remain viable. Ongoing challenges included achieving a high rate of feedback sessions on the road and gradually increasing involvement of management.

CAB demonstrated an ability to withstand external pressures and ultimately became more ingrained into the organization, with influence extending beyond its own activities and SASU. This culminated in a company-wide endorsement of CSA and its principles of proactive, nondisciplinary safety.

6.2 CAB Process Metrics

6.2.1 Training

Figure 14 graphs the cumulative percent of workers trained between July 2005 and October 2007.

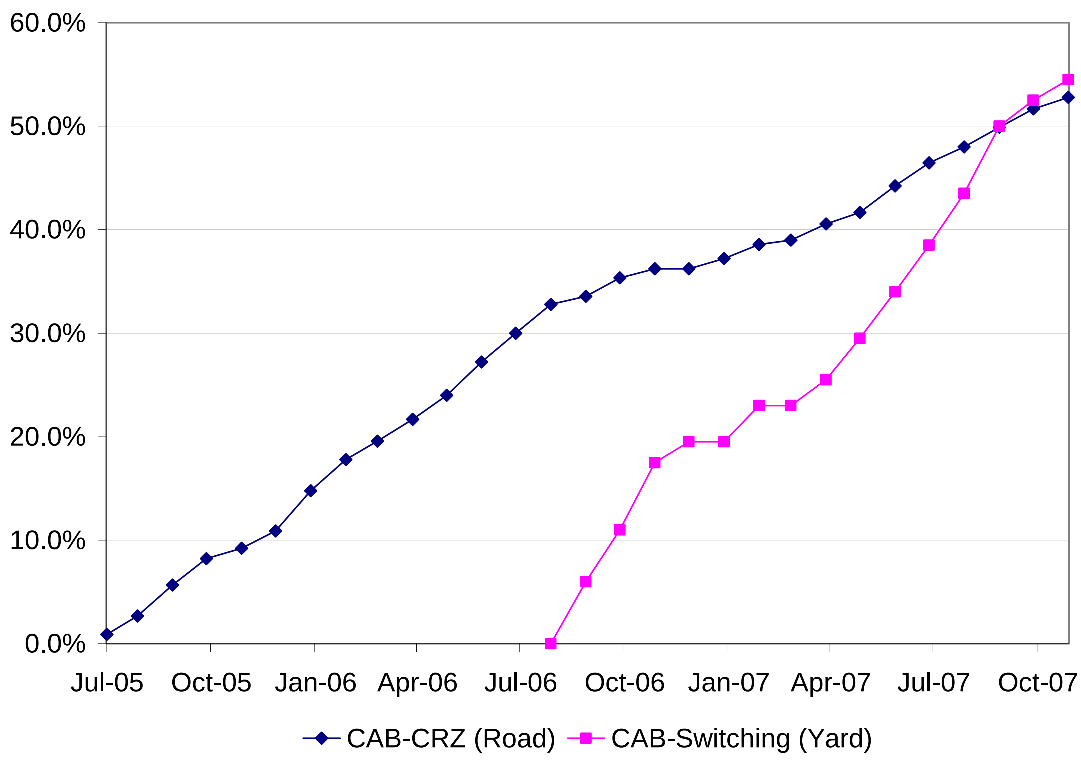


Figure 14. Cumulative Percent of Workers Trained for CAB–CRZ (Out of 900 Road Workers) and CAB–Switching (Out of 200 Yard Workers)

Both CAB–CRZ and CAB–Switching trained over 50 percent of road and yard workers by the end of the evaluation period. The sharper climb for CAB–Switching training relative to CAB–CRZ training was caused by the larger number of road workers at SASU than yard workers. In raw numbers, 475 road workers and 109 yard workers were trained by November 2007. This exceeded initial estimates, based on two classes of eight trainees per month. Relying only on volunteer trainees, all classes were always fully enrolled. The frequency of classes was limited by the amount of manpower that each location could spare for training. During months when train traffic was low, such as January and February, additional classes were held.

The rate for CAB–CRZ slowed with the introduction of CAB–Switching training in summer 2006, when CAB resources were divided between the two. The acceleration of both rates, in late spring 2007, corresponded to the increase in budget and reduction in administrative costs for CAB (see CAB Budget Increase, page 71).

Figure 15 illustrates the cumulative percent of workers trained within CAB–Switching between August 2006 and November 2007 for Eagle Pass and for all other yards (the latter corresponds to the San Antonio Complex alone until August 2007).

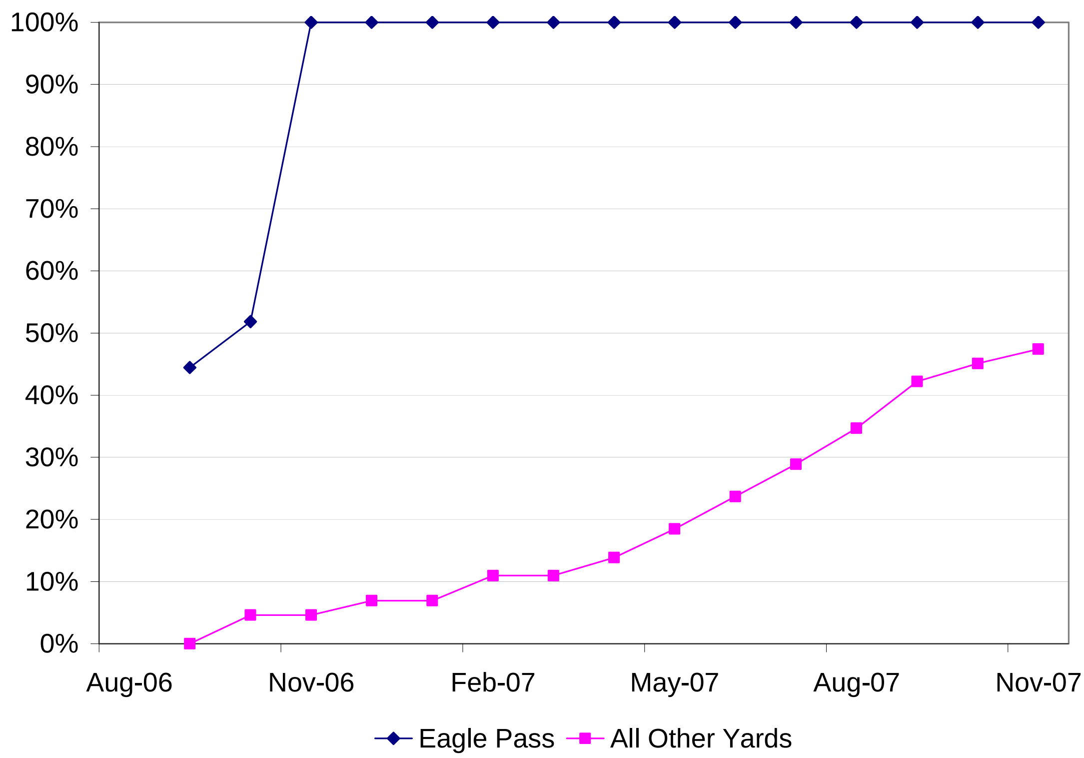


Figure 15. Cumulative Percent of Workers Trained for CAB–Switching at Eagle Pass (Out of 27 Workers) and All Other Yards (Out of 173 Yard Workers)

Eagle Pass was subject to a burst of activity with four classes in almost 2 months with all 27 workers completing CAB–Switching training. The training rate at other locations (predominantly at the San Antonio Complex), which had a larger base of workers, proceeded slowly at first when scarce training resources were split between CAB–Switching and CAB–CRZ. Training accelerated in spring 2007, when CAB reduced administrative costs and received an increased budget. By the end of the evaluation period, 82 yard workers outside Eagle Pass were trained.

6.2.2 Contact Rate

As mentioned earlier (see Contact Rate, page 41), contact rate is the number of feedback sessions completed in a month expressed as a percentage of the targeted workforce, namely 900 road workers for CAB–CRZ and 200 yard workers for CAB–Switching. Figure 16 illustrates the contact rate for CAB–CRZ and CAB–Switching over time between July 2005 and October 2007.

The contact rate of CAB–Switching was usually higher than the CAB–CRZ rate. This is partially due to the smaller number of yard workers at SASU. In raw numbers, there were more feedback sessions every month for CAB–CRZ than for CAB–Switching during the evaluation period. Nonetheless, additional time and resources were necessary for each CAB–CRZ third-party feedback session compared to a CAB–Switching third-party session. The reasons were: 1) the more complicated logistics necessary for the dispersed work setting associated with road work, and 2) the focus on CRZ, which required samplers to wait for a CRZ condition before a session could start. This meant the contact rate for CAB–CRZ was substantially below CAB–Switching.

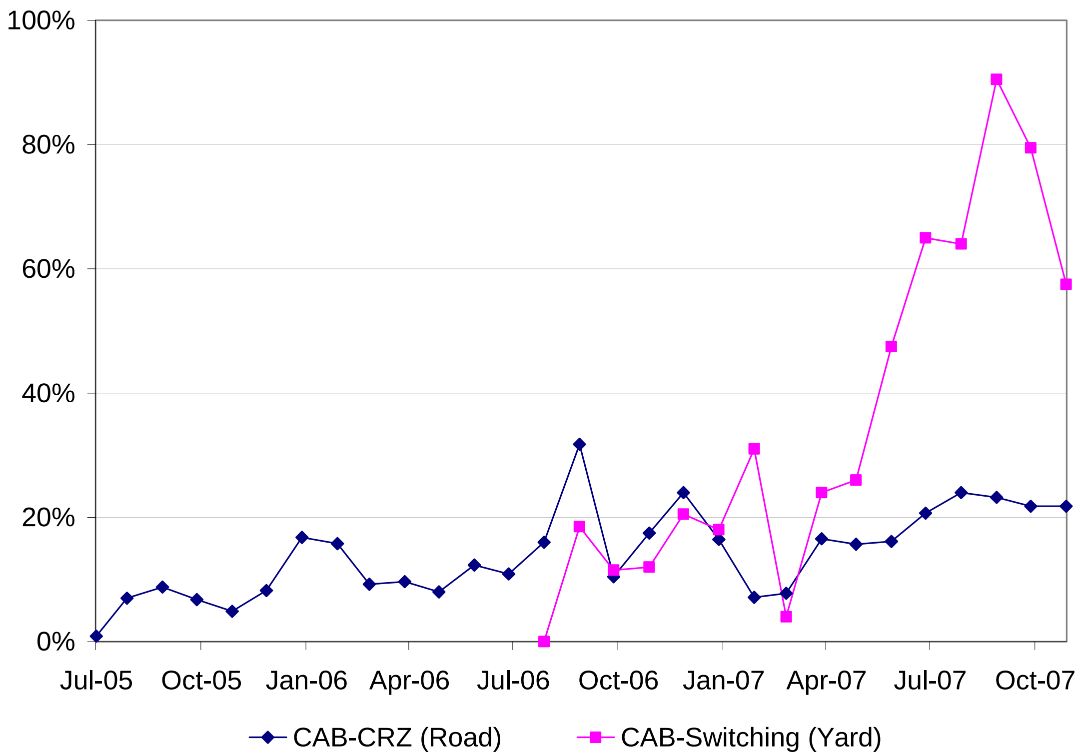


Figure 16. Contact Rate for CAB–CRZ (Out of 900 Road Workers) and CAB–Switching (Out of 200 Yard Workers)

Working with the BST consultant, the CAB Design Team chose a target contact rate of 100 percent. CAB–Switching approached this rate from June 2007 to the end of the evaluation period, after CAB had increased resources for third-party feedback sessions from reducing its administrative costs and receiving an increased budget. In general, CAB–CRZ had a low contact rate throughout the implementation. Occasional spikes, such as during September 2006, corresponded to bursts in third-party sessions aimed at encouraging cross-crewmember sessions (see Multiple Efforts Attempt to Raise Contact Rate, page 69).

Only in late spring 2007 was the contact rate consistently above 20 percent. This corresponded to both an increase in resources for third-party sessions and the introduction of the sampler pool system, where opportunities for third-party sessions were used as an incentive to conduct cross-crewmember sessions.

Although the contact rate was lower than the Design Team target, especially for CAB–CRZ, it was sufficient since most employees were likely to have completed repeated feedback sessions. By November 2007, 3,660 CAB–CRZ feedback sessions had been completed, corresponding to more than four times the number of road workers. For CAB–Switching, 1,139 sessions had been completed, corresponding to almost six times the number of yard workers.

6.3 Feedback Session Data

6.3.1 Consistency

Table 10 provides the results of the consistency analysis of the samplers' ratings of videos used in training and coaching, indicating the psychometric suitability of these ratings as an outcome measure.

Table 10. Assessment of At-Risk Feedback Session Data for Consistency Based on Ratings of Training/Coaching Videos

Measurement	Calculation	Results
Reliability	Percent of agreement between a sampler and the standard observation for each CRZ event.	Average percent agreement = 83.06%, significantly greater than 80.0% ($t(107) = 4.56$, $p < 0.0001$).
Bias	Difference in percent at-risk scores between samplers and the standard observation for each CRZ event.	Average difference = –9.94%, significantly different than 0.00% ($t(107) = -7.61$, $p < 0.0001$).
Time Drift	Changes in bias over time as indicated by the correlation of bias with the date the video is rated.	$r = 0.0312$, $n = 108$, $p = 0.749$, no significant time bias.
Retest Drift	Changes in bias in comparing ratings at the end of training with ratings during coaching.	Difference in average bias = –7.50%, significantly different than 0.00% ($t(106) = -2.05$, $p < 0.0427$).

Complete ratings were collected for 108 samplers at the end of training and for 13 samplers during subsequent coaching. Reliability was significantly higher than 80 percent. This is above the minimal standard for interjudge reliability that is often used in evaluation research (Rossi et al., 1999). Samplers showed significant negative bias, meaning they tended to rate the CRZ events in the videos as more at-risk than the standard ratings set by the steering committee. This implies that the scores for actual feedback sessions may be less at-risk than were reported.

Time drift was not significant, indicating that samplers trained at the beginning of the evaluation period rated the same sessions as equally at-risk as the samplers trained at the end of the evaluation period. However, samplers rated the videos as more at-risk during coaching than during training, suggesting that experienced samplers may have acquired a negative bias. The proportion of experienced samplers increased throughout the evaluation period, because samplers trained in the beginning stayed to conduct feedback sessions at the end. A greater proportion of experienced samplers implied more negative bias on average. This suggests the possibility that the true average at-risk scores may be lower than reported for later periods in the evaluation and that the improvement in practices is greater than indicated.

Overall, the feedback session ratings appeared to be adequate or perhaps even conservative for measuring positive changes in practices.

6.3.2 CAB–CRZ

Figure 17 illustrates average at-risk scores for CAB–CRZ feedback sessions between August 2005 and December 2007.

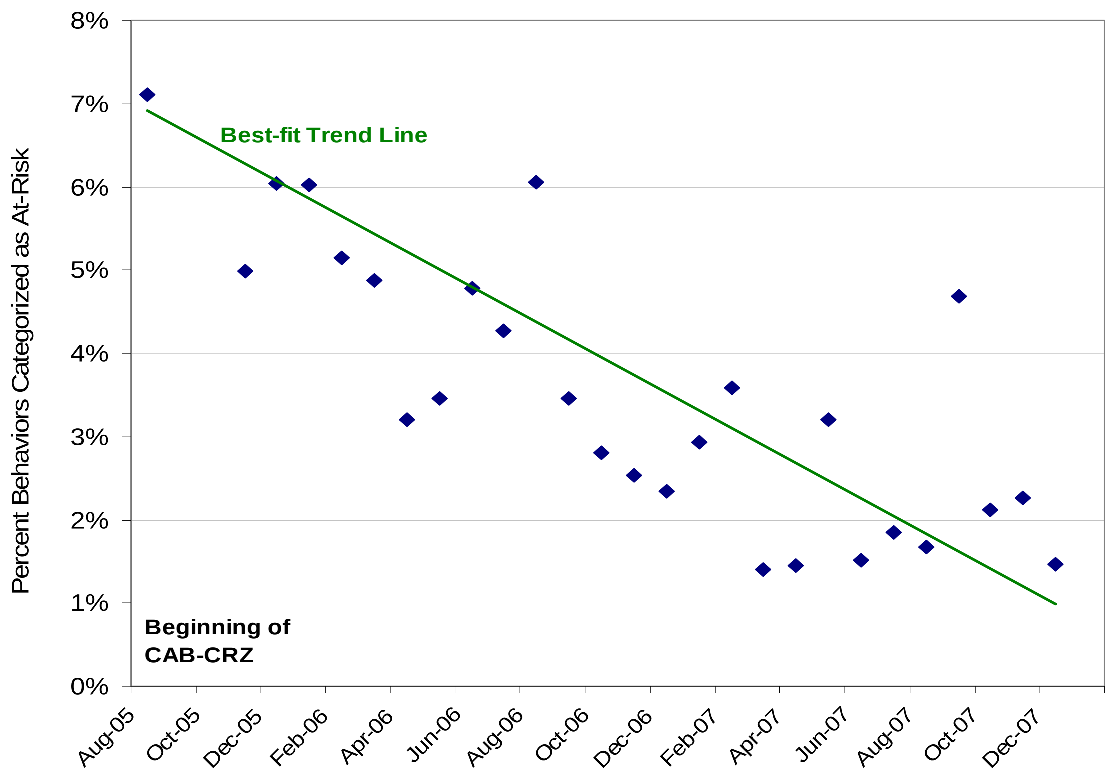


Figure 17. Monthly Average At-Risk Scores for CAB–CRZ Feedback Sessions

Diamond shapes represent the average percent at-risk for each month, whereas the straight line represents a least-squares best-fit straight line to the monthly averages.

Throughout the evaluation period, observations by CAB samplers show decreasing tendency for at-risk CRZ practices ($r = –0.797$, $n = 29$, $p < 0.0001$). On the basis of a linear regression of these data, the percentage of at-risk practices decreased from 7.14 percent in August 2005 to 1.05 percent in December 2007. In other words, the rate of at-risk practices at the end of the evaluation period was less than one-sixth the rate at the beginning.

Improvements were also seen for most individual items of the CAB–CRZ observation feedback checklist. Twenty of 33 list items significantly negatively correlated with time, and none significantly positively correlated with time.

6.3.3 CAB–Switching

Figure 18 illustrates average at-risk scores for CAB–Switching feedback sessions between August 2005 and December 2007. The diamond shapes represent the average percent at-risk for each month, whereas the straight line represents a least-squares best-fit straight line to the monthly averages. Figure 18 uses the same *x* and *y* scales as used in Figure 17 to facilitate comparisons between the two. As a result, the data start midway through the *x*-axis at the beginning of the Second Phase.

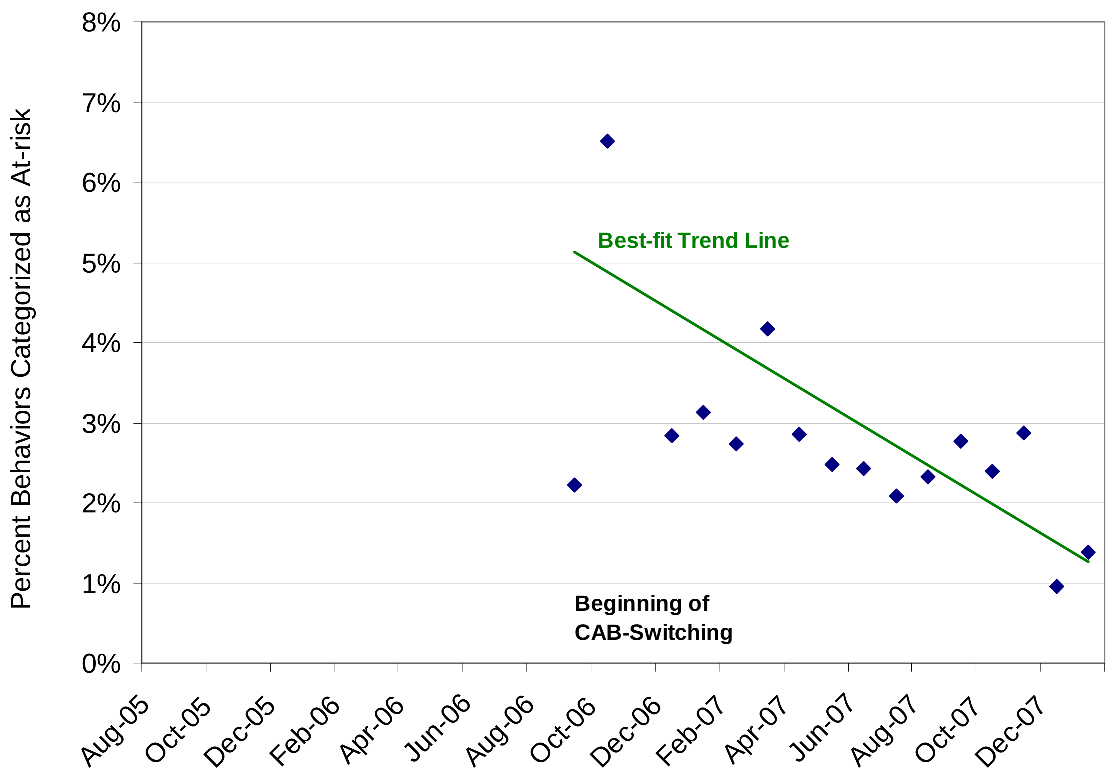


Figure 18. Monthly Average At-Risk Scores for CAB–Switching Feedback Sessions

Following the trend for CAB–CRZ scores, observations by CAB samplers show a decreasing tendency for at-risk switching practices from the beginning of CAB–Switching ($r = –0.569$, $n = 17$, $p = 0.0157$). On the basis of a linear regression of these data, the percentage of at-risk practices decreased from 5.13 percent in August 2005 to 1.05 percent in January 2008. In other words, the rate of at-risk practices at the end of the evaluation period was approximately one-fifth the rate at the start of CAB–Switching.

6.4 Corporate Safety Data

6.4.1 FTX

The CAB process was expected to decrease FTX failures relative to comparison groups, whether the comparison was with other service units or among categories of rules tested within SASU. Correlations of failure rates of FTX among the original service units of the Southern Region were low and inconsistent, averaging 0.039. This suggests that FTX failures are generally independent across service units and not subject to general region- or corporate-wide effects. As a result, service units were not used for comparison. In contrast, within each service unit, there were relatively consistent positive correlations in failure rates among the categories of rules (CRZ, Switching, and Other). Correlations averaged 0.1730, indicating a weak tendency for failure rates to rise and fall together, consistent with each category responding to general factors within the service unit. This suggests that the best comparison for CRZ Rules and Switching Rules failures within SASU would be with the failures for Other Rules.

Figure 19 illustrates the mean percent failures for FTX at each phase of the evaluation and each category of Rules Tested in the FTX tests for San Antonio.

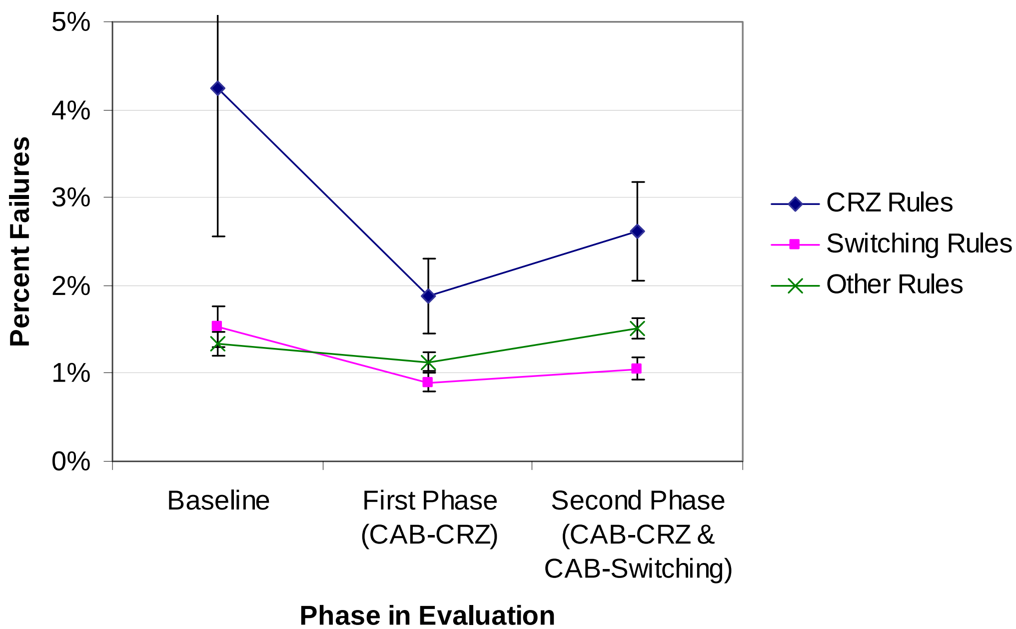


Figure 19. Mean Percent FTX Failures across Implementation Phases at SASU (Error Bars Are 95 Percent Confidence Intervals)

Failures decreased for all categories of rules from Baseline to the First Phase, corresponding to the initiation of CAB–CRZ at SASU. CRZ Rules, which were most closely associated with CAB–CRZ, show a reduction in failures that is greater than Other Rules from Baseline to the First Phase ($t(386) = 2.421$, $p = 0.0160$). Surprisingly, Switching Rules associated not with CAB–CRZ but with CAB–Switching (which did not start until the Second Phase) also experienced a reduction in failures greater than Other Rules ($t(386) = 2.697$, $p = 0.0073$). The change in failures for CRZ Rules was not quite significantly different from the change in failures for Switching Rules from Baseline to the First Phase ($t(386) = 1.930$, $p = 0.0543$).

Failures increased for all categories of rules between the First Phase and the Second Phase. However, the increase for the Switching Rules was significantly less than for Other Rules ($t(358) = 2.013$, $p = 0.0449$). The change for CRZ Rules between the First and Second Phases was not significantly different than the change for CRZ Rules in Switching Rules ($t(358) = -1.549$, $p = 0.1223$) or Other Rules ($t(358) = -0.938$, $p = 0.3490$).

In summary, both CRZ Rules and Switching Rules improved more than the comparison Other Rules between the Baseline Phase and the First Phase. Switching Rules worsened less than Other Rules between the First Phase and the Second Phase. The change for CRZ Rules over this time was uncertain.

There were no significant linear changes in failures for any category of rules during the First Phase (r's range from 0.1303 to 0.2044, lowest $p = 0.0505$), whereas all categories of rules showed significant linear increases in failures during the Second Phase (r's range from 0.2255 to 0.3111).

6.4.2 Injuries

Analyses were performed on all injuries because it proved to be infeasible to separate yard-related injuries from other transportation injuries. Figure 20 shows the normalized cumulative incidence plot of injury data for the SASU versus the other original service units of the Southern Region combined.

All service units exhibit a gradually decreasing rate of injuries from Baseline through the Second Phase, as indicated by the curved slopes in the normalized cumulative incidence plot. The rate of reduction for SASU in the sample of injuries appears close to the rest of the Southern region. Furthermore, any divergence of SASU from the rest appears to precede the start of CAB-Switching at SASU.

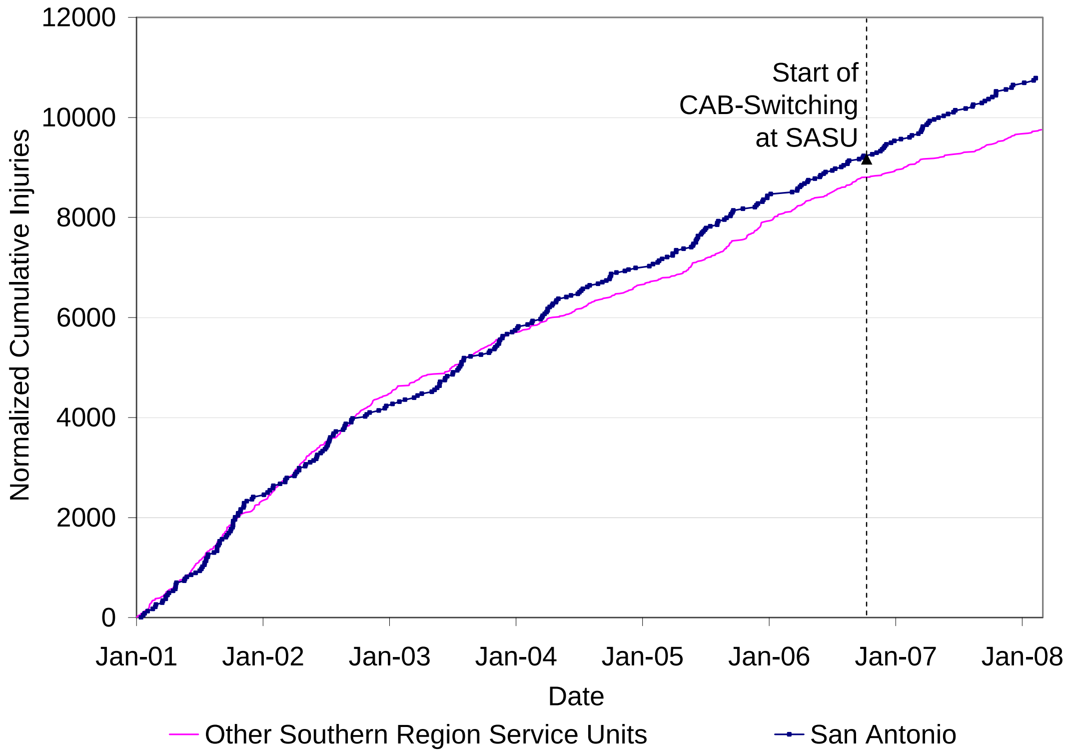


Figure 20. Normalized Cumulative Incidence Plot in the Original Southern Region

Autocorrelations of the log-transformed gaps between injuries were low and on average not significantly different from zero (range –0.009 to 0.178, average $r = 0.087$, $t(7) = 0.578$, $p = 0.581$), which is consistent with the gap times being statistically independent. Weibull plots were generally straight lines, consistent with identically distributed gap times. The distribution of the gaps was approximately exponential for all service units with means nearly equal to standard deviations, and skewness averaged about 2. These characteristics suggest the injury gap data may be analyzed using standard survival analysis methods.

Table 11 lists results of Cox regressions of data on the gaps between injuries for the four service units. Analyses combined the Baseline and First Phase data for comparison with the Second Phase to contrast the time before CAB–Switching with the time during CAB–Switching. Correlations between weighted Schoenfeld residuals and the gaps was very low and nonsignificant indicating the proportional hazard assumption of the Cox regressions was met.

Table 11. Cox Regression of Injuries over Time for Each Phase and Service Unit

Phase		Service Unit			
		San Antonio	Fort Worth	Houston	Livonia
Baseline and First Phase Gradual Trend	*n*	248	212	284	208
	b	-0.000443	-0.000424	-0.000361	-0.000513
	Δ*r**	-61%	-59%	-53%	-66%
	p	0.000062	0.000401	0.000214	0.000017
Second Phase Gradual Trend	*N*	43	34	26	25
	b	-0.000597	-0.00104	-0.000404	0.00183
	Δ*r***	-27%	-42%	-19%	158%
	p	0.569	0.402	0.775	0.225
	* Change in incidence rate over 2099 days (Baseline Phase and First Phase). ** Change in incidence rate over 517 days (Second Phase).				

All service units experience significant gradual improvements in their injury rates of about the same magnitude from Baseline through the First Phase. These trends continued into the Second Phase for most service units, including SASU, but were not significant given the smaller sample sizes.

A Cox regression compared SASU's trend against the other service units combined by using date, service unit (SASU or not), and their interaction as predicting variables. This regression detected no significant service unit by date interactions for either the Baseline and First Phase together ($\chi^2(1) = 0.11$, $p = 0.740$) or the Second Phase alone ($\chi^2(1) = 0.12$, $p = 0.732$). San Antonio does not have a significantly different rate of change.

6.4.3 Incidents

Incident data were filtered to exclude the relatively few incidents associated with road operations (9 percent of all incidents) in order to focus on the occurrences most likely to be impacted by CAB-Switching.

6.4.3.1 San Antonio versus Other Service Units

Figure 21 shows the normalized cumulative incidence plot of the incident data for the SASU and each of other original service units of the Southern Region. For clarity, the first two years of the baseline data are omitted. All analyses include baseline data back to January 2001.

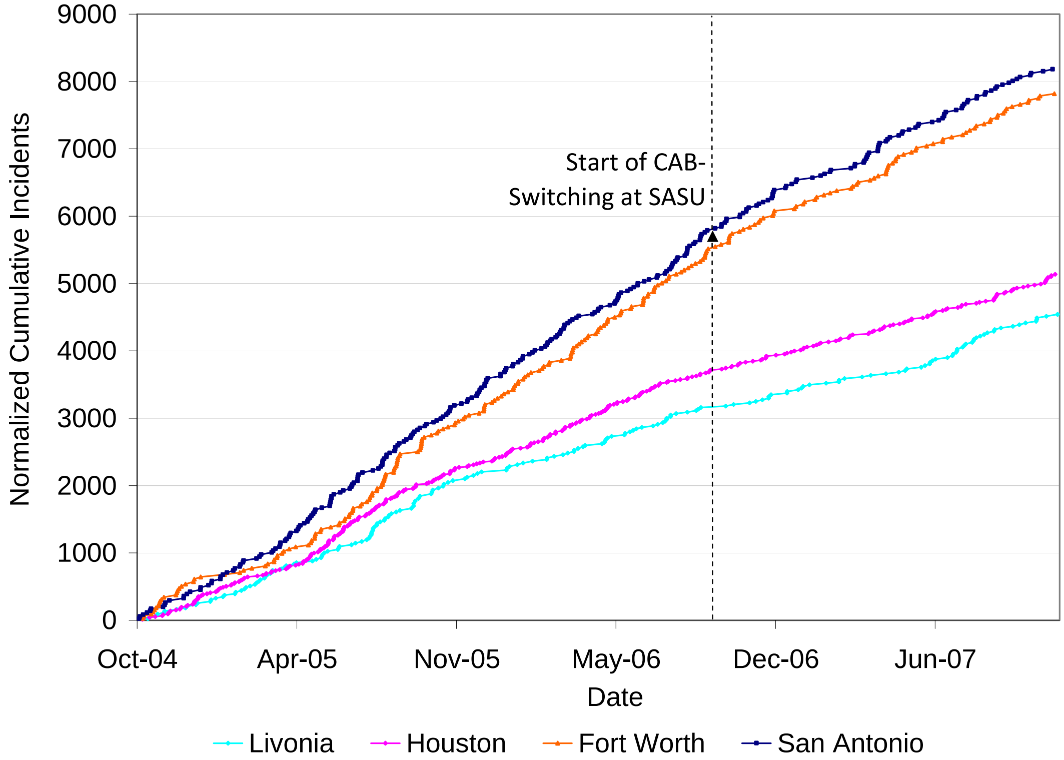


Figure 21. Normalized Cumulative Incidence Plot of Incidents in the Original Southern Region

Incident rates are generally steady for all service units from the Baseline Phase through the start of the Second Phase, with the exception of a spike in incidents at all service units in the summer of 2005. Qualitative data did not provide an explanation for these spikes. Changes in the incidents rates since the start of CAB-Switching are very weak, if present at all, although it appears that San Antonio and Fort Worth may have gradually decreasing rates, while Livonia and Houston may each have had spikes of incidents in 2007.

Autocorrelations of the log-transformed gaps between incidents were low and on average not significant (range 0.004 to 0.124, average $r = 0.064$, $t(7) = 0.668$, $p = 0.526$), which is consistent with the gap times being statistically independent. Weibull plots were generally straight lines, consistent with identically distributed gap times. The distribution of the gaps was approximately exponential for all service units with means nearly equal to standard deviations, and skewness averaged about 2. These characteristics suggest the incident gap data may be analyzed using standard survival analysis methods.

Table 12 lists results of Cox regressions of date on the gaps between incidents for the four service units. Analyses combined the Baseline and First Phase data for comparison with the Second Phase to contrast the time before CAB–Switching with the time during CAB–Switching. Correlation between weighted Schoenfeld residuals and the gaps was very low and nonsignificant indicating the proportional hazard assumption of the Cox regressions was met.

Table 12. Cox Regression of Incidents over Time for Each Phase and Service Unit

Phase		Service Unit			
		San Antonio	**Fort Worth**	**Houston**	**Livonia**
Baseline and First Phase Gradual Trend	*n*	479	429	579	295
	b	0.000211	0.000221	0.000248	0.000233
	Δ*r**	56%	59%	68%	63%
	p	0.0070	0.0065	0.0005	0.0182
Second Phase Gradual Trend	*N*	82	70	95	58
	b	-0.00133	-0.00107	0.00148	0.00227
	Δ*r***	-43%	-39%	89%	228%
	p	0.145	0.283	0.080	0.019

* Change in incidence rate over 2099 days (Baseline Phase and First Phase).
** Change in incidence rate over 429 days (Second Phase).

All service units showed approximately the same increase in the rate of incidents during the Baseline and First Phase, which may be attributed to the summer 2005 spike in incidents. During the Second Phase, however, San Antonio and Fort Worth showed a nonsignificant decrease in rate, while Houston and Livonia showed an increase, with only Livonia's increase being significant.

To compare these changes among the services units, a Cox regression tested for a site by date interaction on these trends using the following two models:

$$\textit{Model 1: } h(g) = h_0(g)\exp(\,b_d d + b_s s\,)$$

$$\textit{Model 2: } h(g) = h_0(g)\exp(\,b_d d + b_s s + b_{ds} ds\,)$$

where:

g = The gap times.

$h(g)$ = Hazard function for the gaps between incidents.

$h_0(g)$= Base hazard function

d = Date

s = Site (dummy coded)

b_j = Cox regression coefficients.

This regression showed that the second model, which included the date × site interaction term, represented a significantly better fit to the data (Log likelihood from -1422.838 to -1417.664, $\chi^2(3) = 10.343$, $p = 0.016$) than did the first model. This indicated that the change in the rate of incidents significantly depends on the service unit in the Second Phase.

Pairwise comparisons showed that the reduction at SASU was significantly greater than Livonia ($\chi^2(1) = 7.469$, $p = 0.006$) and Houston ($\chi^2(1) = 4.197$, $p = 0.040$), but not Fort Worth ($\chi^2(1) = 0.080$, $p = 0.777$). The reduction at Fort Worth was significantly greater than Livonia ($\chi^2(1) = 6.299$, $p = 0.012$), but not significantly greater than Houston ($\chi^2(1) = 2.977$, $p = 0.084$). Livonia and Houston were not significantly different ($\chi^2(1) = 1.024$, $p = 0.312$).

In summary, SASU showed the greatest reduction in the rate of incidents, significantly greater than two out of the three comparisons service units. SASU was not significantly different than the next best service unit, Fort Worth, which was significantly better than only one other service unit. In an analysis that combined the three comparison service units together, there was a significant date by site interaction ($\chi^2(1) = 4.325$, $p = 0.038$), indicating that SASU's reduction in the rate of incidents during Phase 2 was significantly greater than all comparison sites together.

During the 4-year Baseline Phase, all service units showed significant increases in the rate of incidents, which may be attributed to the spike in incidents in the summer of 2005. As shown in Figure 21, rates for all service units returned to normal rates before the Second Phase. The decreased rate exhibited by SASU and Fort Worth do not appear to be a continuation of an earlier trend.

6.4.3.2 Eagle Pass, San Antonio Complex, and Laredo within SASU

UP provided monthly data for cars moved for each yard, allowing comparisons to be made among the large yards within SASU: the San Antonio Complex, Eagle Pass, and Laredo. All three locations received implementations of CAB–Switching but featured different levels of implementation during Second Phase, thus providing a natural quasi-experiment for comparing the impact of CAB–Switching. Eagle Pass and the San Antonio Complex both began the CAB process at the beginning of the Second Phase, but Eagle Pass reached advanced implementation sooner than the San Antonio Complex (see CAB Scope Expanded Beyond CRZ and the San Antonio Terminals, page 68 and CAB Process Metrics, page 80). Thus Eagle Pass represents a strong CAB implementation and the San Antonio Complex represents a moderate CAB implementation. Laredo began CAB–Switching in the last few months of the evaluation period. It represents the weakest CAB implementation when comparing the three locations over the Second Phase.

Figure 22 shows the normalized cumulative incidence plot of the incident data for the three yard locations.

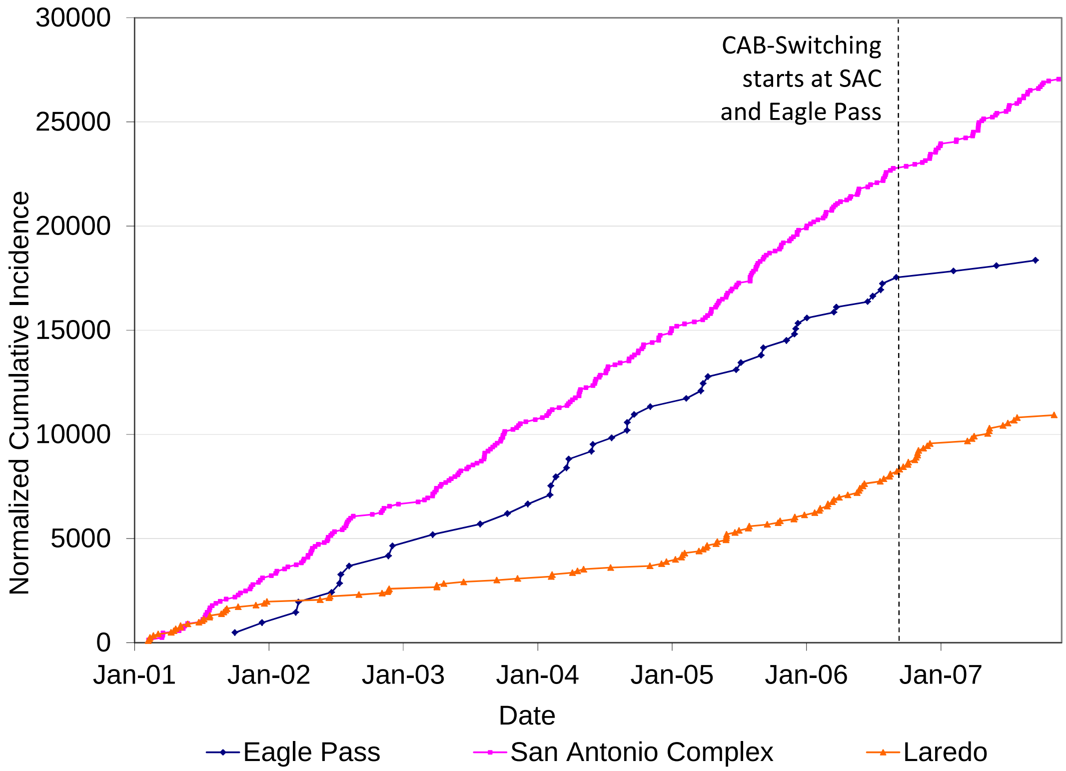


Figure 22. Normalized Cumulative Incidence Plot of Incidents at the Largest SASU Yards

Over the four years of the Baseline Phase, Eagle Pass and the San Antonio Complex showed relatively constant rates of incidents. Data for Laredo, however, showed that the rate from January 2002 through September 2004 was steady at 1.66 incidents per 100,000 car moves on average, an uncharacteristically low value for any yard. It was about a quarter of the rate for Laredo for October 2004 through the end of the Baseline Phase in September 2006, when the rate averaged 6.50 incidents per 100,000 car moves (Poisson rate comparison (Nelson, 1982) $F(42,96) = 3.734$, $p < 0.001$). We have no explanation for the significantly low rate in the earlier portion of Baseline and field notes provide no indication of Laredo ever having remarkably low rates. Possibly some data were excluded from our data set, although there were no obvious missing categories of data. In any case, in order to provide a stable baseline for Laredo, the baseline for all locations was limited to August 2004 through September 2006.

Figure 23 shows the normalized cumulative incidence plot of the incident data for three yard locations with the modified baseline. Each location has a different overall rate of incidents, with the San Antonio Complex having the highest rate and Laredo have the lowest rate, as represented by the different steepness of the lines. These differences may be due to characteristics of the geometry of the yards or type of work performed there.

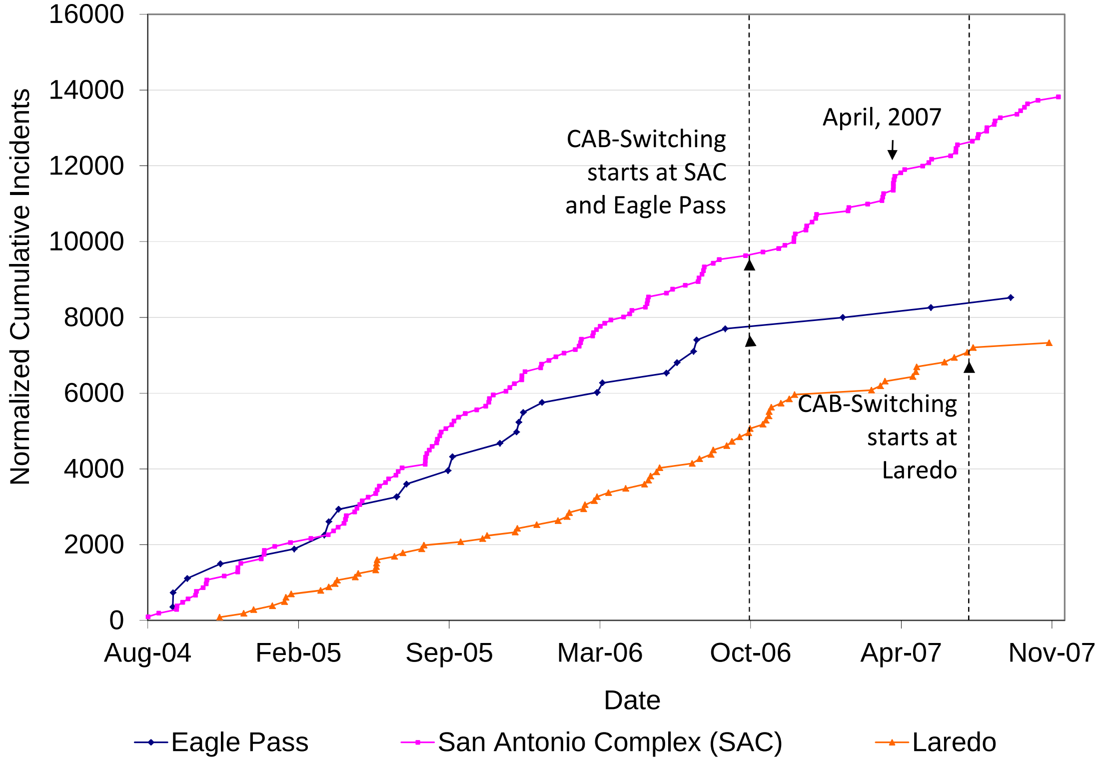


Figure 23. Normalized Cumulative Incidence Plot of Incidents Since August 2004

At the start of the Second Phase, San Antonio Complex, which had a moderately strong implementation during the Second Phase, appears to have a slight decrease in the incident rate, except during April of 2007, for which there is a cluster of incidents. In April 2007, SASU experimented with a program that brought furloughed non-yard workers to assist with yard work in the San Antonio Complex. Other yards did not participate. The resulting spike in derailments led management to abandon the program. Data for April 2007 were excluded from all locations to make them more comparable to each other.

Eagle Pass, which had the strongest implementation during the Second Phase, experienced a dramatic decrease in the incident rate. The yard had only three incidents after implementing CAB-Switching. Laredo does not appear to have a clear change in the rate of incidents, although there were a couple streaks of no incidents during the Second Phase, one shortly after Laredo started CAB-Switching. However, the time following the start of CAB-Switching at Laredo was too brief to allow a meaningful comparison of incidents before and after the start of CAB-Switching there.

Autocorrelations of the log-transformed gaps between incidents were low and on average not significantly different from zero (range –0.074 to 0.112, average $r = 0.029$, $t(5) = 0.162$, $p = 0.878$), which is consistent with the gap times being statistically independent. Weibull plots were generally straight lines, consistent with identically distributed gap times. The distribution of the gaps was approximately exponential for all service units with means nearly equal to standard deviations, and skewness averaged about 2. These characteristics suggest the incident gap data may be analyzed using standard survival analysis methods.

The changes in the gaps between incidents for each location were analyzed with a parametric Weibull regression. A parametric model was favored over a Cox regression due to the small number of incidents at Eagle Pass during the Second Phase, which can lead to imprecise parameter estimation (Box-Steffenmeier and Jones, 1997), and, indeed, a Cox regression of these data resulted in implausibly large coefficients for Eagle Pass. Weibull plots of the data for each site resulted in relatively straight lines implying that a Weibull distribution would adequately fit the gap data. The shape parameter (α) for all regressions was not significantly different from 1, so the regressions were re-run as exponential regressions (α fixed at 1). Table 13 shows the results of separate exponential analyses for each location.

Table 13. Exponential Regressions for Human Factors Incidents at SASU Yards

Phase		Service Unit		
		Eagle Pass	**San Antonio Complex**	**Laredo**
Baseline and First Phase	n	24	234	92
Gradual Trend	b	0.00046	–0.00026	–0.00144
	Δr*	–30%	23%	213%
	p	0.615	0.566	0.031
Baseline and First Phase	n	28	136	69
Versus Second Phase	b	1.681	0.392	0.030
	Δr**	–81%	–32%	3%
	p	0.006	0.042	0.909

* Change in incidence rate over 791 days (Baseline Phase and First Phase).

** Change in incidence rate between Baseline through First Phase versus Second Phase.

Figure 24 shows the incident rates per 100,000 car moves at the three locations during the Second Phase and before the Second Phase (Baseline and First Phase). These rates were calculated as the total number of incidents divided by the total number of car-moves during the relevant stretches of time. Both Eagle Pass and the San Antonio Complex experienced a significant decrease in incident rates between the two phases. The strong implementation of CAB–Switching at Eagle Pass was associated with an 81-percent decrease in the rate of incidents over a 1-year period while the San Antonio Complex experienced a more modest 32-percent decrease. The changes at Laredo were not significant.

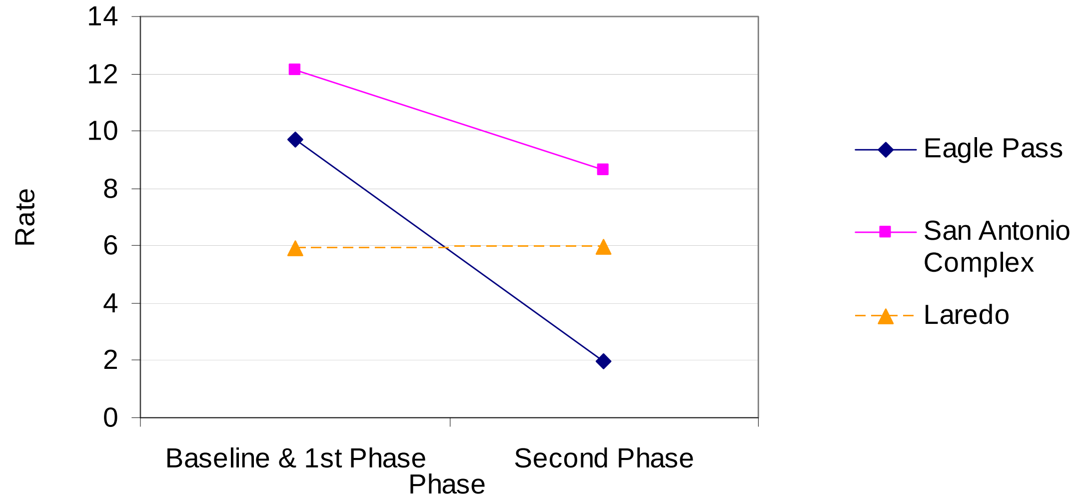


Figure 24. Rates of Human Factors Incidents at SASU Yards (Dashed Lines Indicate No Significant Difference)

To compare the changes in rates among the three locations, a Weibull regression analysis was performed. The parametric models for the regression were as follows:

$$\textit{Model 1: } h(g) = g^{\alpha} \exp(b_p p + b_s s)$$

$$\textit{Model 2: } h(g) = g^{\alpha} \exp(b_p p + b_s s + b_{ps} ps)$$

where:

g = The gap times

$h(g)$ = Hazard function for the gaps between incidents

p = Phase (0 = baseline and First Phase, 1 = Second Phase)

s = Location (dummy coded)

b_j = Regression coefficients

α = Shape parameter for the Weibull distribution

The regression showed that the second model, which included the phase × location interaction term, represented a significantly better fit to the data (Log likelihood from -365.437 to -369.312, $\chi^2(2) = 7.750$, $p = 0.0208$) than did the first model. This indicated that the change in the rate of incidents within SASU differed across the locations. Specifically, the change at Eagle Pass was significantly better than the change at the San Antonio Complex ($\chi^2(1) = 4.19$, $p = 0.0406$) and at Laredo ($\chi^2(1) = 6.06$, $p = 0.0138$), while the change at the San Antonio was not significantly better than Laredo ($\chi^2(1) = 1.20$, $p = 0.2733$). In other words, the location with the strongest CAB-Switching implementation experienced a significantly greater reduction in the incident rate than the locations with moderately strong implementation or the weakest implementation in the Second Phase.

During the 4-year Baseline Phase, all locations showed no significant changes in the rate of incidents, except that Laredo exhibited a significant increasing trend. The decrease in rates at Eagle Pass and the San Antonio Complex in the Second Phase does not appear to be a continuation of an earlier trend.

6.4.4 Decertifications

The decertification data were filtered to limit the analysis to those decertification types most closely associated with CRZ (namely decertifications for failure to stop, excessive speed, and violation of main track authority) and to those decertifications in which the engineer is considered at fault in management's investigation of the decertification. This excludes decertifications for drugs and alcohol, for failure to conduct an air brake test, and for tampering with the locomotive alerter, all of which involved behaviors not targeted by CAB–CRZ. It also excludes decertifications in which an investigation determined that the engineer was not at fault and decertifications for which the investigation had yet to determine fault.

Figure 25 shows the normalized cumulative incidence plot of the decertification data for the SASU and all other original service units of the Southern Region combined. For clarity, the first two years of the baseline data are omitted.

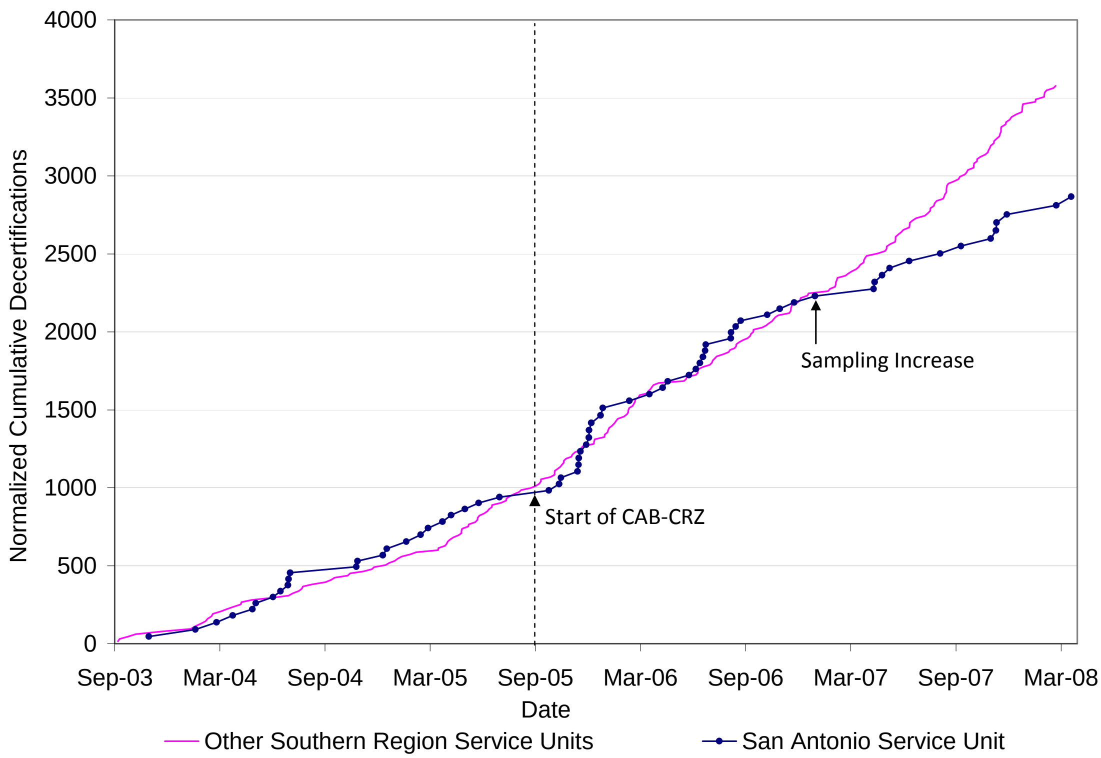


Figure 25. Normalized Cumulative Incidence Plot of Decertifications over Time

Through the start of CAB-CRZ in September 2005, Figure 25 shows that the service units exhibited a slightly increasing rate of decertifications. However, sometime after the start CAB-CRZ, the decertification rate for SASU began to decrease, departing visually from the other service units by January 2007, which corresponds to the increase in training and sampling due to the budget increase (see CAB Budget Increase Gives Process Needed Boost on page 71). Plots of separate normalized cumulative incidences for each service unit in the Southern Region showed that SASU has the same divergence from every other service unit, but not consistently around January 2007.

Autocorrelations of the log-transformed gaps between decertifications were low and nonsignificant on average (range –0.072 to 0.166, average $r = 0.0757$, $p = 0.7093$), which is consistent with the gap times being statistically independent. Weibull plots were generally straight lines, consistent with identically distributed gap times. The distribution of the gaps was approximately exponential for all service units with means nearly equal to standard deviations, and skewness averaged about 2. These characteristics suggest the decertification gap data may be analyzed using standard survival analysis methods.

Table 14 lists results of Cox regressions of time on the gaps between decertifications for the four service units. Correlation between weighted Schoenfeld residuals and the gaps was very low and nonsignificant indicating the proportional hazard assumption of the Cox regressions was met.

Table 14. Cox Regression of Decertifications over Time for Each Phase and Service Unit

		Service Unit			
Phase		**San Antonio**	**Fort Worth**	**Houston**	**Livonia**
Baseline Phase	*n*	45	50	54	37
Gradual Trend	*b*	< 0.00001	0.00007	-0.00003	0.00036
	Δr*	0%	13%	-5%	83%
	p	0.997	0.817	0.924	0.965
First and Second Phase	*N*	42	72	74	37
Gradual Trend	*b*	-0.00168	< 0.00001	-0.00114	0.00116
	Δr**	-79%	0%	-22%	193%
	p	0.014	1.000	0.543	0.109

* Change in incidence rate over 1690 days (Baseline Phase).
** Change in incidence rate over 943 days (First and Second Phase).

SASU showed a significant 79-percent decrease in the rate of decertifications through the First and Second Phases from September 2005 (when CAB started) through the end of available data in March 2008. The other three service units of the original Southern Region (Fort Worth, Houston, and Livonia) showed no significant change in the rate of decertifications. To compare the changes between SASU and the other services units, a Cox regression tested for a site by date interaction on these trends using the following two models:

$$\textit{Model 1: } h(g) = h_0(g) \exp(\, b_d d + b_s s \,)$$

$$\textit{Model 2: } h(g) = h_0(g) \exp(\, b_d d + b_s s + b_{ds} ds \,)$$

where:

g = The gap times

$h(g)$ = Hazard function for the gaps between decertifications

$h_0(g)$= Base hazard function

d = Date

s = Site (1 = SASU, 0 = Other)

b_j = Cox regression coefficients

This regression showed that the second model, which included the date × site interaction term, represented a significantly better fit to the data (Log likelihood from -981.697 to -979.256, $\chi^2(1) = 4.882$, $p = 0.027$) than did the first model. This indicated that the rate of decertifications at SASU decreased significantly more than at the other service units.

During the 4-year Baseline Phase, all service units showed no significant changes in the rate of CRZ-associated decertifications. The improvements at SASU do not appear to be a continuation of an earlier trend.

During the First and Second Phases, there were no significant changes in the rate of other decertifications (drugs and alcohol, air brake testing, and alerter tampering) for any of the original Southern Region service units (the smallest observed p value was 0.1526). There tended to be fewer of these kinds of decertifications.

Fort Worth and Houston service units showed significantly lower rates of decertifications during the Baseline Phase than during the First and Second Phases combined. This difference appeared to be associated with lower decertification rates early in the Baseline Phase before 2004. There were no significant Service Unit by Phase interactions in the Baseline Phase.

6.5 Forced-Choice Survey

6.5.1 Unsafe Behaviors-CRZ

Table 15 lists the number of respondents for the Unsafe Behaviors–CRZ scale.

Table 15. Number of Respondents for the CRZ Scale for Each Phase and Respondent Type

		Respondent Type	
		Worker	**Manager**
Phase	First	174	16
	Second	82	26

The smaller numbers (compared with overall survey returns) reflect the fact that only the subset of respondents performing roadwork (or supervising others performing roadwork) were instructed to answer the Unsafe Behaviors–CRZ questions.

The Unsafe Behaviors-CRZ scale had a Cronbach's alpha of 0.837, indicating acceptable reliability (Rosenthal and Rosnow, 1991).

A 2×2 ANOVA evaluated changes in Unsafe Behaviors-CRZ with Respondent Type (Worker or Manager) and Phase (First and Second) as independent variables. The analysis showed that self-reported CRZ practices were significantly more compliant with the operating rules for the Second Phase ($M = 3.988$) than for the First Phase ($M = 3.749$, $F(1, 294) = 5.739$, $p = 0.017$). There was no significant effect of Respondent Type ($F(1, 294) = 0.049$, $p = 0.825$) or interaction between Respondent Type and Phase ($F(1, 294) = 1.079$, $p = 0.300$).

6.5.2 Unsafe Behaviors-Switching

Forty-two workers and 19 managers completed the Unsafe Behaviors–Switching scale. The smaller numbers (compared with overall survey returns) reflect the fact that only the subset of respondents performing yard work (or supervising those workers performing yard work) were instructed to answer the Unsafe Behaviors–Switching questions.

The analysis of the Unsafe Behavior–Switching scale is different than for the Unsafe Behavior–CRZ scale, because no data are available from the First Phase. Instead, respondents were asked first to indicate their current switching behaviors and then to indicate retrospectively their behaviors before the CAB process started. The Unsafe Behaviors–Switching scale was found to have a Cronbach's alpha of 0.762 for current behaviors and 0.787 for retrospective behaviors, indicating weak but adequate reliability (Rosenthal and Rosnow, 1991).

Change in Unsafe Behaviors–Switching was evaluated with a 2×2 mixed-design ANOVA with Respondent Type (Worker or Manager) as a between-subjects independent variable and Time (Current and Retrospective) as a within-subjects independent variable. Workers reported overall greater compliance for themselves with switching rules than managers did for their subordinates ($M = 4.655$ and 4.342, respectively; $F(1,59) = 4.919$, $p = 0.030$). There were no effects for Time ($F(1,59) = 0.014$, $p = 0.906$) or the Respondent-Type-by-Time interaction ($F(1,59) = 0.277$, $p = 0.601$). Respondents reported no difference in their compliance with switching rules between the time of when the survey was administered and what they recalled doing before the start of CAB. Scores in general were very near the maximum value mathematically possible for both scales (5.0).

6.5.3 Propensity to Safe Behaviors

The Propensity to Safe Behaviors scale had a Cronbach's alpha of 0.500, indicating inadequate reliability. Evaluating the reliability for managers only or workers only did not improve Cronbach's alpha sufficiently (0.611 and 0.524, respectively).

Consistent with inadequate reliability, there were no significant effects in a 2×2 ANOVA with Respondent Type (Worker or Manager) and Phase (First and Second) as independent variables (the smallest observed p-value was 0.410).

6.5.4 Coworker Safety

The Coworker Safety scale had a Cronbach's alpha of 0.710, indicating marginally adequate reliability.

There were no significant effects in a 2×2 ANOVA with Respondent Type (Worker or Manager) and Phase (First and Second). The strongest effect was for the interaction, which had an observed p-value of 0.062. Scores for Coworker Safety exhibited excessive heterogeneity of variance (Levene's $F(3,303) = 8.368$, $p < 0.001$), calling into question the p-values from this ANOVA (Levene's F for the other scales was not significant). Analyses using t-tests with separate variance estimates showed no significant effects, with the interaction's p-value (still having the strongest effect) increasing to 0.201.

In addition to the relatively low reliability, the Coworker Safety scale scores were generally close to the maximum value mathematically possible for both phases, with an overall average of 4.087 on a five-point scale.

6.5.5 Labor–Management Relations

The Labor–Management Relations scale had a Cronbach's alpha of 0.809, indicating modest reliability.

A 2×2 ANOVA evaluated changes in Labor–Management Relations with Respondent Type (Worker or Manager) and Phase (First and Second) as independent variables. The analysis showed that management–labor relations were significantly better for the Second Phase ($M = 2.881$) than for the First Phase ($M = 2.649$, $F(1, 303) = 11.123$, $p = 0.001$). Managers generally saw relations as being significantly better than did workers (Ms = 3.472 versus 2.378, $F(1, 303) = 56.296$, $p < 0.001$). There was no significant interaction between Respondent Type and Phase ($F(1, 303) = 0.050$, $p = 0.823$).

6.6 Interviews

Themes found in the analyses of the interviews are shown in *italics.*

6.6.1 Implementation

Qualitative analyses of the interviews found the following implementation themes:

- Employee acceptance of CAB
- Management commitment to and resources for CAB
- Effectiveness of CAB training
- Inclusiveness of CAB's scope
- Role of CAB communications
- Adaptations with and by CAB's steering committee

6.6.1.1 Employee Acceptance of CAB

During all three annual interviews, workers and managers indicated *increasing acceptance of CAB by employees and the unions.* Resistance was confined to small pockets by the end of the evaluation period. For workers and managers, concerns about acceptance during the final interviews in winter 2007–2008 were about the same as they were during the midterm interviews in winter 2006–2007. This was less than the initial interviews in winter 2005–2006.

Worker, Initial Interviews

> "Those against CAB say 'it's a finger-pointing program'—gives bad information—shows that workers are just so reckless and not so good at what they do."

Worker, Midterm Interviews

> "Workers are committed to a high degree. Only a few don't want CAB. If they know about it, they're positive about it."

Worker, Final Interviews

> "The old heads are more willing to listen. They still think its going to end…but they think if CAB doesn't work we may not get another chance. There is a sense of urgency."

Workers and managers alike ascribed greater resistance to long-time workers than newer workers, a problem expected to solve itself with the shifting demographics of the service unit.

Worker, Final Interviews

> "Newer workers are sold on it. The midlevel seniority guys are somewhat positive and somewhat skeptical still. Then the old heads are the most skeptical. The old heads that are supportive are the ones who have gone to the training."

Manager, Final Interviews

> "The older employees are becoming fewer and fewer…. As these older employees retire, there will be less and less people talking against CAB."

Reasons for resistance may also have changed. During the initial interviews, workers were primarily concerned with the use of CAB data—that it may be turned over to management and be used for disciplinary or FTX purposes. By the final interviews, workers appeared to be largely assured of CAB's control of the data, and their concerns did not gravitate towards any specific issues.

Worker, Initial Interviews

> "People think it's a Nazi police program. There's no way legally to see their name. But if you know the sampler's ID number, the time of day and the weather [which are recorded on the checklist] they think they can be identified."

Worker, Final Interviews

> "There are some guys who have things against the company from 20 years ago—things that prevent someone from getting involved in CAB."

In general, workers seemed to see more employee resistance than did managers.

Manager, Midterm Interviews

> "Workers, they are very involved. I don't see any lack of commitment to making it work."

Worker, Midterm Interviews

"There's a lot more commitment now than when this started. On a scale of 1 to 10, we were at 5, now we're at 8. From the newer guys, I really see them want to be committed."

6.6.1.2 Management Commitment to and Resources for Cab

Across all three annual interviews, workers and managers indicated *increasing support of CAB by management*. By the final interviews, most workers and managers saw management as sufficiently committed to CAB, in contrast to midterm interviews, when workers were more evenly split on whether they thought management was committed or not.

Worker, Midterm Interviews

"CAB is weakening some from the manpower budget side. I feel management is just waiting for the process to take its course so they can do away with it."

Manager, Midterm Interviews

"Their resources are good. Anything they've needed. I've heard no one gripe about it."

Worker, Final Interviews

"Management has opened doors that weren't open before. Management has allowed workers to be involved with the CAB process, and let samplers do the observations a couple days a month, and being paid to do it.… I think that the road blocks—management resistance—have been brought down. Management wants this to become stronger."

Manager, Final Interviews

"Management is committed to CAB…. Some of the managers have been more active in their support and it has worked…. They talk it up. They encourage it."

Concerns about CAB's manpower budget were noticeably absent from the final interviews in contrast to midterm. The increase in the budget (see page 71) did much to increase the perceived commitment from management.

Worker, Midterm Interviews

"What we need is more budget to cover the outlying areas not covered yet. We are committed with classes in San Antonio and Eagle Pass, but there are areas not covered."

Worker, Final Interviews

"At the beginning, it was the budget for CAB. The steering committee couldn't do much because it didn't have much. Now, CAB has a lot better budget, and because of the budget CAB has been able to expand in other areas."

The concern about management commitment in the final interviews was more vague and did not point to any consistent issue.

Worker, Final Interviews

"I don't think that managers have much of a commitment at all. The managers tolerate us. The managers won't interfere with our CAB sampling and won't [FTX] test right after, but that doesn't mean that they trust CAB."

6.6.1.3 Effectiveness of CAB Training

From initial through final interviews, workers consistently stated that *CAB training is especially effective for promoting the CAB process and safety awareness in general*.

> *Worker, Initial Interviews*
>
> "Going through the program really changed attitudes. After going through it, it kind of changed their minds…. To limit employee resistance, they need to go through the class."
>
> *Worker, Midterm Interviews*
>
> "Ninety-nine of those on the fence walk out of training thinking, 'CAB is not what I thought it was.' It's very successful."
>
> *Worker, Final Interviews*
>
> "Going through the class opens your eyes…. They are learning about CRZ and talking about it.… As guys go through class they get committed."

Managers likewise saw the value of CAB training in teaching safety and overcoming resistance, but they did not mention CAB training much in the final interviews, perhaps taking it as a given.

> *Manager, Initial Interviews*
>
> "I think some of the change is from the people that have been through the sampling class. They've nothing but positive things to say about it, and see it relevant to their own behavior, versus being us-versus-them, where the company is responsible for safety."
>
> *Manager, Midterm Interviews*
>
> "The more they teach on CAB, the more we get buy in, educating people on what it is…changing at-risk behavior through a nondisciplinary process."

6.6.1.4 Inclusiveness of CAB's Scope

In both the initial interviews and the final interviews, many workers and managers expressed *a need to broaden the scope of CAB*. This included going beyond CRZ on the road, involving more workers, running more feedback sessions, involving managers more, and expanding CAB to other locations. In the initial interviews, such statements included a tone of concern.

> *Worker, Initial Interviews*
>
> "The CAB process doesn't work in the yard. The CAB–CRZ checklist doesn't apply to the yard. So we don't do it. The intent is to have sampling in the yard but the checklist doesn't work."
>
> *Manager, Initial Interviews*
>
> "The only thing—only criticism—would be more samples. The more often you sample, the more employees will be cognizant of the sampler's own behavior."

By the final interviews, as CAB–Switching was being introduced during the Second Phase, the tone shifted to ambition and to expanding CAB beyond its original vision.

Worker, Final Interviews

"I think that it could be expanded to other locations that don't have it. To make it stronger, get all the employees involved. It will be better."

Manager, Final Interviews

"I think the railroad needs some responsibility—involvement in the steering committee meetings, where the managers can use information gathered in the steering committee meeting to fix the problem. Management doesn't want to use the information to swarm on guys [e.g., with FTX testing], but they need to discover the root causes to address while keeping their hands off the CAB process."

6.6.1.5 Role of CAB Communications

Although responses during all three annual interviews emphasized the value of CAB training, the responses in final interviews also indicated that *additional media were effective in promoting and informing workers about CAB*. This included face-to-face contact with samplers and steering committee members, the CAB newsletter, and the Web site.

Worker, Final Interviews

"The newsletter, the word of mouth and the attending class are all good communication vehicles. Sometimes the Facilitators will put out a safety briefing and announce the next class and other stuff and this always helps."

Manager, Final Interviews

"There is one person who goes in and talks about it. You need guys who are faithful and talk about it. The Facilitators and a steering committee member are among the faithful. They are leaders of it and convert people to their cause."

6.6.1.6 Adaptations with and by CAB's Steering Committee

Absent in the earlier interviews, but present in the final interviews, some workers and managers reported that the *CAB process or steering committee procedures need some tweaking*. Many of these tweaks related to increasing the sustainability of CAB.

Worker, Final Interviews

"The decentralization and getting the locations to be more autonomous is also potentially good…. The way I think it should go is to have each of the locations, Laredo, Eagle Pass, Del Rio, etc., own their own process. Wouldn't necessarily need to have a BLET and a UTU leader at each location. There would just have to be general representation of the two unions across all the locations."

Manager, Final Interviews

"Hopefully they will move away from taking Saturday off [as an incentive for CAB training]. They will tweak things and make the training on a different day of the week or those going to training will have to sign on Friday after class."

Rotating the facilitators and the steering committee members and bringing in new workers systematically in particular were regarded as important changes for long-term viability.

Worker, Final Interviews

"Currently, there are people on the committee who are only there to ride the committee. That is sad. Now the committee is rotating people out and putting new people in. Hopefully, it will be a better committee."

6.6.2 Outcomes

Qualitative analyses of the interviews found the following outcome themes:

- Improvements in worker practices
- Generalization and internalization of new safety attitudes
- Improvements in manager practices
- Improvements in labor–management relations

During the interviews respondents, managers in particular, also recognized decreases in occurrences from their own monitoring of the corporate safety data. These observations largely paralleled the actual decreases described in Corporate Safety Data on page 86.

6.6.2.1 Worker Practices

Most workers and many managers credited CAB with *improving safety behaviors, particularly related to CRZ*. There were several reports of practice improvements from the initial interviews.

Worker, Initial Interviews

"I know it has changed people's behaviors. I heard accounts of, 'This is how I was before, this is how I am now.'"

Manager, Initial Interviews

"If you were to ride a train, everyone is calling signals now. Now they get to calling the train ID, but calling signals and watching trains—that has changed for the better."

By the midterm interviews, descriptions of changes in practices were more numerous and specific.

Worker, Midterm Interviews

"Lately a lot of guys aren't using the cell phone any more. Not everybody, but more are not doing it, and not being distracted in the CRZ and out of it."

Manager, Midterm Interviews

"I've seen a big improvement in the conductor logbooks. The logbook is organized now and filled out with greater detail than before [as required by CRZ procedures]. I think their logbooks are more filled out than at other locations I've seen."

This theme continued into the final interviews, expanding to include other practices with the broadening in CAB's scope.

Worker, Final Interviews

"I've seen conductors becoming more involved and see that the workers are getting safer. Examples include: Conductors calling out the signal, paying attention to what the crew behind their train is doing, showing 100-percent alertness."

Manager, Final Interviews

"The switching procedures seem to be being followed more. There are fewer exceptions [FTX failures]. Everyone here [at a particular yard] is working safer."

6.6.2.2 Generalization and Internalization

At the beginning with the initial interviews, respondents noted a substantial change in worker's safety attitudes in which they credited CAB with *increasing awareness about safety and promoting personal responsibility for ensuring safety*. Starting with the CAB training, workers reportedly began talking more about safety, raising issues, and discussing what they could do to work more safely.

Worker, Initial Interviews

"I've heard more and more people talking about safety and CAB. There's a certain freedom to discuss safety and risk. People describe going from at-risk to safe behavior."

Manager, Initial Interviews

"I think crews have been more safety conscious in general, not just regarding CRZ. Crews tell managers things that are safety concerns now. That may be due to CAB."

This safety awareness theme continued through the midterm interview, with respondents saying workers had a greater sense of understanding and control over their safety.

Worker, Midterm Interviews

"I see guys in class talking about their own behavior and seeing it as something they can choose to do. That has more bang for the buck than CRZ. It's more effective than FTX by talking about behavior rather than rules."

Manager, Midterm Interviews

"People are more aware of everything. They always knew it's a dangerous job. Now they're more aware that it's a dangerous job. Training and education, this tells you that. They give great examples. Before, it was just, 'This is a dangerous job. Don't do this.'"

During the final interviews, workers echoed this awareness.

Worker, Final Interviews

"CAB has made me aware of doing things around the house more safely. My wife has told me to wear my safety goggles, and being careful when working around ladders to take off Christmas tree lights."

Worker, Final Interviews

"The managers are seeing the difference out there and hear the workers talking about safety and say, 'Hey that's what we are concerned about out there.'"

In contrast, managers no longer reported increases in awareness and personal responsibility, perhaps viewing such behavior as the normal state of affairs.

6.6.2.3 Manager Practices

Management safety practices take two forms. The first form is safety coaching of workers, which is closely tied to labor–management relations and is described as a separate outcome. The second form is removal of safety barriers in the workplace through changes in training, procedures, tools, and the environment.

In the initial interviews, workers were asked to describe their chief safety concerns. Two concerns topped the list: fatigue, especially as related to the work schedule, and the conditions of the facilities and equipment.

To the workers, fatigue on the job often results from an inability to adequately plan rest before or after work because of an unpredictable work schedule. Workers monitor train lineups (expected times when trains will arrive at a terminal) using UP computers and try to estimate when they are likely to be called for work. The initial feedback was that lineup data were unreliable.

> *Worker, Initial Interviews*
>
> "Train lineups—the accuracy of departure times for when workers are to leave compared to what it should be—is off by a couple of hours. This affects rest and family time."

Conditions of the facilities and equipment were the second greatest concern.

> *Worker, Initial Interviews*
>
> "The condition of the yard. There is debris all over the yard, the gondola cars are overfull, and then there are gigantic pieces of metal lying around. There are holes two feet deep from work done by Maintenance-of-Way. You could break a leg if you fell in one of these holes."

Consistent air conditioning for the train cabs was also mentioned as an equipment problem.

> *Worker, Initial Interviews*
>
> "AC [air conditioning] in cabs—this is a problem. It has improved but not where it should be…. There is a big difference when one gets off a train with AC than one without it. Lots of the fleet are not equipped."

During the final interviews, workers were asked specifically about the work schedule and the facilities, and asked if there had been any changes. Fatigue control was not raised as an issue in the CAB barrier removal process, yet it appeared management had worked on this issue. Many workers and some managers reported *improvements in the predictability of work schedules*, which helped workers better manage fatigue. In particular, respondents reported improvements in the train lineup and in the use of relief crews (although some of the improvement is attributed to a transitory surplus of workers).

> *Worker, Final Interviews*
>
> "You have people, for instance calling hours in advance so that you can better plan your day, and getting better sleep. More courtesy calls regarding their schedule are made well ahead of time to better prepare you for work."

In contrast to the work schedule, the CAB Barrier Removal Team processed barriers related to the conditions of the facilities and equipment, and workers appear to have noticed the difference. Most workers and managers reported *improvements in facilities and equipment since the initial interviews*. Locomotive air conditioning, yard cleanup, and building repair in particular were mentioned.

Worker, Final Interviews

"I have seen a lot of improvements in the facilities, in providing employees with more stuff, such as locker rooms, more stuff in the engine such as AC in it. That has totally changed in the last three years. I've seen the facilities are now clean."

6.6.2.4 Labor–Management Relations

From the initial through midterm and final interviews, respondents reported improvements in the relations between management and labor. For workers, an important change was seeing a shift toward greater commitment to safety by management. Many workers and managers report that *management commitment to safety has improved since the baseline phase*, although some would have liked to see an even greater shift.

Worker, Initial Interviews

"Management has a great safety attitude if it benefits them. They push cars through even if they are not safe. Then they strong-arm employees with safety when they want to. All management at all levels is the same."

Worker, Final Interviews

"I see that 80 percent management have commitment but don't have the money. The lower management can only execute what upper management says. I really don't know where the money is."

Management's commitment toward safety implies rules are enforced consistently, and managers can be relied on for safety. Some workers and many managers reported *greater fairness by management and greater trust between workers and managers* since the initial interviews.

Worker, Initial Interviews

"Most people don't trust managers. There are maybe a handful of managers workers trust to talk to."

Worker, Final Interviews

"I have a lot more trust with my managers. I can go to all of them. I know that a lot of the older guys still don't trust managers. I know that my managers want me to work safe."

There were several comments specifically about FTX tests, which were targeted by the SLD component of CSA as well as by management's own initiatives, as being more fair and less disciplinary.

Worker, Final Interviews

"Yes, managers are now being very professional out there in talking with employees, and doing the test in areas where they can do it. I assume that the upper managers are talking to their staff and saying, 'We are here to help, not to fire somebody.'"

Instead of discipline, some workers viewed their managers as safety coaches.

Worker, Final Interviews

"I have even had managers take me along the route and point things out…. I was out for a bit and came back but the rules have changed. I glanced over them but didn't really learn them. I was hit on an FTX observation. It stings, but you know that you aren't perfect.

> The manager was good about it. He told me ways on how to fix them. I had heard of horror stories from the other workers, but my experience was good."

Changes in worker perceptions coincided with reciprocation by managers.

> *Manager, Final Interviews*
>
> "When I first started, the mentality between managers and trainmen is that they considered trainmen as oxen that have to be whipped to get them going…. Over the years management's thoughts have changed. The workers are smart people and have smart things to say, too…. There are things that they could teach me."
>
> *Manager, Final Interviews*
>
> "I used to hate to want to come to San Antonio. I heard bad stories. I had the opportunity to come up to San Antonio. Now many of the changes were already happening. But there has been a lot of change here. CAB is a start in improving trust."

With greater trust between workers and managers, some workers and managers reported *improved communication and cooperation between management and labor on safety*. Problems in communication and cooperation were rarely mentioned by either group during the final interviews in contrast with the initial interviews. Workers in particular reported more often during the final interviews than during the initial interviews that managers were listening to them and respecting them more in face-to-face interactions.

> *Worker, Initial Interviews*
>
> "I'll give you an example. A worker was told to pull 100-plus cars out under certain conditions that broke the train in two. It was a manager's idea to not cut away a smaller set. Could have avoided it but he wouldn't listen to the worker's idea."
>
> *Worker, Final Interviews*
>
> "Yes I have seen changes in management making safety a priority. I have experienced one incident myself, where managers approached me after letting them know about problems in getting a switch lined up. The four managers went there to investigate. They talked to us and said thank you for bringing it to [their] attention."

Such openness in turn was associated with more reports of workers initiating communication with managers.

> *Manager, Final Interviews*
>
> "It used to be that the employees wouldn't even talk to a manager without their Union Local Chair, but now I can deal with the employees directly. Before they wouldn't talk to us even on the smallest issue, but now the conversations are more open and frequent."

Overall, many workers and managers indicated that *peer-to-peer relations and relations between workers and managers have improved since the introduction of CAB*. Some workers saw change beginning as early as the time of the initial interviews, sometimes singling out specific relationships that were the exceptions.

> *Worker, Initial Interviews*
>
> "Managers had a bad relationship around here. The Superintendent will not talk to us that way…. Old management was aggressive."

By the final interviews, bad relationships were more the exception than the rule. Many respondents saw CAB assisting in this transformation.

> *Worker, Final Interviews*
>
> "You see more people talking to each other. When I get to work, everybody used to just type in their work on the computers in the building and not interact with one another. The CAB process has changed the atmosphere and tries to make people feel that they are friends of one another. More people talking to each other and become a family—a one group that is working together and interacting with one another. CAB is totally changing this."
>
> *Worker, Final Interviews*
>
> "There is something that might be partly attributed to CAB but not completely and that is that we now have a softball team comprised of labor and management. Mostly it's comprised of samplers from the labor side and then the superintendent and a couple of managers. The team sucks but that's okay because it's fun."
>
> *Manager, Final Interviews*
>
> "I think CAB has opened up the relationship between management and workers. It may be the one avenue that opened doors that hadn't been available before."

6.6.3 Personal Change

During the personal change interviews conducted in 2008, workers and managers attributed changes in their behavior to CAB. Specifically, those performing third party sampling saw changes in others, and some also saw change in themselves. The most informative stories were from these samplers working with their co-workers, either as formal or informal third party samplers. Some of these stories highlighted the value of having CAB participants conducting intervention with their colleagues. One engineer leveraged his relationships with co-workers to convince them to change their behavior.

> *Engineer: Personal Change Interviews*
>
> "Just the other day I had an old head, 30 years plus, and I was like, "hey man…you got off the van, and when you were standing next to the locomotive you started putting on your glasses, you started putting in your plugs." Cause, you know he went to the rear of the van to get his PPE but didn't put them on till he was standing by the locomotive…and I go "What were you thinking?" And he said, "Well I know I gotta do it." And I go, "Well you *gotta* do it man…. Do you have grandchildren?" And he goes, "You know I always talk about so-and-so I always brag about them." And I go, "Exactly! Isn't the best thing in the world you ever heard was 'I love you dad' or 'I love you grandpa' or anything like that?" He says, "Yea." I say, "Well you ain't gonna hear that if you don't wear those plugs. Don't you wanna see them grow up?" He says, "Yeah, of course I do." I say, "Well you ain't gonna see them if you don't wear those glasses. I'm not telling you to wear them. I'm not a *manager* or anything like that. I'm here to help *you* out. You know what I mean?" I go, "You can't be anti-railroad, okay, when you work, 'the hell with the PPE.' You gotta realize it's for your benefit. We are out here to make it safer for you." And he was like, "Okay, yeah, I see where you're coming from." So, do I see him now? Yeah, he even has those little things where his glasses are hanging around his neck and stuff like that. And he'll just wink at me like "I got em!" [all emphasis in original tone of voice]"

In this instance, a sampler's behavior reportedly had an impact on another worker. The result: the more experienced worker began to act more safely on the job by using personal protective equipment more often. Less overt is the sampler's apparent internalization of the CAB principles generalization to his everyday job. His handling of this conversation was not the activity expected in the feedback of a CAB sampling session. This conversation was the sampler's way of accomplishing the goals of CAB and reflects how internalization of CAB can influence activities outside the confines of traditional CAB feedback sessions.

Although most workers did not articulate when they started to change their behavior, many saw the CAB process as another in a long line of safety-related awareness programs that help "fine-tune" their safety sensibilities. Talking with other workers, attending safety films, and attending CAB class all serve to bring safety back to the front of their thinking and keep it on their radar.

Some people described actual behavior changes. An engineer talked about being more rule compliant. Another worker mentioned not answering his cell phone in the CRZ. Others described a general atmosphere of rule compliance as the norm. Many workers and managers talked of acting more safely off-duty.

Workers reported internalizing the lessons of CAB and generalizing them to areas not directly related to CAB implementation. For instance, a senior locomotive engineer said crews are safer in how they embark and disembark from the train—paying more attention to handing bags up and down the ladder. Others mentioned paying more attention to details such as carrying an extra microphone in their bag for the conductor for cases when trains are missing one. A confrontation between a younger conductor and a more senior engineer highlighted how safety is taking on a more overt role in how workers operate. Prior to CAB, one would be very reluctant to challenge a more experienced worker on safety.

> *Manager: Personal Change Interviews*
>
> "Well, I've had a couple of times where I don't think it would have happened before [CAB]…. But, I've had crews come in [to the terminal] and I mean the engineer and conductor are about to fight. The engineer's an old head and the conductor's a new guy that may be promoted. In one of the instances… they were arguing about the right way to do something. Where before it would have just been 'hey you know you need to do this' and the old head would have said 'I'm not doing it like that, this is the way I do it.' And, I mean it turned in to probably not the best scenario because they were, you know, they were about to 'come to blows'. But, it all ended up getting defused and the engineer eventually agreed that you know, maybe that wasn't 'the smartest thing for me to do' and I don't think that ever would have happened 3 years ago. I think 3 years ago they would have, either nothing would have been said to each other or they'd have went off the property and gotten in a fistfight. You know? And I attribute all of it to the CAB process because they had, the young guy had enough gumption to stand up to the old head and say this is the safe way to do it, regardless of whether or not that's the case and had enough gumption to stand his ground because he knew that he had backing whether it be with the CAB class or me or the other managers that's out here. He knew that the most important thing was to do it safe. Period. And I don't think 3 years ago that was the case."

Respondents to the personal change interviews in general reported more communication or sharing of information as an outcome of CAB. More communication among workers both in and out of the cab and a more open space for dialogue around safety prevailed. This was different than the atmosphere before CAB where “you were actually just across the cab but seemed like you were 2–3 miles apart.”

Detailed findings for the personal change interviews are provided in Appendix A: Changing At-Risk Behavior (CAB) SASU: Individual Stories of Change, starting on page 162.

7. Findings and Recommendations

7.1 Summary of Findings

With CAB being a pilot demonstration of CSA, the evaluation team assessed specific activities, outcomes, and impacts (refer to the logic model in Figure 7 on page 38) of CAB to answer evaluation questions about CSA in general. These questions concerned:

- Implementation.
- Outcomes.
- Sustainability.
- Transformation.

In addition to the findings related to each of these evaluation questions, there were also serendipitous findings regarding the *context* of a site and organization and their readiness to accept a CSA implementation.

Table 16 lists the evaluation questions and summaries of the answers implied by the analysis of CAB. Detailed findings for each of these evaluation questions of CSA along with the supporting results from the analyses of CAB are summarized in the following subsections 2 through 5. Note that, where findings are inconsistent with the predictions in the logic model in Figure 7 on page 38, it generally appears to be due to a lack of adequate data being available during the summative evaluation rather than to outright contradictions of predictions.

Table 16. Evaluation Questions and Answers

Evaluation	Question	Answer
Context	What are the characteristics of the initial context of a site (e.g., attitudes, labor relations, etc.) that prepares it to accept a CSA implementation?	A site is likely to accept CSA if the context makes it open to nontraditional approaches and leadership consistently promotes labor–management cooperation.
Implementation	What characteristics of a CSA implementation allow its activities to be completed as planned in the railroad industry? How should the CSA process be improved to more effectively establish it in railroad settings?	CSA can be successfully implemented within railroad industry transportation departments. A successful implementation depends on overcoming worker resistance, obtaining management commitment, and acquiring a sufficiently high contact rate. The implementation in this demonstration was sufficiently strong to justify an evaluation for outcomes.
Outcomes	What are the effects of the CSA process on safety and safety culture?	CSA has demonstrated improvements for transportation departments in safety practices, conditions, culture, and occurrences such as incidents and engineer decertifications.
Sustainability	What characteristics of the implementation and organization (e.g., economic and political) are necessary for CSA to be sustained once it has been successfully implemented?	CSA implementation is sustainable in railroad transportation departments through enlisting the support of workers, unions, management, and other organizations and providing benefits at an acceptable cost.
Transformation	How does the CSA process influence the broader organizational environment of the pilot demonstration, if at all? Specifically, does the process make the stakeholder organizations more likely to adopt proactive, nondisciplinary approaches to safety, thereby encouraging the spread of processes similar to CSA to other parts of the organization or industry?	If senior management values CSA enough to decide to spread it to other locations, CSA has the potential to transform the transportation departments across service units in a carrier to use effective, proactive, nondisciplinary approaches to safety.

7.1.1 Context

7.1.1.1 Cultural Unfreezing and Openness

Finding: An openness to suspending traditional organizational patterns in favor of a modicum of labor–management cooperation facilitates the establishment of CSA and cultural change in general.

Table 17 summarizes the results describing the context before the implementation of CSA that made SASU inclined to try it.

Table 17. Summary of Results on Cultural Unfreezing and Openness.

Method	Result
Field notes	A string of serious accidents at SASU produced the perception that the current organizational culture was not adequate for promoting safety. Contrary to traditional union–management relations, local union officials and top management chose to cooperatively address safety issues at SASU after the accidents. This included initiating interventions before CAB. Many new workers who were not indoctrinated into the traditional railroad culture and those who were not prepared to accept this culture joined the service unit.

Field notes and interview data indicate a context of cultural disequilibrium before the establishment of CAB at SASU, with some workers and managers deviating from their usual patterns of interaction in the culture. The series of fatal accidents at Macdona and elsewhere led the unions and management to question the cultural assumptions about the safety of their organization. This questioning apparently led unions and management to break away from their traditional adversarial relations and to make attempts to cooperate. The field notes show that before CAB started, managers and workers cooperated on the joint Crew Resource Task Team and Safety Command Center. NTSB noticed this unusual level of cooperation and reported their observations to FRA.

The logic model of Figure 5 on page 34 shows cultural outcomes as distal outcomes occurring well after implementation. By nature, cultural change is a slow process that takes years. Even though unions and management demonstrated unusual cooperation and trust *before* CAB implementation, evident to outsiders such as the NTSB, this behavior was apparently limited to immediate safety concerns resulting from fatal accidents. It seems unlikely that instances of cooperation before CAB implementation represent a consolidated cultural change—that management and workers had fundamentally changed the way they interact. Instead, it more likely represents the *start* of cultural change, manifested by an *openness* to new approaches and new relations.

Kurt Lewin (1947, cited in Schein, 1992) characterized the start of change as an "unfreezing" of the current culture in which an organization experiences disequilibrium. The driving force behind the unfreezing is a perception that the old culture is inadequate in some areas, leading individuals to try alternatives. Such unfreezing, according to Schein (1992), is followed by cultural and psychological restructuring, where alternatives are tried and selectively accepted or discarded. Eventually, the culture "refreezes." The culture is modified and equilibrium regained by the institutionalization of acceptable alternatives that appear to address the inadequacy.

In the case of SASU, the perception of inadequacy resulted from the series of accidents. However, the demographic shift in the workforce was also a factor. The service unit experienced an influx of new, younger workers who did not have the same traditional expectations about labor–management relations. These two factors may have encouraged

SASU's preimplementation venture into labor–management cooperation. These ventures, such as establishing the Safety Command Center and Crew Resource Task Team, in turn triggered a series of events that helped to implement CAB. They also provided the kernel of labor–management cooperation that a CSA process needs to get started at a site. At SASU, this cooperation was noticed by NTSB and FRA and expanded as CAB was implemented, changing safety and the culture (see Outcomes, page 123).

Given that SASU experienced cultural unfreezing followed by CSA-induced expansion of labor–management cooperation, it appears that cultural change is not necessarily a distal outcome of CSA, as depicted in Figure 5 on page 34, nor is it a prerequisite for CSA that occurs solely before implementation. Rather, a successful CSA process is embedded in the course of cultural change. The change starts with the unfreezing of the culture before implementation and an openness to alternatives, specifically cooperation between labor and management. Ultimately, this leads to a refreezing of a modified culture under the influence of the CSA process. Figure 26 on page 117 is a revised logic model that illustrates these effects. In this case, there is an openness to cooperation at these facilities before the CSA implementation. Broad levels of labor–management trust and cooperation, measured by such survey scales as the MLR, become quantitatively apparent later as distal outcomes. They are visible as the culture refreezes into a new form.

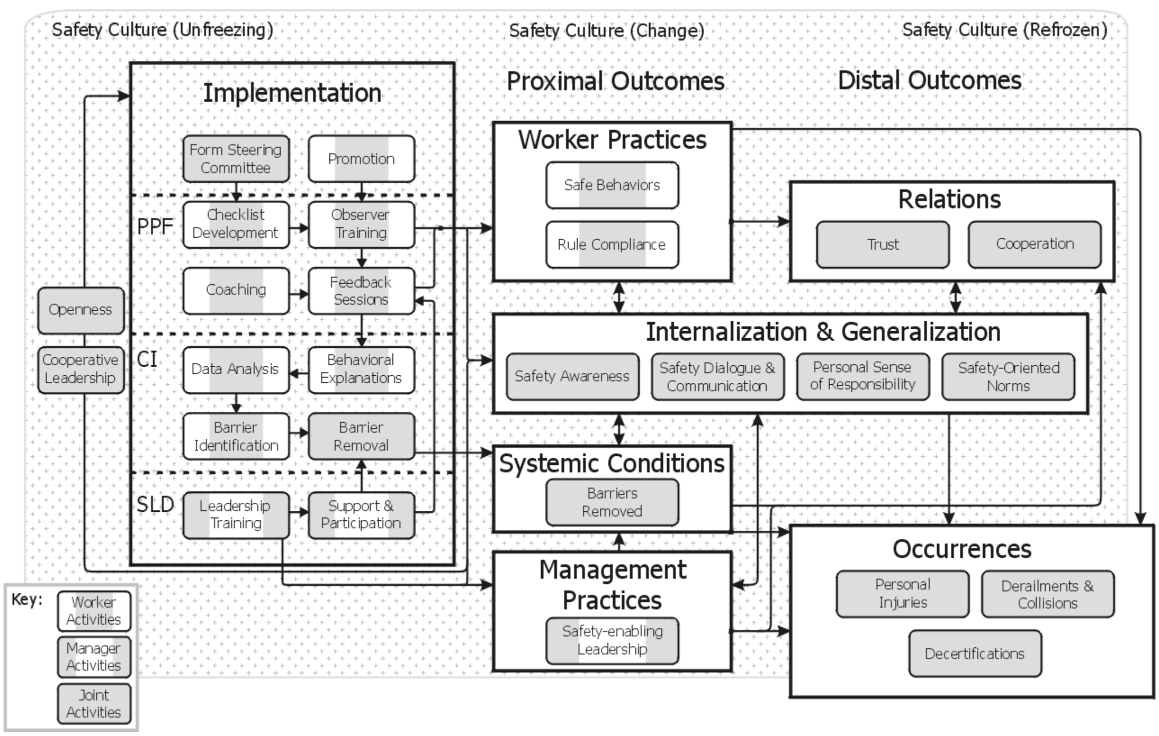


Figure 26. Modified Logic Model with Inclusion of Cultural Disruption and Direct Influences of CSA Training on Internalization and Generalization

7.1.1.2 Leadership and Support in Preparing for CSA

Finding: Local leadership and external support of CSA facilitates its establishment.

> Not ideas, but material interests, directly govern men's conduct. Yet very frequently the "world images" that have been created by "ideas" have, like switchmen, determined the tracks along which action has been pushed by the dynamic of interest.
>
> –Max Weber (1991)

Table 18 summarizes the results regarding a context of cultural leadership that supported CSA at SASU.

Table 18. Summary of Results Demonstrating Leadership and Support in Preparing for CSA

Method	Result
Field notes	Local union officials and top management cooperate to address safety issues at SASU before CAB. FRA approaches UP to conduct a CSA demonstration pilot at SASU. UP corporate commits to supporting CSA for at least 2 years. Local union officials and top management at SASU focus their cooperation on implementing CAB. BST provides instruction and advice on establishing all components of CSA at SASU.

Field notes indicate that serious road and switching accidents, such as Macdona, substantially promoted Weber's "material interest" to improve safety at SASU. However, even though workers and managers may have had a common goal of improving safety, this does not guarantee cooperation between labor and management. In the culture of the railroad industry, workers tend to adopt a "technology" orientation (Hale, 2000) toward safety, attributing accidents to facilities and equipment. In contrast, managers tend to adopt an opposing "procedural" orientation, attributing accidents to inadequate rules or to worker failure to follow rules. When opposing safety orientations are combined with the emotionally charged atmosphere generated from a fatal accident, it is entirely possible that each side will lash out against the other for causing the accident. In the traditional railroad culture, accidents could worsen relations between labor and management, making each more entrenched in their respective opposing orientations of safety.

Discontent with the SASU approach to safety appeared to combine with the attitude of a younger workforce in unfreezing the culture. An unfrozen culture can be mercurial, with individuals pursuing many possible, perhaps contradictory, countercultural actions. With an unfrozen culture implying uncertainty, individuals need direction toward a meaningful alternative, or eventually, the gravity field of the traditional culture pulls behaviors back to where they were.

The accidents at SASU occurred at the same time as growing numbers in new labor and new leadership in management. The approach taken by these leaders to improving safety was cooperation, resulting in safety initiatives such as the Safety Command Center and the Crew Resource Task Team. The latter became the nucleus of the Design Team for CAB, implying that improvements in labor–management cooperation may have helped start the CAB process. CAB appeared to start at a fortuitous time, when labor and management leadership was more disposed to cooperate and when the rank-and-file workers and lower managers tended to support countercultural actions by this leadership because of the recognized need to improve safety.

Leadership and labor support can focus the countercultural tendencies into a consistent process that ultimately becomes part of the new culture. At SASU, top local management and union officials were united in their support for cooperation (as shown by the Safety Command Center and Crew Resource Task Team) and for CSA in particular. The long-term focus on CSA over other means of cooperation appears to be the result of support from outside the service unit. Promising results from another CSA pilot site encouraged FRA to support a demonstration pilot at SASU. With FRA backing, UP corporate management committed its support to the CSA pilot. BST, through FRA sponsorship, provided technical expertise to make an effective and sustainable implementation. This included training in implementing PPF, CI, and SLD. The training contributed to the cultural change by offering new safety activities for workers and managers at a time when both were seeking ways to improve safety. In the event of a cultural unfreezing, positive cultural change is facilitated by a consistent and compelling support shared by all leaders.

7.1.2 Implementation

The implementation evaluation question relates to what happens *if* CAB were implemented as planned in addition to which characteristics allow CSA to be implemented as planned. There are four findings related to CSA implementation. The first finding concerns implementing the process as planned, and three remaining findings described characteristics associated with a complete implementation: worker resistance, management commitment, and contact rate for a dispersed work setting.

Implementing CAB as planned is important for two reasons:

- It demonstrates that CSA can be successfully implemented in railroad transportation departments despite the railroad culture and other challenges, such as a dispersed work setting (see Evaluation Scope, on page 28), that suggest that CSA may not be possible for railroad transportation.
- It justifies the evaluation for CAB outcomes. If CAB were implemented incompletely, then improvements in practices, occurrences, and culture could not be plausibly attributed to CAB.

All components of CSA are included in evaluating the CAB implementation. This includes the PPF activities completed by workers, the SLD activities completed by management, and the CI activities completed by workers and managers jointly.

7.1.2.1 Completion of Implementation Activities as Planned

Finding: Implementing CSA as planned is possible within transportation departments of the railroad industry. Implementation of the pilot demonstration was not entirely complete, but it was sufficient to justify an evaluation of outcomes.

Table 19 summarizes the results concerning completing the implementation activities.

Table 19. Summary of Results for Implementation Completeness

Method	Result
Process metrics	Training of samplers was completed at or ahead of schedule. The rate of feedback sessions ("contact rate") for CAB–Switching was near the targeted rate. The contact rate for CAB–CRZ was low relative to the target rate but adequate to allow all workers on average to be observed repeatedly.
Field notes	The steering committee created feedback session checklists for both CAB–CRZ and CAB–Switching. Stakeholders promoted CAB at the San Antonio terminals and at the outlying terminals of the service unit. Steering committee members coached samplers regularly. The steering committee analyzed observation data regularly and identified barriers. The steering committee removed barriers regularly by directly contacting managers, participating in the KEYS Safety Committee, and working with the Barrier Removal Team. Managers completed SLD training.
Interviews	Managers reportedly supported CAB efforts including removing barriers and conducting feedback sessions.

All CAB implementation activities were completed to a certain extent (see Figure 7, page 38, left side for Implementation box). Field notes showed that the workers developed, deployed, and used checklists for both CRZ and switching operations. CAB process metrics indicated that over half the workforce was trained to conduct observations at the end of the evaluation period, meaning that training was completed at or above the planned rate. The process metrics also showed that the contact rate for CAB–CRZ at the end of the evaluation period was steady at approximately 20 percent. Although this was below the target, it was frequent enough so that each worker was likely to receive feedback multiple times during the evaluation period. The contact rate for CAB–Switching was consistently 50 percent or more and approached the Design Team's target of 100 percent. However, this was only observed in the last 6 months of the evaluation period, thus diminishing the likelihood of detectable outcomes from CAB–Switching.

Field notes show that, as instructed by the BST consultant, samplers regularly collected the reasons for at-risk behaviors and that workers reported at-risk conditions during feedback sessions. The steering committee routinely analyzed these data for barriers to safety. Workers and managers regularly removed these barriers by, for example, the steering committee conducting educational interventions and directly notifying appropriate managers and submitting the barrier to the KEYS Safety Committee and the CAB Barrier Removal Team. Management completed safety leadership training on schedule and supported feedback sessions and the barrier-removal process.

In summary, the demonstration pilot showed that CSA can be implemented as planned within a railroad transportation department. CAB implementation occurred despite the challenges from the culture of command and control management, reactive operations, personnel with opposing and outdated orientation on safety, adversarial labor–management relations, and rotating management (refer to Context: The Problem of Railroad Safety and Culture on page 8). CSA was successfully implemented in both dispersed road settings and semidispersed yard settings. Stakeholders managed the competition between the two unions involved in the process and developed effective behavioral checklists for transportation operations, including for CRZ conditions with difficult-to-observe attention and awareness practices.

Because CAB implementation was sufficient, there is a reasonable chance for detectable outcomes if CAB is effective. As a result, it is appropriate to evaluate CAB for outcomes and impacts (Rossi et al., 1999). However, the implementation was not ideal for evaluating outcomes. With a strong but not entirely complete implementation, one may expect detectable outcomes and impacts in some but not necessarily all areas, assuming that CSA is potentially effective in the railroad industry.

7.1.2.2 Overcoming Worker Resistance

Finding: Worker education and management respect for the process overcomes worker resistance to the CSA process.

Table 20 summarizes results for overcoming worker resistance to CAB.

Table 20. Summary of Results for Overcoming Worker Resistance to CAB

Method	Result
Field notes and interviews	CAB training educated workers on the CAB process and overcame resistance by allaying concerns about management using it for discipline.

Field notes and interviews indicated that worker resistance was primarily over concerns that CAB observations of at-risk behaviors might be used against workers for disciplinary purposes. This resistance decreased after workers learned that their peers controlled the CAB process and that the data did not identify the observed workers. Given the tradition of adversarial labor–management relations, these attributes of CAB made it attractive to workers. Workers learned about CAB through the CAB newsletter, the CAB Web site, face-to-face dialogue with steering committee members and samplers, and most importantly, training for samplers. Management did not intrude on or interfere with feedback sessions or data storage and analyses, assuring workers of the integrity of the process and reinforcing their impressions about CAB.

Workers who were part of the CAB process, from the facilitators down, believed in the value of CAB, and there were no signs of disenchantment. During the interviews across the evaluation period, steering committee members stated that the primary problem with CAB was not having enough CAB—that they needed more feedback sessions, more presence in outlying locations, or more inclusion of yard work.

7.1.2.3 Achieving Management Commitment

Finding: Management commitment of resources to the CSA process improves in response to a greater awareness of process needs. Management commitment to CSA also increases as the process became more visible to regional management.

Management commitment includes providing adequate resources for CAB and showing workers visible management support of CAB, the latter being important for worker acceptance given the history of stalled safety programs. Table 21 summarizes results.

Table 21. Summary of Results for Obtaining Management Commitment to CAB

Method	Result
Field notes and interviews	Management initially did not anticipate the scale of resources necessary for CAB. Management responded by committing resources as the needs of CAB to expand to entire service unit became apparent.
Field notes	Management commitment was associated with the visibility of the CAB process. Managers who attended Oversight Committee telcons or visited with CAB facilitators and trainers expressed commitment to CAB. Support from regional and corporate managers was important for obtaining the commitment of local managers.

Field notes and interviews indicated that management did not interfere with the workers' PPF activities, allowing them to conduct feedback sessions and maintain ownership of the data. Following directives from top management, front-line supervisors allowed workers to complete their activities independently. This illustrated management's commitment to CAB and the concept of a worker-driven safety process. Field notes and interviews also indicated that initially management was not fully prepared to provide the resources (time and space) necessary for a CSA process. This resulted in resource shortfalls, such as the lack of computer support and physical space for CAB training and operations. In interviews, workers interpreted such shortfalls as a lack of commitment.

Management commitment improved during the evaluation period, culminating in the spring 2007 increase in the CAB budget. The increase allowed CAB to expand from several locations to the entire service unit providing a comprehensive safety program for both labor and management to promote and manage.

Commitment at the corporate through regional and local levels seemed related to maintaining CAB visibility with management through such means as opportunities for managers to witness sampler training and participate in Oversight Committee telcons. Inclusion of managers above the local level seemed important, perhaps giving local management implicit authorization to commit resources to CAB. At least one manager at the regional or corporate level attended 75 percent of the telcons. Regional managers also visited the facilitators at SASU and expressed favorable attitudes toward CAB. The steering committee encouraged managers at all levels to attend CAB training. Combined with regional management support of CAB, this appeared to maintain local management commitment through manager rotations.

Demonstrations of the value of CAB also appeared to encourage management commitment. Within a year of starting CAB, BST and the evaluation team reported signs of positive outcomes to managers (e.g., decreases in FTX failure rates). Reduced administrative costs in response to the budget challenge in spring 2007 showed managers that CAB can control costs. Worker perception of management commitment also improved in view of the budget increase and management interest in CAB (e.g., attending telcons and trainings).

7.1.2.4 Contact Rate for the Road

Finding: Conducting feedback sessions for a dispersed workforce remains a challenge, but worker motivation to make safety contributions can be used to multiply contacts.

As expected, the dispersed nature of the work setting made it difficult to achieve a high contact rate (feedback sessions per month). Table 22 summarizes CAB contact rate results.

Table 22. CAB Contact-Rate Results

Method	Result
Field notes	Samplers initially did not conduct a sufficient number of cross-crewmember feedback sessions despite encouragement from the steering committee. Samplers volunteered regularly to conduct third-party feedback sessions, indicating that workers valued participation in a safety process in addition to liking the at home 8-hour workday associated with the activity. The steering committee eventually required that samplers first conduct cross-crewmember sessions in order to be allowed to conduct third-party sessions. This was effective in increasing the contact rate while not sacrificing quality.

Achieving a high contact rate with limited manpower depended on using cross-crewmember feedback sessions, but in general, samplers were reluctant to conduct these sessions. In response, the steering committee used the opportunity to conduct third-party observations to motivate samplers to conduct cross-crewmember observations. The use of extrinsic motivators to increase feedback sessions is risky because samplers may sacrifice quality to achieve quantity goals. However, in the case of CAB's use of third-party observations to motivate cross-crewmember observations, contact rate greatly improved without any apparent reduction in the quality of the observations or feedback. One of the main reasons that workers reportedly enjoy third-party sampling is that it allows them to directly contribute to improving safety. It seems reasonable that workers who want to contribute to safety are likely to value high-quality feedback sessions. This implies that they will pursue high quality in their cross-crewmember sampling. It is an ingenious solution. However, although it doubled the contact rate from approximately 10 to 20 percent, the design team sought a contact rate of 100 percent. Achieving such a rate remains a challenge that will face any CSA implementation for dispersed (e.g., road) workers.

7.1.3 Outcomes

There were six potential outcomes: worker practices, manager practices, systematic conditions, occurrences, safety culture—relations, and safety culture—internalization and generalization (see Figure 7, page 38).

7.1.3.1 Worker Practices

Finding: Worker practices targeted by the CSA process improve with the introduction of the process.

Data from process metrics, the survey, the interviews, and corporate safety data all showed evidence of improvements in worker practices targeted by CAB. Table 23 summarizes these results.

Table 23. Summary of Results for Worker Practices

Method	Result
Process metrics	The percentage of at-risk behaviors observed by workers decreased over time for both CAB–CRZ and CAB–Switching after the start of feedback sessions.
Survey	Workers and managers reported more safe CRZ practices in 2008 than in 2005. In 2008, workers and managers reported no significant difference in switching practices between those done currently and those they recalled doing before the start of CAB.
Interviews	Workers and managers reported improvements in both CRZ and switching practices over the evaluation period.
Corporate safety data	Failure rates of FTX tests for CRZ rules improved from baseline to the first phase more than rules not related to CAB PPF checklists. This corresponds to the start of CAB–CRZ. Failure rates of FTX tests for switching-related rules also improved from baseline to the first phase more than rules not related to CAB checklists. This was before the start of CAB–Switching. Failure rates for all FTX tests worsened from the first to the second phase, but failure rates for tests related to switching worsened less than failure rates for tests unrelated to CAB checklists. This corresponds to the start of CAB–Switching.

The process metrics showed a decrease across the evaluation period in the percentage of at-risk behaviors reported by samplers for both CAB–CRZ and CAB–Switching. Samplers were checked for consistency in their judgments over time by analyzing their ratings of training videos of safe and at-risk behaviors. This consistency check indicated that the measured decrease in the percent at-risk of actual behaviors cannot be attributed to observers changing their criteria for at-risk and safe behaviors. To the contrary, the negative drift detected between initial training and coaching of samplers suggests that the decrease in at-risk behaviors is underestimated by the process metrics.

In surveys, workers and their managers reported more safe CRZ practices in 2008 than in 2005. There were no differences in current and retrospective switching practices in 2008. One explanation is that, because retrospective data rely on workers accurately recalling their practices from years ago and because retrospective questions immediately followed the current-practices questions on the survey, workers may have felt encouraged to mark the same values for each question. Results may have been different if the switching questions were asked in 2005 and 2008 for contemporary practices rather than having relied on retrospection.

The scores for switching practices were also, on average, close to the mathematically maximum value, suggesting a ceiling effect. To detect changes in practices, the scale may need different items with wider distributions of ratings. Given reliance on retrospection and the apparent ceiling effect, the results from the switching scale appear to be inconclusive, indicating neither a change in practices nor a lack of change.

Interviews backed up the survey results for CRZ, with progressively better CRZ practices reported by workers and managers from initial (2005–2006) through midterm (2006–2007) and final (2007–2008) interviews. The interviews likewise indicated improvements in switching practices, in contrast to the survey results for switching.

For corporate safety data, FTX test failure rates for SASU improved more for CRZ practices than for practices not associated with CAB checklists from baseline to the first phase (2005–2006). This suggests that something related specifically to CRZ practices at SASU, such as the CAB–CRZ process, improved these practices more than others.

FTX failure rates for switching practices also improved more compared with failure rates for practices not associated with CAB checklists from baseline to the first phase, even though this was before the start of CAB–Switching. It is not clear what caused this improvement. There is qualitative evidence for internalization and generalization of safety that may improve practices beyond those related to CRZ (see the Safety Culture – Internalization and Generalization subsection below). However, internalization and generalization should improve practices not associated with CAB as much as switching practices. There is also qualitative evidence that managers have improved the FTX process by providing workers with more constructive feedback that may have resulted in improvements in practices. This, too, should improve practices not targeted by CAB–Switching as much as practices targeted by CAB–Switching. It is possible that the facility improvements made by KEYS and other safety processes in parallel with CAB have made it easier for workers to comply with rules and have resulted in reduced FTX failures for switching. However, because KEYS did not exist until late in the first phase, it is unlikely that it made a large contribution to switching-practice improvements this early in CAB. One possibility unique to switching is that the fatalities at East Yard and Crystal Cold Storage (NTSB, 2004, 2005b) raised safety awareness in switching through the first phase.

From the first phase (2005–2006) through the second phase (2006–2007), FTX failures increased for all tests. However, FTX failures increased less for practices targeted by CAB–Switching than for practices not targeted by CAB. The relative change for CAB–CRZ practices was ambiguous. The performance of switching-related tests relative to other tests suggests that something specific to switching practices counteracted whatever was responsible for the general worsening in FTX performance. Such a counteraction cannot be attributed to cultural changes or management improvements to the FTX process, because these would affect tests for practices not targeted by CAB as much as tests for practices targeted by CAB. The counteraction is most plausibly attributed to CAB–Switching or the KEYS safety committee or both. Both CAB–Switching and KEYS were operating in the second phase, and both were specifically concerned with yard work.

The general worsening of FTX performance between the first and second phases is difficult to explain. With qualitative evidence for internalization and generalization, one would expect FTX for all rules to continue to improve through 2007 as CAB strengthens. Inspection of a smoothed graph of FTX failures shows a sharp increase in failures around December 2006 for all of the original service units of the Southern Region, suggesting a regional or corporate-wide cause. This timeframe corresponds approximately to a change in corporate discipline policy mandating harsher penalties for selected rules violations, but it is unclear how this is connected to FTX failure rates. Other service units of the Southern Region recovered from the increase in failures seen in winter 2006 but not SASU.

Presumably, there are SASU-specific reasons to explain this lack of recovery of FTX, perhaps something about SASU's FTX process itself. The worsening in practices indicated by FTX failures is not replicated by other measures used in this evaluation. The process metrics show only continuous improvement, and no worsening was reported by managers or workers in the final interviews relative to the midterm interviews. Given these results and the lack of any psychometric checks for FTX's soundness as a measure of practices (unlike the process metrics), the apparent worsening of the FTX failure rates is more likely an artifact of FTX as a measure than an actual worsening in safety practices. The final conclusion is that, as predicted in the logic model, the practices targeted by CAB–CRZ and CAB–Switching improved in response to CAB.

7.1.3.2 Manager Practices

Finding: In a successful CSA implementation, management promotes the CSA safety process, removes barriers to safety, and improves safety leadership in other areas, such as in coaching of workers in field operations tests.

Table 24 summarizes results for manager practices.

Table 24. Summary of Results for Manager Practices

Method	Result
Field notes	Managers removed barriers to safety, both those identified through the CAB process and those identified by other means. Managers supported the CAB process and complimentary safety efforts such as KEYS.
Interviews	Managers increasingly promoted the CAB process. Managers improved the FTX process by providing more coaching to workers in their interactions.

Interviews indicated greater management commitment to and promotion of CAB over time. By the end of the evaluation period, promoting CAB was the primary method of involving front-line and middle managers in CAB. Workers also reported improvements in manager coaching during FTX tests through the evaluation period. By the final interviews, some newer workers wondered aloud why others viewed FTX tests so negatively. In light of the SASU Safety Leadership training targeting CAB support and

FTX coaching, and with managers reportedly finding the training useful, it appears reasonable to conclude that manager practices have improved in response to this training.

Field notes also indicate that management is fulfilling its role in removing safety barriers that greatly affect many workers, as described in the following subsection. They also instituted with labor's cooperation the KEYS safety committee, which streamlined the process of removing safety barriers, whether they were identified through CAB or some other means.

7.1.3.3 Systemic Conditions

Finding: Systemic conditions improve as management removed barriers to safety with workers contributing.

Table 25 summarizes results for systemic conditions.

Table 25. Summary of Results for Systemic Conditions

Method	Result
Field notes	Work conditions at the Eagle Pass yard improved extensively after the steering committee alerted management of issues there. In response to information supplied by the steering committee, corporate policy was changed to ensure that lead locomotives were equipped with air conditioning. Several yard issues were reportedly resolved through the KEYS safety committee. Some but not all of these issues were discovered during CAB feedback sessions.
Interviews	Workers reported improvements in conditions of yards and other facilities. Workers reported improvements in the predictability of train lineups so that crews can more effectively plan their rest.

Field notes verify that managers and workers improved systemic conditions. Some of these improvements were the direct results of the barrier-removal process of CAB such as management's ensuring that head-end locomotives have air conditioning and making various facility improvements, including a substantial investment at Eagle Pass. These improvements also included CRZ-related barriers the CAB steering committee chose to remove themselves through education of the workers.

However, some barrier removals by management were more general and extended beyond the activities of the CAB Barrier Removal Committee, suggesting that management is actively cooperating with labor and doing its part to improve safety. The KEYS safety committee became a vital means for labor and management to cooperatively improve facilities. At a corporate level, enhancements to the procedure for updating train lineups increased the predictability of rest time for road crews, a goal traceable back to the SASU Crew Resource Task Team before the start of CAB (see Accidents Motivate Change on page 61).

An exact count of the barriers removed was not available, but such a count would be misleading in any case because of the size range of the corrective actions. Some of these actions were small, such as peer education by steering committee members and local yard repairs handled through direct contact with the managers responsible. Larger actions included local yard repairs handled through KEYS, whereas major actions included Eagle

Pass yard work and the change to corporate policy to provide air-conditioned locomotives on the head-end of trains. Worker interviews corroborate the field notes, with workers reporting noticeable improvements in work conditions, both those associated with the CAB barrier removal process and those not associated with CAB.

In conclusion, even though there are no available quantitative values for systemic improvements, there have been improvements at SASU relating to both the CAB barrier-removal process and safety leadership training, along with management's own initiatives.

7.1.3.4 Occurrences

Finding: Occurrence rates decrease in the presence of CSA. A decrease in the rate of engineer decertification is associated with a CSA process for road practices. A decrease in human factors incident rates is associated with strong implementation of CSA for yard practices.

The results indicate improvements in occurrence rates specifically where CAB is theoretically most likely to improve occurrences. Table 26 summarizes results for occurrences.

Table 26. Summary of Results for Occurrences

Occurrence	Related CSA Process	Results
Decertifications	CAB–CRZ	After the start of CAB–CRZ, SASU experienced gradual improvement for decertifications associated with CRZ practices while other service units did not. There was no improvement at SASU for decertifications not related to CAB–CRZ.
Incidents	CAB–Switching	Since fall 2006, the rate of incidents for SASU did not decrease significantly, but did decrease more than the comparison sites combined.
		Within SASU, yards with moderate and strong CAB–Switching implementation had improved incident rates while yards weak CAB–Switching implementation had not changed.
Injuries	CAB–Switching	Injury rates at SASU and the other service units of the Southern region gradually improved 79% throughout the evaluation period, with no apparent change after the CAB-Switching started in SASU.

The evidence for improvements was most unambiguous for CAB–CRZ, which was subject to a full 2-year evaluation period. The evidence for CAB–Switching was more ambiguous, but it was evaluated only 1 year after it started.

From the beginning of CAB–CRZ until the end of the evaluation period, there was a gradual 79-percent decline in the overall incidence rate of engineer decertifications for failure to stop for red-signal aspects, speeding, and violating main-track authority. CAB–CRZ promoted practices specifically designed to prevent these kinds of decertifications. The decline in the decertification rate appears to be associated with CAB–CRZ activities rather than being a continuation of a preceding trend. Other service units of the original Southern Region, which had no CSA process for CRZ practices, did not experience such a decline, implying that the cause was unique to SASU, the one service unit with a CAB–CRZ process. Furthermore, there was no decreased rate at SASU in decertifications for drugs and alcohol, alerter tampering, or brake testing—decertifications unrelated to the practices promoted by CAB–CRZ.

In conclusion, only decertifications associated with CAB–CRZ declined, and only at SASU, the one service unit with a CAB–CRZ process. The decline coincided with CAB–CRZ activities at SASU. Consistent with predictions, a strong CSA implementation appears to have improved occurrence rates associated with the behaviors targeted by the process.

The overall human factors incident rate at SASU did not decrease significantly after the start of CAB–Switching. At the same time, the human factor incident rate did not increase as was observed at two out of the three comparisons service units suggesting SASU may have fared better than these two. As a result, SASU rate of incident decreased significantly more than other service units of the original Southern Region combined. While this is consistent with CSA improving safety, the differences among the comparisons sites casts doubt on whether they adequately represent a single comparison group. In particular, the difference between Fort Worth and Livonia suggests that different factors are at work among the comparisons sites, and thus the basis of comparison becomes unclear.

In general, comparison among the service units on yard safety is problematic due to confounds that imply the non-CSA service units were not appropriate comparisons for SASU. Two of them, Livonia and Houston, had initiated their own CSA-like processes that coincided with CAB–Switching and drew from the experiences at SASU. The third service unit, at Fort Worth, had no CSA process at the time but apparently encountered a major turnover at all levels of management. The new managers appeared to approach safety proactively rather than reactively, as is the traditional approach of the railroad industry. For example, in November 2007, Fort Worth management aggressively pursued the initiation of its own CSA-like process in the form of UP's TSC for the transportation department. This new management team and its approach to safety may account for declines in incidents at the service unit. The role of CSA in the decline in incidents is therefore ambiguous when comparing service units.

Although comparisons between SASU and other Southern Region service units revealed somewhat ambiguous improvements related to CAB–Switching, there were clearer improvements related to CAB–Switching *within* SASU. The rate of human factors incidents decreased 81 percent at Eagle Pass, the yard with the strongest CAB–Switching implementation. Meanwhile, the incidence rate decreased 32 percent at the San Antonio Complex, which had a moderate implementation. The change in rate at Eagle Pass, with its strong implementation, was significantly more than the San Antonio Complex. Rates did not significantly change at the Laredo yards, which had weak CAB–Switching implementations at the time of the evaluation. Thus, the differences in rate changes among the yards matched the strength of implementation, This is consistent with CSA impacting the incident rates.

That being said, the improvement at the San Antonio Complex was modest and not significantly different from the change at Laredo. It is not surprising that there was little change in the incident rate at the San Antonio Complex. The evaluation period included only the first year of CAB–Switching, whereas at least 2 years are generally necessary for detecting significant changes in occurrences. In light of a slow CAB–Switching implementation with a low contact rate at the San Antonio Complex for the first several months (see CAB Process Metrics, page 80), one could argue that CAB–Switching had only half a year to affect incident rates at the San Antonio Complex.

The results for incidents are consistent with results from an analysis of the CSA process within the Livonia service unit, where one yard had a CSA implementation for switching operations with a training rate and contact rate comparable with that of CAB–Switching for Eagle Pass. The Livonia yard experienced the same decline in human factors incidents as Eagle Pass, whereas comparison yards at Livonia without CSA implementation had no change in incident rates (Ranney, Zuschlag, Coplen, and Wu, 2009).

Contrary to predictions, there were no improvements in injury rates at SASU relative to the comparison sites. All service units in the original Southern Region, including SASU, had declining injury rates through Baseline and the First Phase. This suggests that injury improvements at SASU and other service units are due to factors existing before CAB, possibly a region-wide factor attributable to regional safety leadership. The other service units are therefore inappropriate for comparison with SASU (Rossi et al., 1999).

Like human factors incidents, injuries were expected to be associated with CAB–Switching. However, unlike the case for incidents, injuries specifically associated with switching could not be isolated in the data. Non-switching transportation injuries may have diluted any effects on the switching injuries. Furthermore, the evaluation period covered only the first year of CAB–Switching. CAB may not have had enough time to improve injury rates at SASU. Only Eagle Pass had a strong CAB–Switching implementation. The larger San Antonio Complex had only a moderate CAB–Switching implementation. The remaining large yards at Laredo were just beginning their CAB–Switching implementation at the end of the evaluation period. Given the moderate strength of CAB–Switching across SASU (on average), the period of evaluation could be too short to detect a change in injury rates for the entire service unit against a background of generally improving rates for the entire Southern Region.

Rather than analyzing changes in injury rates for all of SASU, the effects of CAB–Switching could be determined by comparing SASU yards with different strengths of CAB–Switching, specifically, by comparing injury levels for Eagle Pass, the San Antonio Complex, and Laredo yards. This was the approach used to examine the effects of CAB–Switching on human factors incidents. However, the worker-hours for each yard in SASU were not available to control for exposure to the risk of injury in each yard. If these data were available, the analysis would have questionable value because there was only one FRA reportable injury at Eagle Pass since the start of CAB–Switching. Statistical analyses are difficult with such a small sample size. If the analysis data set included nonreportable as well as reportable injuries there might be sufficient data for an analysis. However, nonreportable injury data were not available from UP.[11]

In conclusion, analysis of the injury rates provides no evidence that CAB reduces occurrences of this type. However, given the measurement and analytical limitations for injuries in this evaluation, this should not be regarded as evidence that CSA fails to reduce occurrences. Instead, the effect of CSA on injuries is inconclusive.

7.1.3.5 Safety Culture—Relations

Finding: Labor–management relations improve with a successful CSA implementation. There is greater trust and cooperation between workers and managers.

Table 27 summarizes results for labor–management relations.

Table 27. Summary of Results for Labor–Management Relations

Method	Result
Survey	The average scores of the Labor–Management Relations scale improved from 2005 to 2008.
Interviews	Workers and managers reported improvements in relations between labor and management. Workers reported greater commitment to safety by management. Workers reported greater fairness by management. Workers and managers reported greater trust between workers and managers. Workers and managers reported better communication and cooperation between management and labor on safety. CAB was credited with this improvement.
Field notes	Workers and managers cooperated on CAB, KEYS, and other initiatives.

The Labor–Management Relations scale on the surveys showed that both managers and workers reported improvements in labor–management cooperation from 2005 to 2008, consistent with the prediction that CAB improves labor–management relations. This result is also consistent with the results reported by BST from its own scales on the survey (BST, 2008). The BST scales indicated that, from 2005 to 2008, workers on average:

- Trusted UP management more.
- Believed that UP management values safety more.
- Considered UP management to be more concerned about worker well-being.

BST scales showed these improvements in these worker attitudes regarding both immediate supervisors and the organization as a whole.

---

[11] That the evaluation team succeeded in obtaining nonreportable as well as reportable *incident* data was largely due to a communication error within UP.

Both workers and managers perceived improved relations, although managers tended to report more labor–management cooperation than did workers. The difference in perceptions between workers and managers did not change from 2005 to 2008, implying a continued lack of consensus between workers and managers. Furthermore, although workers reported better relations in 2008 than in 2005, improving from an average of 2.65–2.88, average perceptions in 2008 were still unfavorable, being less than 3 (neutral) on a 5-point scale. In summary, although the Labor–management Relations scale shows improvements associated with CAB at SASU, there is room for further improvement.

Interviews and field notes corroborated survey results. From initial through midterm to final interviews, workers reported trusting management more. This increase was associated with workers perceiving a greater management commitment to safety and greater management fairness regarding safety. Managers' opinions of workers likewise improved. Along with greater mutual trust, workers and managers reported greater cooperation and communication between labor and management. New avenues for labor–management cooperation included KEYS and other safety initiatives outside of CAB. The joint labor-management softball team mentioned by an interviewee likewise indicates an unprecedented level of cooperation and camaraderie among workers and managers at SASU.

As discussed in Context on page 115, field notes and interview data indicate that labor–management relations were already changing before CAB was initiated. One may wonder if relations would have improved without CAB. Field notes and interviews suggest that CAB was necessary for improving relations. In interviews, respondents specifically pointed to CAB as being responsible for improved relations. Labor and management leadership may have wanted cooperation, but CAB was *the means for ongoing cooperation*. Specifically, the barrier-removal process, whether through KEYS or CAB's own Barrier Removal Team, provided a continuous improvement process with sustainable opportunities for managers and workers to address safety through cooperation rather than bargaining.

Without CAB to allow labor and management leaders to cooperate over the years, the nascent cooperation that followed the accidents may have withered. The Safety Command Center and Crew Resource Task Team safety initiatives probably would not have had as much of a long-term impact on labor–management relations as CAB. Despite both of these alternative initiatives being successful in producing cooperative safety improvements, neither initiative had mechanisms such as continuous improvement and barrier removal for maintaining the cooperative process over multiple years. The Crew Resource Task Team was concerned with implementing responses to issues raised specifically by the Macdona accident, such as printing the CRZ checklist. Once they were implemented (e.g., the CRZ checklist printed), the team's purpose would be completed and cooperation would end. Originally, the Safety Command Center provided innovative exercises to train workers on safety. During the evaluation period, its activities apparently degraded to providing simple safety briefings only. One worker suggested that personnel running the Safety Command Center ran out of ideas. It appears that neither the Crew Resource Task Team nor the Safety Command Center was structured to *continuously* improve safety.

In contrast to the Crew Resource Task Team and the Safety Command Center, the KEYS committee seems to have provided an effective forum for ongoing labor–management cooperation and deserves credit for some of the longer-term improvement in labor–management relations at SASU. However, since CAB and KEYS were closely related, with steering committee members serving on KEYS and using it for some of CAB's barrier-removal process, it is hard to separate the effect of CAB from KEYS on labor–management cooperation.

Although this demonstration pilot indicates that CSA can improve labor–management relations, it would seem that other cooperative safety processes may also be effective. These processes, such as a confidential close-call reporting system or joint labor-management root-cause analysis of incidents, may improve relations much as CSA has done. The necessary feature appears to be an institutionalized process for labor and management to jointly and continuously contribute to safety. A site may combine multiple cooperative processes. At SASU, CAB was combined with KEYS, suggesting that CSA can be used with other cooperative safety processes for increased effect on relations.

7.1.3.6 Safety Culture – Internalization and Generalization

Finding: Qualitative data indicate improvements in safety awareness, personal responsibility for safety, and safety dialogue among site personnel. These improvements appear to result directly from CSA implementation activities and to mediate the improvements in practices and relations.

Table 28 summarizes results for internalization and generalization.

Table 28. Summary of Results for Internalization and Generalization

Method	Result
Survey	Scales measuring safety-oriented norms (the Propensity to Safe Behaviors scale and the Coworker Safety scale) indicate no changes from 2005 to 2008.
Interviews	Workers and managers reported increases in safety awareness and personal responsibility for safety among workers from 2005 through 2007. Informal safety dialogue among peers reportedly raised safety awareness and personal responsibility for safety among workers. Managers also encouraged safety awareness and personal responsibility through coaching the workers on safety.
Field notes	Managers reported examples of workers taking more responsibility for safety, including reporting their own at-risk behaviors. Workers reportedly began to have more safety awareness and personal responsibility after attending CAB training.

The survey administered in 2005 and 2008 included two scales, Propensity to Safe Behaviors and Coworker Safety, which were intended to measure internalization and generalization of safety culture. Both scales measured safety-oriented norms. Neither scale showed any significant changes between the 2005 and 2008 administrations of the survey. Both scales exhibited low inter-item reliability, with the reliability of Propensity to Safe Behavior being inadequate. The low reliability may account for the lack of a detectable change between 2005 and 2008. The Coworker Safety scale also may have had a ceiling effect that could be equally responsible for the lack of a detectable change. The Coworker Safety means for 2005 were close to the maximum possible, effectively restricting the scale's ability to measure improvement in 2008. Both scales were based on scales used successfully in industries other than railroad, and it is possible that they may not always be appropriate for that industry. Unfortunately, they were the only quantitative measures of internalization and generalization of safety culture in this evaluation.

Both field notes and interviews provide qualitative evidence for the development of safety awareness and personal responsibility for safety among SASU personnel. From the initial interviews in 2005 through the final interviews in 2007, workers reported greater awareness of their capacity to control work safety through their behavior. This awareness extends beyond the initial behaviors target by CAB to include other situations both at and away from work. In field notes and interviews, managers reported that workers were taking greater responsibility for safety and showing greater awareness of hazards. There were even cases of workers reporting their safety missteps or deficiencies to managers.

Like the rest of safety culture, a change in safety awareness was hypothesized to be a distal outcome (see Figure 7, page 38), apparent later in the evaluation period after CAB-targeted practices and systemic conditions changed. Contradicting this hypothesis, increased safety awareness was reported in the initial interviews, suggesting that safety awareness coincided with or even preceded changes in practices and conditions.

As with labor–management cooperation, it seems likely Macdona and the other accidents helped to raise workforce awareness about safety and heightened the perception of risk in the workplace, shocking workers and managers out of complacency. However, workers not only recognized the risk but also felt empowered to reduce it. The presence of accidents alone cannot account for this feeling of empowerment. One can easily imagine that a high-profile accident would have the opposite effect, leading to feelings of helplessness or fatalism among the workers.

What accounts for feeling empowered? In the case of SASU, field notes and interviews indicated that the increase in safety awareness and empowerment started with CAB training. Early in the training, students discussed with peers their own safety and at-risk behaviors at work and at home. They also explored the reasons for engaging in at-risk behaviors, leading them to realize that many behaviors were a choice. The students watched video examples of on-the-job at-risk behaviors. They reportedly had seen such behaviors on the job (some may have even committed them), but viewing the video allowed them to appraise the behaviors objectively and consider alternatives.

These classroom exercises provided an opportunity for workers to engage in safety dialogue with each other. By talking about safety and behavior with peers, away from threats of discipline, workers recognized their capacity to choose behaviors that contribute to safety. According to one CAB trainer, it was a "tremendous" new way of thinking about safety, quite different from workers' traditional technological orientation.

The personal change interviews completed in 2008 indicated that safety dialogue continued to stimulate safety awareness beyond the classroom. CAB feedback sessions included a formalized safety dialogue in which workers and samplers discussed the observed safe and at-risk behaviors. The personal change interviews showed that CAB also encouraged informal safety dialogue in a variety of contexts. Workers conducting third-party feedback sessions seemed likely to engage in such informal dialogue. In the third-party sessions, a worker took the role of a safety specialist for the day. This may have encouraged the sampler to raise safety topics with peers, and it may have encouraged other workers to approach the sampler with questions or comments about safety. Such one-on-one or small-group conversations allow topics to be personalized, which encourages internalization. Because the sampler is a peer, such discussions can be frank, without concern about discipline.

The peer-to-peer aspect of the CAB process encourages workers to engage in safety dialogues with each other. SLD may also provide managers with a role in safety dialogue. Including managers in conversations about safety, as reported in the personal change interviews, provides another voice to encourage change. This can be effective if managers use these opportunities to apply their safety leadership training. Informal safety dialogue will be most effective if managers listen to worker concerns and coach them rather than command and discipline them.

It would seem, then, implementation activities, such as observer training, feedback sessions, and management safety leadership training, apparently directly encourage an increase in safety dialogue, leading to more safety awareness and personal responsibility early in the process. The effect of implementation activities is shown in the modified logic model (see Figure 26) by the addition of an arrow from Implementation to Internalization and Generalization. Internalization and generalization gradually change in response to the implementation and other factors, with safety awareness leading to safety dialogues that ultimately produce solid changes to social norms as the culture refreezes.

The addition of the connection between implementation and internalization and generalization implies that internalization and generalization is central to all outcomes. It is affected by nearly all activities and outcomes, and it also affects all outcomes. At a personal level, safety dialogue and the resulting feelings of empowerment and responsibility are relatively inconspicuous in the scheme of a service-unit-wide process. The significance of safety dialogue is easy to miss in the search to detect broader outcomes like safety-oriented norms, labor–management relations, and reductions in occurrences. However, the small-scale one-on-one conversation between workers or between a manager and a worker may be the linchpin around which all other outcomes revolve. With this evaluation lacking any quantitative measures of safety dialogue or personal responsibility, further research is needed.

7.1.4 Sustainability

Field notes suggest that CAB is sustainable given its support from four groups of stakeholders: workers, unions, management, and other safety programs. The support of these organizations apparently resulted from CAB's continuous efforts to:

- Promote CAB visibility.
- Engage the organizations in CAB activities.
- Address the organizations' interests.

Long-term sustainability requires that CAB continue to adapt to the challenges presented by changes in the organizational environment. Fortunately, CAB has a history of recognizing challenges and making such adaptations. The reactive nature of the railroad culture works to the benefit of sustainability of CAB. Railroaders are inclined to be vigilant for problems and to quickly and creatively marshal resources to address them. This tradition allows CAB to overcome challenges before they become unmanageable.

7.1.4.1 Worker Support

Finding: CSA training, feedback sessions, and integrity of the data protection facilitated sustaining the support of workers.

CAB sustained worker support by raising the visibility of CAB through the feedback sessions, the newsletter, and informal communications. The rotation of steering committee members resulted in periodic drives to recruit new members, which in turn increased the visibility of CAB to workers.

The CAB process engaged workers in CAB activities thereby promoting CAB sustainability primarily through training and opportunities to conduct feedback sessions. CAB encouraged workers to conduct feedback sessions by requiring that samplers conduct cross-crewmember sessions before they could conduct third-party sessions. The rotating pool of third-party samplers made the process fair and predictable for the workers. With such clear and open rules for selecting third-party samplers, workers felt encouraged to participate. The regular rotation of steering committee members was also a factor, because it helped workers view the steering committee as made up of workers like themselves, which provided them with vicarious engagement and promoted the credibility of the committee. As CAB expanded to outlying terminals and yards, more workers at SASU became involved in these activities. Steering committee members were added from outlying areas, with some areas such as Laredo receiving some autonomy (e.g., control of their own budget), thus engaging local workers more in CAB.

The primary interest for workers was that CAB data not be used against them, either directly for discipline or indirectly in selecting areas for FTX testing. Sustained worker support depended on maintaining the integrity of the protection of CAB data. During the evaluation period, there were no cases of observation data being used against workers, and just as importantly, no *perception* that such data were being used against them. As revealed during the interviews, when the integrity of data protection was maintained over several years, the primary worker concern over CAB faded.

7.1.4.2 Union Support

Finding: CSA can enlist union support by including officials and members from all concerned unions on the steering committee and barrier removal team, through regular education of union officials and members, and through the integrity of the data protection.

CAB engaged union leaders by including them in the steering committee and Barrier Removal Team. To allay union concerns that CAB might add to the discipline applied to workers, union leaders and key union members attended CAB training. When three BLET members moved against CAB in 2005 by threatening a facilitator with official sanction, the training helped to change the attitude of two of them. Maintaining data protection also resulted in continued union support. Educating the unions about the CAB process also showed how it was consistent with other union interests. Given their interest in improving workplace safety conditions, unions favored the CI component of CAB. In 2006, this awareness apparently warded off an effort by the union to pull out of CAB in retaliation for a change in the disciplinary policy. Balancing the membership of the steering committee and the Barrier Removal Team between BLET and UTU, in addition to including a facilitator from each union, was consistent with each union's political interest.

Despite this support, there were also apparent conflicts for union officials between defending worker interests and supporting CAB. CAB requires cooperation with management, but unions historically exist to bargain with management. Resolving this conflict remains important for the long-term sustainability of CAB. More research is needed to on the impact of cooperative processes like CSA on unions.

7.1.4.3 Management Support

Finding: Management support for CSA can result from regularly scheduled contacts with CSA personnel, metered participation of management in the CSA process, positive outcomes by CSA, and the willingness of CSA workers to negotiate and achieve an effective balance of cost and value.

Like union officials, management attendance at CAB training appeared to have improved the visibility of CAB and its subsequent support. CAB facilitators met with management leadership regularly, also improving the visibility of CAB to management. The Oversight Committee telcons were an important venue for this, providing visibility at the regional and corporate levels of management, which encouraged local management to support CAB. The telcons also allowed regional and corporate management to demonstrate CAB support and encouraged local management buy-in. The telcons were discontinued at the end of the evaluation period. Long-term sustainability will likely require replacement scheduled contacts between CAB personnel and at least local management.

Management engagement requires manager participation in CAB. Membership on the Barrier Removal Team was one effective role. Throughout the evaluation period, managers became more involved in other ways, primarily in promoting and protecting CAB, including representing CAB to corporate management at critical times. When faced with the transition to TSC, local management attended the TSC training with the CAB facilitators and helped them to integrate the two programs without sacrificing the core elements of CAB. Local management also championed the corrective actions at the corporate level, as seen with the policy change regarding air conditioning.

At the end of the evaluation period, the culture at SASU may have been intolerant of greater management involvement in PPF activities, as seen in 2007 with the experiment of a manager conducting a feedback session. However, long-term sustainability may require a more substantial role for managers. The best evolution of manager engagement in CSA remains an open issue that warrants future research.

The cost of CAB was of major interest for management. However, when concerned with costs, regional management chose not to take over budgeting CAB, but left it in the hands of the workers, and instead provided motivation to CAB to control costs on its own. In response to a challenge from regional management, CAB was able to cut administrative costs to receive a matching increase in budgeting offered by management. This negotiated agreement between management and workers not only allowed CAB to cover at an acceptable cost all of transportation of the entire service unit but also demonstrated to management the workers' capacity to adjust the operation of the process to meet critical management interests. This willingness to negotiate and compromise was essential for management feeling the process was a worthy long-term investment. For CAB, it meant less time in meetings and more time training workers and in conducting observations. It was a win-win solution resulting from open communication and problem solving.

As long as CAB can continue to keep costs comparable with those of other safety initiatives like TSC and to show similar results in variables such as occurrences, management support should continue.

7.1.4.4 Mutual Support of Other Safety Programs

Finding: CSA sustainability improved through mutual support with other safety programs both within and outside the site.

CAB connections to other organizations, in addition to management and unions, suggested good sustainability. Locally, steering committee participation in KEYS helped to justify the steering committee's existence beyond its roles in conducting feedback sessions, collecting and analyzing data, and developing corrective actions. CAB extended its influence through the steering committee, not only submitting problems to KEYS but also offering solutions through its own resources, such as offering video expertise and equipment to make a film on a new mentoring system.

Communication exchanges between CAB and other CSA-like processes also bode well for sustainability. By exchanging knowledge and experiences with similar organizations at Houston and Livonia, each can quickly evolve into a sustainable process. CAB personnel initially viewed the expansion of TSC as a threat, potentially ending CAB as it was known. However, management and CAB personnel slightly modified CAB to allow it to use TSC resources without losing the elements of CAB that are key to workforce acceptance. Inclusion under the umbrella of TSC improved CAB's long-term sustainability instead of threatening its viability. As a member of a larger movement within UP, CAB has access to TSC's resources and support from corporate management as well as to knowledge exchange opportunities with TSC sites.

7.1.5 Transformation

Finding: CSA can be effective in promoting a transformation in the broader corporate organization toward more proactive, nondisciplinary approaches to safety in transportation departs across multiple sites. The transformation can result from CSA making its successes known to decision-makers.

According to the field notes, from the middle to the end of the evaluation period, UP embraced CSA-like processes for transportation. Similar processes were initiated at other service units in the Southern Region, and eventually, UP rolled out a plan to establish a

CSA-like process under the name of TSC for the transportation departments of all service units in the corporation.

The TSC expansion appeared to be a direct result of UP decision-makers recognizing the apparent beneficial outcomes of CSA, which some had recognized even before CAB was initiated. The SASU superintendent and Southern Region vice president, at the start of CAB in 2005, had seen positive outcomes with TSC at the North Platte mechanical department, and they supported CAB at SASU on transportation. Midway through the evaluation period, BST and the evaluation team announced positive outcomes from CAB to decision-makers, which seems to have encouraged decision-makers to expand CSA to transportation departments across UP.

The following attributes of CAB appeared to assist in the transformation of the organizational environment:

- A safety process with the potential for positive outcomes.
- A thorough implementation to realize these outcomes.
- Sufficient time and commitment of resources for the outcomes to be detectable.
- A means to detect the outcomes.
- A means to bring the outcomes to the attention of decision-makers.

The successes of PPF in other industries indicated the potential of CSA for positive outcomes in the railroad industry. Following BST direction, CAB personnel continuously assessed its own implementation. The ability to self-assess, bolstered by the formative aspect of this evaluation, allowed CAB to detect weaknesses in the implementation and correct them to various degrees, resulting in a thorough implementation.

The reactive tendencies of the railroad culture do not favor long evaluation periods. With CAB stakeholders feeling pressure from the organization for immediate results, transformation depended on BST and the evaluation team obtaining management commitment to evaluate CAB over a sufficiently long period. The 2-year-plus length of this evaluation period was minimally adequate for successfully detecting outcomes. An evaluation period of 1 year appears to have been insufficient, as suggested by the failure to detect significant differences in occurrences between the San Antonio Complex and Laredo yards.

CAB used an independent evaluation team to detect outcomes, and BST also conducted analyses for outcomes, including improvements in occurrences. CAB thus coincided with a plan to identify and measure specific outcomes, which provided input for transforming the organization. The railroad industry has a tradition of numerically evaluating performance, so CAB’s association with statistical safety improvements helped to persuade decision-makers of its value. Reports of qualitative changes can also influence decision-makers. The inherent random nature of occurrences can make detecting statistical trends difficult, whereas qualitative changes may be apparent sooner. The ability of CAB to transform UP was assisted by the presence of both quantitative and qualitative means of detecting safety improvements associated with CAB.

Finally, CAB received regular formative evaluation feedback through sessions organized by BST and the evaluation team plus the regular Oversight Committee telcons. These meetings provided opportunities not only to monitor and improve the implementation but also to disseminate the latest outcome information, providing CAB with the means to bring outcomes to the attention of UP decision-makers.

7.1.6 Summary of Findings at SASU

Consistent with Lewin and Schein's (1992) theory of cultural change, fatal accidents and new members of the workforce at SASU triggered cultural unfreezing that spurred cooperation between labor and management to improve safety. With labor and management leadership, the CAB implementation capitalized on this unfreezing by working to achieve worker acceptance and management commitment for CAB. The CAB process then changed the culture at SASU by providing a new safety approach including changing worker and manager practices, reducing safety occurrences, raising safety awareness, safety dialogue, and personal responsibility, and increasing cooperation and trust in labor–management relations. Cultural refreezing began when top management sought to spread CAB-like processes throughout the corporation as the TSC process and then accepted and budgeted CAB under the TSC program. This institutionalization of CAB-like processes ensures CAB will be sustained and indicates CAB's transformation of the organization.

7.2 Validity of the Evaluation

This evaluation is perhaps the most comprehensive evaluation carried out by FRA to date. The demonstrated safety process was selected based on an understanding of the railroad industry context (Stufflebeam, 1983). The evaluation followed a logic model to assess not only the distal outcomes of interest, such as occurrences and safety culture, but also the pilot implementation, intermediary outcomes, and impacts of the broader organizations. It used a quasi-experimental mix-method design. The evaluation team collected data for 13 quantitative variables over a 2.5-year period, using implementation process metrics, observational data, surveys, and archival corporate sources. These data were reinforced by two types of qualitative data: field notes and interviews, including an in-depth case study analysis of personal change. Where possible, the team made comparisons with other data to separate the influence of factors other than CSA. The findings reported here are supported by multiple data sources, representing a convergence of evidence. In summary, the research documented here comprises features of the most rigorous evaluations (GAO, 2009).

7.2.1 Limitations of the Evaluation

Despite the comprehensive nature of the evaluation, there are some limitations to take into account when drawing conclusions. The findings were from a specific service unit at a specific point in history. Whether or not the same results can be achieved at other sites and times depends on the sites sharing the right features with SASU. Specifically, both labor and management at the sites must be open and willing to commit to the stringent approach of the CSA process (Initial Conditions on page 61 describes these features of the demonstration site). Beyond that, the CSA process will need to be adapted to the specific conditions at each locality to obtain the results documented in this report.

The results for CAB–Switching are tentative because of the following limitations of the CAB–Switching evaluation:

- A short 1-year period of evaluation.
- Difficulties in finding suitable comparisons for occurrence data.
- Lack of a scale for switching practices at the time of the initial survey administration.

These limitations probably account for CAB–Switching results being weaker than CAB–CRZ results. CAB–CRZ likewise had weak results after 1 year (Coplen, Ranney, and Zuschlag, 2008).

Injury data failed to show an effect, perhaps because of the limitations in evaluating CAB–Switching. It may also be due to the data including only FRA-reportable injuries. Nonreportable injuries are minor, requiring only first aid. However, they may also represent "close calls" with more severe injuries and can be considered a proxy measure for them. The inclusion of nonreportable injuries with reportable injuries in the data set boosts the sample size and improves the chance of statistically detecting an outcome. In the evaluation of a separate safety process on another railroad, the inclusion of nonreportable injuries made the difference between a significant finding and an insufficient sample size (Ranney, Shuang, and Zuschlag, in preparation). In the CAB–Switching evaluation, the dataset for human factors incidents included both reportable and nonreportable incidents, the latter being incidents where damage was below a threshold dollar value. The inclusion of nonreportable incidents increased the sample size, facilitating the detection of a significant improvement in incident rates for Eagle Pass. Future evaluations should pursue the obtaining of data on nonreportable injuries.

Data on worker practices consistently showed improvements after implementation of CAB, with the exception of FTX data. The FTX data indicated a decline in rule-compliant practices from the first to the second phase that was not corroborated in the process metrics or the interviews.[12] This lack of concurrent validity casts doubt on FTX as a measure of practices. Future research using FTX should include a systematic investigation of its fitness. The decrease in FTX performance approximately coincided with a change in corporate discipline policy. However, it is not clear how increasing the consequences of discipline for certain rules violations would make FTX failures more likely. One possible issue to investigate is manager selection of rules and locations for conducting FTX tests. Apparently, managers are not attempting to be random or even representative with these selections. Conceived as a training tool, FTX tests tend to be conducted for rules and locations where it is believed that workers are most likely to fail

[12] The survey included scales for worker practices, but there was no mid-evaluation administration of the survey to provide an additional check for a decline in practices between the first and second phases. The survey analyses could only compare baseline and the second phase.

the test. If management becomes more effective in identifying problematic rules and locations, measured FTX performance may decline when actual practices are unchanged. The process metric measures of practices (percent at-risk) may also be influenced by the CSA steering committee deliberately targeting locations of concern. However, this targeting uses a single data set that remains constant—the historical at-risk data itself. The decision process for FTX targeting, in contrast, is unknown, perhaps relying on manager intuition. FTX targeting may thus be based on shifting criteria, which makes it unreliable.

Quantitative measures for safety culture internalization and generalization appear to be inadequate for this evaluation, exhibiting low reliability and/or ceiling effects. The Propensity to Safe Behavior in particular showed inadequate reliability. This three-item scale includes items related to trusting one's supervisor for safety information. This may not be suitable for measuring safety-oriented norms in the railroad industry, where there is substantial distrust of management interest in worker safety and different orientations toward safety between workers and managers. The Coworker safety scale had low but adequate reliability. However, it was also hampered by a ceiling effect, which, when combined with low reliability, will make detecting improvements difficult. This scale may be improved for use in the railroad industry by adding items to increase reliability. New items should be written such that they exhibit a range of responses to reduce ceiling effects.

Field notes and interviews both indicated improvements in internalization and generalization. However, since only qualitative data are included, the evidence for internalization and generalization is moderate rather than strong. Future research should include quantitative measures of internalization and generalization. In particular, the frequency of safety dialogue should be measured because qualitative analyses in this evaluation suggest that safety dialogue mediates nearly all other outcomes.

Even though this evaluation found converging evidence that CSA improves safety and culture, one wonders if other kinds of safety processes would have provided better outcomes for a given cost. CSA costs can be estimated. For example, at the end of the evaluation period, monthly labor costs for CAB at SASU were 170 worker-days per month for a 1,100-person workforce.[13] In other words, CSA activities accounted for approximately 0.8 percent of the work time in the transportation department. However, deriving an aggregated value for the outcomes is an involved process. One approach includes summing the savings associated with a reduction of incidents. Railroads routinely collect the cost of damage because of incidents (e.g., NTSB (2006) reports that the Macdona accident caused $5.8 million in damage) but other costs are not. The total cost of an incident should include the labor costs to handle incidents, the costs due to schedule disruptions caused by incidents, and liability costs, including those associated with any injuries (Coplen and Ranney, 2003; Ranney and Nelson, 2007). The value of

[13] At the end of the evaluation period, CAB had a budget of 130 worker-days per month. Add to this two facilitators working approximately 20 worker days per month each brings the total to 170 worker-days per month.

cultural change should also be assessed and included in the savings associated with a process. More labor–management cooperation can mean less discipline and therefore fewer costs for handling discipline cases (e.g., time for hearings). Despite the difficulties in valuing the outcomes of a safety process, future research should include a systematic cost-benefit analysis to provide a basis for comparison with other safety approaches.

7.2.2 Alternative Explanations

7.2.2.1 Corporate and Regional Confounds

Because this evaluation was a field study and did not randomly assign participants to conditions, it is inevitable that confounds would emerge that could conceivably account for outcomes and impacts. CSA implementation at SASU coincided with other multiple events that might be considered an alternative cause of these outcomes and impacts.

CAB–CRZ startup nearly coincided with three safety-related changes in the region:

- *CRZ rules*. Shortly before startup, the operating rules changed to establish CRZ practices. It is reasonable to assume that this rule change had the intended effect of improving locomotive crew attentiveness and that this alone is responsible for the decline in decertifications at SASU.
- *Compliance agreement*. Shortly before startup, SASU and two other Southern Region service units came under a compliance agreement with FRA, requiring the service units to overhaul their FTX process, in addition to other changes. It is possible that these changes prompted the safety improvements documented at SASU.
- *Discipline policy change*. In November 2006, about 1 year after startup, UP corporate management increased the penalties for infractions of certain safety rules. Safety improvements since then could be attributed to this policy change.

All of these alternative explanations are implausible, given the results from the occurrence data, which show occurrence improvements specifically associated with CAB. Each of the three alternative causes above would uniformly affect all service units and all locations within each service unit. For example, the risk of decertification decreased only at SASU and not at other service units that also had changes in rules, had a new compliance agreement, and had a change in discipline policy. This implies that improvement is due to a specific event or process at SASU. Engaging in the new CRZ practices may indeed improve attentiveness, but only SASU had a process that encouraged workers to rapidly adopt the new practices. The greatest improvements in incident rates were observed for the Eagle Pass yard within SASU, the one location with a strong CAB–Switching implementation. Changes to rules, the compliance agreement, and the change in discipline policy should have affected all yards within SASU.

7.2.2.2 Local Confounds

There were also some local confounds with CAB that were present at SASU but not at the other service units. These confounds may provide alternative explanations for the findings. The confounds include:

- *Fatal accidents*. In the 2 years before CAB initiation, there were three fatal accidents at SASU, raising concerns about safety among the personnel.
- *New leadership*. New local leadership emerged in both labor and management shortly before the start of CAB. At Eagle Pass, the start of CAB–Switching coincided with new local leadership.
- *Other safety initiatives*. The Safety Command Center and Crew Resource Team were implemented shortly before the start of CAB, and KEYS was established shortly after the start of CAB.

Any of these confounds may be responsible for the findings reported in this evaluation. In fact, there is qualitative evidence that change was already under way at SASU before CAB conducted its first feedback session.

Any field setting has multiple factors occurring and influencing one another. It is misleading to identify one feature as responsible for all change. Saying, for example, that safety improved only because of leadership changes is as overly simplistic as saying that safety improved only because of the steering committee. Both leadership changes and creation of the steering committee, along with many other CAB-related factors, contributed to the successes at SASU. Many events at SASU played a role in outcomes and impacts, making CAB implementation possible and positive outcomes likely, providing sustainability, and encouraging organizational transformation.

A high-profile accident can raise concerns about safety, but translating that concern into effective action requires the right leadership to pursue the right approach. Any random leadership change will not necessarily result in safety improvements; these result from how leadership is exercised. In this case, at both SASU and at Eagle Pass, leadership chose to strongly support CAB and proactive, cooperative safety approaches in general. These actions appear to be responsible for improvements in safety. A spirit of cooperation can motivate labor and management to pool their resources and expertise to improve safety. However, a sustainable process like CAB is necessary to allow labor and management to effectively apply their resources to improve safety.

The other safety initiatives at SASU may have independently improved safety, but they were also related to the creation of CAB. The Crew Resource Team primarily contributed to safety by providing the personnel and organization to become the Design Team for CAB. As described in the subsection Safety Culture—Relations, on page 131, the Safety Command Center appeared to have a short-term role in direct safety improvements, but as described in the subsection Previous Successes with PPF Known to Key Players on page 63, its existence caught the attention of NTSB. This ultimately led to FRA sponsorship of CAB. Although the KEYS safety committee developed independently of CAB, the two collaborated extensively. If there had not been a committee like KEYS to use as a resource, to be successful, CAB would have had to create this type of committee, probably through modifying and expanding its own Barrier Removal Team.

If CAB did not exist at SASU, the local confounds might have contributed to safety on their own. However, in practice these “confounds” were the context that helped to make CAB a reality. They are all part of the chain of causation that made CAB what it is, and in this regard, they are part of CAB.

A railroader interested in replicating the outcomes of CAB at another site ought not search for the one unique element at SASU that brings about change or assume that events must exactly match those at SASU. Although an appropriate site context appears important for CSA success, one does not necessarily need new leadership, or a fatal accident, for that matter. The best approach is to appraise the unfolding events at SASU, from the initial context through the constant evolution of CAB, and consider how to achieve or exploit equivalent conditions and processes at the intended site.

7.2.2.3 Effects of the Evaluation and FRA Involvement

An additional feature present at SASU and not at other service units was the activities of the evaluation team. Rather than function as passive observers, the team conducted a formative evaluation with the specific aim of producing the best implementation possible. A railroader considering the development of a CSA process at another site may ask if such a comprehensive formative evaluation is necessary for success. The evaluation lacked systematic measures of its own impact. For example, the interviews and surveys did not include questions about the evaluation team. This makes the question of the importance of the formative evaluation difficult to answer, except through professional judgment of the events at SASU.

Looking back, the evaluators believed that team activities improved the quality of the implementation but only slightly. This is not because the team's formative feedback contained little useful information. Rather, much of the feedback from the evaluation team was redundant with that from the BST consultant.

The evaluation team contributed to BST's usual approach to implementing safety programs in the following areas:

- *Established a 2-year evaluation period (and then extended it to 2.5 years) that included BST involvement throughout*. Normally, BST spends only 1 year with a client, establishing an implementation. A longer evaluation period was important to complete a conclusive analysis. However, evaluation-team involvement complicated the task for BST, which also necessitated additional time for starting CAB.
- *Emphasized timely focus on the implementation and outcomes*. Like BST, the evaluation team promoted assessing and maximizing the implementation strength early in the evaluation period. Also, like BST, the team set expectations with the stakeholders that outcomes, particularly for occurrences, would not be immediately apparent. In working with the logic model with stakeholders, the evaluation team encouraged stakeholders to first look for implementing CSA as planned and then look for proximal outcomes before expecting distal outcomes like occurrences.
- *Establishment of oversight telcons that involved regional and corporate management*. Through FRA encouragement, corporate and regional leadership received regular briefings on CAB through monthly telcons and participated in resolving issues. Although other BST implementations have involved top managers in improving safety performance, such regular telcons were a first.

During implementation, the evaluation team often reinforced the information provided by BST to CAB stakeholders. During formative feedback sessions, the evaluation team identified areas of concern in the implementation, but BST and the CAB stakeholders identified how to address them. This gave CAB stakeholders and BST a substantially larger role in CAB implementation than the evaluators. Finally, involvement of the evaluation team required learning to coordinate BST and evaluation team activities without one interfering with the other. In the politically charged organizational environment of the railroad industry, where there are high levels of suspicion among players, feedback and support must sometimes be carefully crafted to avoid misunderstandings. Early in the implementation, the evaluation team may not have been as sensitive as necessary, and a few of their actions apparently conflicted with BST designs, thus disrupting rather than facilitating the implementation.

The evaluation team may have made a larger contribution to organizational transformation than to implementation. As in implementation, the evaluation team's role in transformation overlapped with that of BST, which also conducted analyses for outcomes and disseminated results to decision-makers in UP. The difference was in the evaluation team's effort, which was larger and more comprehensive than that of BST, which probably gave it more impact. They replicated some of the BST-reported results and were also able to report important results that BST could not, such as occurrence rates for comparison data. Furthermore, because the team did not have a financial interest in the success of CAB at SASU, it may have had more credibility in reporting the findings than did BST.

Organizational transformation was also enhanced by BST and the evaluation team association with FRA, which raised the profile of the project at UP. By paying BST to consult, FRA had leverage to encourage corporate cooperation and commitment. FRA established the Oversight Committee telcons to communicate CAB findings to decision-makers. It also gave CAB facilitators and local labor leaders direct access to management at the vice-president level and increased CAB visibility at high levels.

In summary, evaluation team activities were important to the success of CAB, in particular in encouraging organizational transformation. However, a successful CSA process does not necessarily require a full-blown independent formative evaluation any more than it requires precursor fatal accidents. To maximize the chances of successful implementations and transformations, a CSA process should include the *equivalent* of the activities of the evaluation team that helped CAB. These include:

- Collecting and analyzing data on implementation performance and outcomes.
- Feeding analysis results back to stakeholders to improve the implementation.
- Disseminating the results to decision-makers to promote transformation.

7.2.3 External Reviews

The evaluation in this report was subjected to two independent reviews, one a metaevaluation (Scriven, 1991; Stufflebeam, 2001) of the analytic procedure used on the occurrence data, and the other as part of the process of receiving an international award.

7.2.3.1 Metaevaluation of Analytic Procedure

A national expert in time-to-events data analysis was asked to review the statistical analysis procedures for the occurrence data, involving survival analysis of gap times (see Occurrence Data Analysis, page 48). The evaluation team contracted the expert to advise and instruct on the most statistically sound and powerful approach to analyzing the occurrence data, and to review and advise on specific analyses concerning:

- Suitably preparing the data for analysis.
- Selection of the specific analyses.
- Testing for assumptions of the analyses.
- Interpretation of the results of the analyses.
- Consideration of alternative approaches.

The expert's review included face-to-face interviews with the evaluators and detailed inspection of the data, statistics, and written interpretation of the results in this report. In written and verbal feedback, the expert concluded that the analytical procedures used in this evaluation were suitable for the data.

7.2.3.2 American Evaluation Association Award

This evaluation received the American Evaluation Association's (AEA) 2011 Outstanding Evaluation Award. AEA is an international professional association of evaluators devoted to the application and exploration of program evaluation, personnel evaluation, technology, and many other forms of evaluation (AEA, 2012). The award is given for the successful completion of a single evaluation project representing high quality methodology, high usefulness of the findings, and the ability of evaluation to serve as an exemplar for other evaluations. To win the award, an AEA panel of experts reviewed the evaluation for compliance with the following standards for evaluations (Yarbrough, Shulha, Hopson, and Caruthers, 2011):

- Accuracy, the dependability and truthfulness of evaluation representations, propositions, and findings.
- Utility, the extent to which program stakeholders found the evaluation valuable.
- Feasibility, the effectiveness and efficiency of the evaluation.
- Propriety, the degree the evaluation was proper, fair, legal, right, and just.
- Accountability, the adequacy of the documentation of the evaluation for the purposes of metaevaluation.

The expert evaluation panel based its decision on a review of an earlier draft of this report and other shorter write-ups of the findings. The panel also requested and obtained independent letters from stakeholder groups (FRA, railroad management, and unions officials) on value they perceived in the evaluation.

7.3 Moving Forward with CSA and Demonstration Pilots

The pilot demonstration at SASU illustrated the potential for CSA to improve safety and safety culture. It also demonstrated the value that FRA development projects can offer the industry.

7.3.1 Lessons Learned

To replicate CAB's success at other sites, the conditions and implementation must be correct. A review of the events related to CAB yields lessons learned in establishing a complete, sustainable, and ultimately transforming implementation. Railroad management, unions, workers, and FRA can use these lessons learned to achieve effective CSA processes at other sites. They expand on the basic features of CSA described in CSA Concept on page 13.

7.3.1.1 Context

The following lessons learned relate to the initial context required for a successful CSA implementation.

- *Motivation for change facilitates the success of CSA.* Before implementation, stakeholders should see a need for change so that they are willing to invest in CSA. Labor and management must jointly be open to cooperate on a new approach to safety. A fatal accident is a dramatic way to stimulate an urgent desire for change, but it is not necessarily the only way. Dissatisfaction with safety at a site may be the result of a gradual decline in safety performance or a chronically poor ranking relative to other sites. Although the most problematic sites are often motivated to change, other sites can also be considered. An interest in change may also be stimulated by exposure to established CSA sites with apparent positive outcomes. Other candidates are sites where unions and management are at a bargaining impasse on safety for which either side offers CSA as a "compromise" alternative approach.
- *CSA may best be started small, at the most favorable locations*. Given that railroad culture can challenge the implementation of CSA, success is most likely if it is first established at a few small sites with the most conducive context, then expanded to others as successes become increasingly apparent.
- *Effective leadership is important for a successful CSA process.* Workers and managers responsible for the CSA process must be committed to cooperating with each other and to completing the activities necessary to establish and sustain the CSA process. They must be creative, socially adept, and willing to take risks to overcome obstacles to the implementation. Committed and capable facilitators are especially important. They, along with other steering committee members, can be found by unions and management working together to identify workers who are pro-safety, favorable to labor–management cooperation, and credible with managers and workers alike. It is preferable that facilitators are not union leaders so that there is no conflict between cooperating with management on the CSA process and bargaining with management to protect worker interests. The management sponsor should be predisposed to work with workers and unions on

CSA and have the authority to address issues and be accountable for their resolution.

- *Support is important for the success of CSA.* In addition to those directly responsible for the CSA process, key individuals at the site, such as local and regional union leaders and upper managers, must be educated on the CSA process and understand what it involves. Their interests and concerns must be addressed, and preferably, they should have a role in the process to maintain their engagement. Managers seeking to implement CSA should communicate with union leaders at the start, providing compelling assurances of the integrity of the data protection, and soliciting their suggestions and support. All site managers should know their role in CSA particularly in relation to SLD. Managers can expect that the process implies spending less time on oversight of workers (since workers will look after themselves more) while providing better-quality oversight. Managers need to be aware that detectable safety improvements cannot be expected until a substantial portion of the workforce is trained and high contact rates are achieved, which may take years. Unions seeking to implement CSA should prepare their local officials and the rank and file for the dual roles of cooperating with management on safety through CSA while still protecting worker interests outside of safety.
- *Expertise on CSA-like implementations is very valuable*. Errors in detecting and addressing concerns or in otherwise managing resistance can delay or derail an implementation. Expert advice and direction, such as from a trained and experienced consultant, can prevent such errors. Consultation should begin at the start of the implementation and continue for about 2 years after feedback sessions are started. The cost to the company will likely be justified by a faster and smoother implementation and better sustainability.
- *The costs for CSA are not trivial.* In addition to the cost of the facilitator (probably the single largest cost), CSA includes smaller but significant other costs for which management must plan. An effective implementation includes timely investment in worker and manager time, along with space and supplies. Management willingness to provide time and resources is important for gaining worker acceptance. Management needs to plan for the resources from the outset, including being prepared to quickly address safety barriers when they are identified in order to demonstrate management commitment to CSA.
- *Promoting CSA at a site as an FRA demonstration pilot provides effective impetus.* FRA involvement can help stakeholders commit to implement CSA as planned and continue the process for the period necessary to assess its effectiveness.

7.3.1.2 Implementation

The following lessons learned relate to characteristics of a successful implementation:

- *Education is the primary tool to overcome resistance.* Training for conducting feedback sessions is especially effective for education. It can be used not only for workers but also for union leaders and managers throughout the hierarchy. Face-to-face informal communications and mass media, such as newsletters and fliers, can also assist with education.

- *The CI and PPF components of CSA can be promoted to gain the support of labor and management, respectively.* In a traditional railroad setting, CSA can be attractive to both labor and management because it includes components consistent with the safety orientations and interests of each organization. The PPF component is consistent with the procedural orientation of railroad management, whereas the CI component is consistent with the technology orientation of railroad unions.
- *Effective behavioral safety processes can be run by workers.* Railroad workers are enthusiastic about contributing to safety, such as through conducting third-party feedback sessions. As long as the improvement process is nondisciplinary and under their control, workers will embrace the notion that their behavior affects safety. Although management should have a role in supporting and advising the workers in running a CSA implementation, in general, the more control that workers have over the implementation the better. In particular the workers should have sole control of the PPF data collected, and management should avoid any appearance of a connection between management discipline and the CSA process. Management efforts instead should focus on barrier removal and SLD rather than PPF.
- *Workers need management support.* Although the process is primarily worker-driven, workers will need management support to be successful. This support includes making verbal statements of support, providing an adequate budget, actively avoiding interference with feedback sessions (e.g., by field operations tests), coordination of other safety processes (e.g., safety committees and safety hotline), and prompt removal of barriers.
- *Observation and feedback with a dispersed workforce remains difficult and costly but is essential for success.* The use of cross-crewmember feedback sessions alone may not result in a sufficient contact rate. Management should expect third-party feedback sessions to play a significant role and budget accordingly. Using third-party sessions to motivate the completion of cross-crewmember sessions is effective, but other approaches may be even more effective. These approaches should aim to reduce the workload and the interpersonal threat associated with cross-crewmember sampling.
- *Cultural change can start early.* Cultural "unfreezing," particularly visible in internalization and generalization, is not necessarily a downstream distal outcome but rather a natural result of implementation activities such as training, observation, and safety leadership practices. Given its role in influencing other outcomes, such early cultural change is desirable.

7.3.1.3 Sustainability

The following lessons learned relate to characteristics of a sustainable implementation:

- *Sustainability is strengthened by regular connections, communication, and coordination of organizations invested in CSA.* An open process enhances sustainability. Examples of connections, communication, and coordination include both regular meetings and informal contacts between the facilitators and management sponsors, and among the workers running CSA, local managers, and local union leaders (such as the oversight telcons for CAB). New top management should be briefed early on the benefits of the process. Sustainability is aided by contacts outside the site: with the superiors of the local managers and union leaders, with other sites implementing CSA, and with other safety organizations, including those with members from other trades.

- *Management's role in CSA must be metered in proportion to the cultural change for sustainability*. Long-term sustainability requires a role for managers, but initially, CSA needs to be primarily worker-driven. The integrity of the data protection in particular is important. There must not be even the appearance of management seeking CSA data for its own purposes. Generally, workers should share information on all safety barriers they find so that management resources may be applied to their prompt removal. However, even here some barrier-removal may be strictly conducted by workers, if they have the resources and it is necessary for worker acceptance. Over time, managers should be phased in to all aspects of CSA, first promoting and eventually assisting in CSA activities.
- *Top management's role includes holding the steering committee accountable for performance of its CSA activities, and championing the process at the corporate level.* In addition to providing the workers with the resources, management should hold the workers accountable for how effectively those resources are used. In keeping with workers running the process, management should avoid directing the use of the resources, but rather set incentives and realistic goals for the workers to achieve. Top management also serves as the interface between the corporation and the CSA process, protecting it from threats (e.g., by integrating the CSA process with competing safety programs) and promoting barrier removal that requires company-wide action.
- *Sustainability is strengthened by including activities to maintain the visibility of the process and its outcomes.* The CSA steering committee should publicize to workers and managers (e.g., through newsletters, or union and company publications) the feedback session frequencies, the barriers removed, the detected improvements in practices, and significant activities (searches for new members). Management should likewise publicize its activities in SLD and barrier removal, listing barriers removed and updating workers on progress being made on long-term barriers. Company and union leaders can build support for the process through symbolic public actions indicating commitment to it, such as attendance at CSA-related events and issuance of supporting statements. Preserving the name of the process as it evolves also helps to maintain support.
- *Sustainability is enhanced by expanding the scope of CSA implementation, such as including multiple operations (e.g., road and switching) and multiple locations.* Scope expansion raises visibility and engages more workers and higher members of the union and management hierarchy, leading to greater support.

- *Local contacts with personnel running CSA strengthen sustainability by increasing visibility and engagement.* Each location within a service unit benefits from having a local CSA expert available to address concerns and to encourage activities. Larger locations may each have a certain degree of autonomy in carrying out activities.
- *A CSA implementation needs a means of measuring its performance and correcting its problems.* The CSA process should be self-measuring, self-critical, and self-correcting. The steering committee must continuously collect and evaluate process performance information and work out the inevitable adjustments. The implementation should use measures of proximal as well as distal outcomes to detect early progress in safety improvements.
- *Flexibility and creativity are necessary to continuously adapt to new challenges as the process expands and the organization around it evolves.* Each site is different, and sites change with time. Thus, each new site needs to make its own adaptations and have local ownership of its process. A larger authority can provide education and resources for each CSA site, but each implementation should ultimately be the responsibility of each site. Across sites and over time, management may need to adapt its existing procedures to boost credibility of the CSA process among workers (e.g., by foregoing field testing when feedback sessions are in progress). Steering committee members may need to adapt the CSA process to address management concerns, such as administrative costs and consistency with corporation-wide safety programs. Rotation of new facilitators and steering committee members may inject new perspectives to keep the process going. When many sites are established, union or corporate leadership may consider a regular conference for facilitators and other steering committee members where they present their local adaptations and learn from each other.

7.3.1.4 Transformation

The primary lesson learned related to organizational transformation is the importance of forwarding CSA implementation and outcome information to superiors in both the corporate and union hierarchy to influence decision-making. The process should include early and regular updates on both proximal and distal outcomes to make the case for CSA expansion and the embracing of proactive, nondisciplinary approaches to safety. Updates should include qualitative signs of improvement; these may be apparent sooner than quantitative changes in safety. Site visits by superiors can include exposure to the process, such as through attendance of worker training for conducting feedback sessions. Corporate managers can promote managers from the CSA sites, or transfer them to other sites, in order to promote changes in corporate safety culture.

7.3.2 Implications

With the substantial weight of evidence from this evaluation showing that CSA is capable of improving safety and safety culture, the railroad industry and railroad workers would benefit from CSA implementation at sites across the country. If the FRA chooses to promote CSA implementations, it may do so through the following:

- Raise awareness of CSA as a risk reduction strategy and its potential outcomes and impacts through publicizing the results of this and other evaluations. Use interim reports, presentations, and telcons to highlight progress and outcomes to capture the attention of top corporate decision-makers.
- Form a national working group of FRA, labor officials, and railroad representatives charged to expand the use of CSA in transportation and other departments across the industry.
- Identify railroad companies to suggest that they try CSA by searching for sites with conditions conducive to a CSA implementation (e.g., those with an above-average disposition toward labor–management cooperation).
- Provide expertise for establishing the implementation, through either contracting with consultants or acquiring the expertise through education and training.
- Provide third-party formative evaluation services to give credibility that the implementations are complete and that outcomes are measured.
- Sponsor forums for the exchange of ideas and perspectives among companies and unions and for establishing and maintaining effective CSA processes. Help to broker differences between companies and unions to get CSA established.
- Explore means to make CSA more affordable to railroad carriers, especially smaller ones for which consulting fees may be daunting.

8. Conclusions

CAB has demonstrated the potential for CSA to improve safety and safety culture in the railroad industry. CSA can be implemented in railroad transportation departments and adjusted for compatibility with obstacles presented by a reactive command-and-control safety culture, a dated safety analysis orientation, poor labor–management relations, rotating management, a dispersed work setting, and competition among unions. CSA can be sustained through maintaining its visibility with labor and management leaders, enlisting their engagement and addressing their interests.

The site for this demonstration pilot included features that perhaps made it more conducive to establishing CSA than at a typical railroad service unit. However, now that CSA has been established at one site and word is spreading about its successes, this is helping in the transformation of the region and company, making it easier to establish CSA at other sites.

As CSA continues to evolve there are still problems to address, including the challenge of achieving a high rate of feedback sessions on the road and the ultimate role of management. Even so, overall the future looks bright for CSA.

9. Epilogue

By 9:00 a.m., the bright Texas sun was already building the heat over the siding at Macdona a year to the day after the fatal collision between a UP and a BNSF train. A low-key ceremony was underway. Approximately 30 UP workers, managers, and family were in attendance, nearly half of whom were members of the recently created steering committee for CAB or the CAB design team. The local management sponsor for CAB arrived just before the start of the ceremony. He had just come from one of the San Antonio yards where he had dealt with a small derailment.

Passing trains at Macdona had been instructed not to sound their horns that morning as a sign of respect, and ceremony attendees took turns protecting a nearby grade crossing. They dug a hole to plant a tree that would serve as a living memorial. Across the tracks, flowers were blooming in a field that coincidentally had once been owned by the fallen conductor's grandfather. "Hard to believe that one year ago this place was so poisoned we couldn't be standing here," mused a design team member. But there were still some signs of the tragedy. A line of gnarled trees along the tracks stood dead, their leaves burned off by the gas. A steering committee member sampled some dirt from the hole dug for the memorial tree; the smell of chlorine was still detectable. It would take more time before things were truly better.

The tree was planted with compost, mulch, and a generous barrel of clean water. Attendees gathered in a circle and some words were spoken, then they broke off into small groups. Conversation turned to the new CAB process being implemented. Progress had been made toward the planned start of observations-feedback sessions in a couple of months: a brochure had been prepared to introduce the process to the workers, and videos for the training class were nearly complete. At approximately 11:00 a.m. the steering committee members left for San Antonio to meet with the BST consultant to plan the next steps in implementing CAB.

From the chlorine-stained grounds of Macdona, a small tree now stood, its young branches reaching upward and out.

10. References

Ambrose, S. E. (2000). *Nothing like it in the world: The men who built the transcontinental railroad 1863-1869*. New York: Simon & Schuster.

American Evaluation Association (2012). About us. Retrieved from http://www.eval.org/aboutus/organization/aboutus.asp.

Armstrong, J. H. (1982). *The railroad- what it is, what it does: The introduction to railroading* (revised). Omaha, NE: Simmons-Boardman.

Behavioral Science Technology, Inc. (2008). *Organizational culture diagnostic instrument: prepared for Union Pacific Railroad San Antonio Service Unit*. Unpublished final report submitted to the Federal Railroad Administration and Union Pacific.

Baram, M. (1997). Shame, blame and liability: Why safety management suffers. In: Hale A Wilpert B and Freitag M. (Eds.) *After the Event: From Accident to Organizational Learning*. New York: Elsevier. 161-178.

Bell, K. (2010) Personal communication.

Box-Steffensmeier, J. M., & Jones, B. S. (1997). Time is of the essence: Event history models in political science. *American Journal of Political Science*, *41*(4), 1414-1461.

Boyd, G. (2008). *Best practices in leading health & safety*. Behavioral Science Technology, Inc. Retrieved from http://www.bstsolutions.com/perspectives/article.php?item=2.

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part III: Basic concepts and first analyses. *British Journal of Cancer*, *89*, 605-611.

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: Basic concepts and first analyses. *British Journal of Cancer*, *89*, 232-238.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cook, R. J. & Lawless, J. F. (2010). *The statistical analysis of recurrent events.* New York: Springer Science + Business Media.

Coplen, M., & Ranney, J. (2003). *The impact of safety rules revisions on safety culture, incident rates, and liability claims in the U.S. railroad industry: A summary of lessons learned.* Federal Railroad Administration Research Results RR03-03. Retrieved from http://www.fra.dot.gov/downloads/Research/rr0303.pdf.

Coplen, M. Ranney, J. & Zuschlag, M. (2008). *Promising evidence of impact on road safety by Changing At-risk Behavior process at Union Pacific*. Federal Railroad Administration Research Results RR08-08. Retrieved from http://www.fra.dot.gov/downloads/Research/rr0808.pdf.

Creswell, J. W. (2003). *Research design: qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Dastmalchian, A., Blyton, P., &Adamson, R. (1989). Industrial relations climate: testing a construct. *Journal of Occupational Psychology*, *62*(1), 21-32.

Deming, W. E. (2000). *Out of Crisis*. Cambridge, MA: The MIT Press.

Department of Defense 1996. Handbook for reliability test methods, plans, and environments for engineering, development, qualification, and production (MIL-HDBK-781A). Washington, DC: U.S. Department of Defense.

Devore, J. L. (1991). *Probability and statistics for engineering and sciences.* Pacific Grove, CA: Brooks/Cole Publishing Company.

Elling, M. G. M. (1991). *Veiligheidsvoorschriften in de industrie (Safety rules in industry)*. Faculty of Philosophy and Social Sciences Publication WMW No.8. Netherlands, University of Twente.

Federal Railroad Administration. (1999). *Switching operations fatality analysis: Findings and recommendations of the SOFA working group*. Federal Railroad Administration: DOT-VNTSC-FRA-00-08, DOT/FRA/ORD-00/04.

Federal Railroad Administration. (2001). *Switching operations fatality analysis: Severe injuries to train and engine service employees: Data description and injury characteristics.* Report from the Switching Operations Fatality Analysis (SOFA) working group, Federal Railroad Administration.

Federal Railroad Administration. (2006). *Railroad safety statistics 2005 annual report.* Federal Railroad Administration.

Federal Railroad Administration. (2008). *The FRA Risk Reduction Program: a new approach for managing railroad safety.* White paper of Federal Railroad Administration Office of Research and Development Division, Human Factors Program, December. Retrieved from http://www.fra.dot.gov/downloads/safety/ANewApproachforManagingRRSafety.pdf.

Frederick, J., & Lessin, N. (2000). Blame the worker the rise of behavioral-based safety programs. *Multinational Monitor*, *21*(11).

Gamst, F. C. (1980). *The hoghead: an industrial ethnology of the locomotive engineer.* New York: Holt, Rinehart, and Winston.

Gamst, F. C. (1982). The development of operating rules. *Proceedings of the Railway Fuel and Operating Officers Association*, *46* 155, 162-171.

Guastello, S. J. (1993). Do we really know how well our occupational accident prevention programs work? *Safety Science*, *16*, 445-463.

Hale, A. (2000). Railway safety management: The challenge of the new millennium. *Safety Science Monitor 4*(1 Article 1), 1-15.

Hale, A. R., Guldenmund, F. W., van Loenhout, P. L. C. H., and Oh, J. I. H. (2008). Evaluating safety management and culture interventions to improve safety:

Effective intervention strategies and lessons learned. *Safety Science*, *48*(8), 1026-1035.

Harrington, J. J. (1987). *The improvement process: How America's leading companies improve quality.* New York: McGraw-Hill.

Hayes, B. E., Perander, J., Smecko, T., & Trask, J. (1998). Measuring perceptions of workplace safety: Development and validation of the work safety scale. *Journal of Safety Research*, *29*, 145-161.

Hays, W. L. 1981. *Statistics* (3rd ed.) New York: CBS College Publishing.

Hofmann, D. A., & Stetzer, A. (1996). A cross-level investigation of factors influencing unsafe behaviors and accidents. *Personnel Psychology*, *49*(2), 307-339.

Howe, J. (1999). Debunking behavior-based safety. *Occupational Health and Safety - Newsletter of the UAW Health and Safety Department*, 1, p5.

Hume, R. (2003) Rail workers battle unsafe remote control technology. *Labor Notes,* July 31.

Juran, J. M. (1964). *Managerial breakthrough: A new concept of the manager's job.* New York: McGraw-Hill.

Lee, E. T. & Wang, J. (2003). *Statistical methods for survival data analysis*. Hoboken, NJ: John Wiley & Sons.

Kopelman, R. E. (1986). *Managing productivity in organizations*. New York: McGraw-Hill.

Krause, T. R. (1995). *Employee-driven systems for safe behavior.* New York: Van Norstrand Reinhold.

Krause, T. R. (1997). *The behavior-based safety process: Managing involvement for an injury-free culture*. New York: Van Nostrand Reinhold, p74.

Krause, T. R., Seymour, K. J., & Sloat, K. C. M. (1999). Long-term evaluation of a behavior-based method for improving safety performance: a meta-analysis of 73 interrupted time-series replications. *Safety Science, 32*(1), 1-18.

Lewin, K. (1947). Group decision and social change. In T. N. Newcomb and E. L. Hartley (eds.) *Readings in Social Psychology.* Troy, MO: Holt, Rinehart, and Winston.

Maguire, B.A., Pearson, E.S., and Wynn, A.H.A. 1952. The time intervals between industrial accidents. *Biometrika*, 39(1-2), 168-180.

McPhee, J. (2005). Coal train—I: Disassembling the planet for Powder River coal. *New Yorker*, October 3.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

National Transportation Safety Board. (2004). *Railroad accident brief (Accident No. DCA-04-FR-004).* Washington DC: National Transportation Safety Board.

National Transportation Safety Board. (2005a). *Collision of Norfolk Southern freight train 192 with standing Norfolk Southern local train P22 with subsequent hazardous materials release at Graniteville, South Carolina January 6, 2005*. NTSB/RAR-05/04, Washington DC: National Transportation Safety Board.

National Transportation Safety Board. (2005b). *Railroad accident brief (Accident No. DCA-05-FR-004)*. Washington DC: National Transportation Safety Board.

National Transportation Safety Board. (2006). *Collision of Union Pacific Railroad train MHOTU-23 with BNSF Railway Company train MEAP-TUL-126-D with subsequent derailment and hazardous materials release Macdona, Texas June 28, 2004*. NTSB/RAR-06/03, Washington DC: National Transportation Safety Board.

Nelson, W. B. (1982). *Applied life data analysis*. Hoboken, NJ: John Wiley & Sons.

Nelson W. B. (2003). *Recurrent events data analysis for product repairs, disease recurrences, and other applications.* Philadelphia: Society for Industrial and Applied Mathematics.

O'Toole, M. (2002). The relationship between employees' perceptions of safety and organizational culture. *Journal of Safety Research*, *33*, p231-243.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text.* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods.* (3rd ed.). Thousand Oaks, CA: Sage.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. (2nd ed.). Fort Worth, TX: Harcourt College Pub.

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies*. Princeton, NJ: Princeton University Press.

Pierce, D. R. (2005). Portland Yard Sale/ UTU Misinformation. Letter to BLET local chairmen, BNSF Northlines/ MRL, August 8.

Pilcher, J., & Coplen, M. (2000). Work/rest cycles in railroad operations: Effects of shorter than 24-hour shift work schedules and on-call schedules on sleep. *Ergonomics*, *43*(5), 573-588.

Railway Age. (2006). One-person crews? No, say UTU, BLET. *Railway Age* February, 9.

Ranney, J., & Nelson, C. (2003). *The impact of participatory safety rules revision in the U.S. railroad industry: An exploratory analysis*. Federal Railroad Administration, DOT-VNTSC-FRA-02-05.

Ranney, J., & Nelson, C. (2007). *The impact of participatory safety rules revision on incident rates, liability claims, and safety culture in the U.S. railroad industry*. Federal Railroad Administration, DOT-VNTSC-FRA-02-05.

Ranney, J., Zuschlag M., Coplen, M., & Nelson, C. (in preparation). Behavior-based safety at Amtrak-Chicago, associated with improved safety culture and reduced injuries and costs. Manuscript submitted for publication (copy on file with author).

Ranney, J., Zuschlag, M., Coplen, M. & Wu, S. (2009). *Safe practices, operating rule compliance and derailment rates improve at Union Pacific Yards with STEEL Process – a risk reduction approach to safety*. Federal Railroad Administration Research Results DOT/FRA/RR09-08.

Reason, J. (1997). *Managing the risks of organizational accidents*. Aldershot, England: Ashgate.

Ricci, M. K. (2003). Evaluating behavior based safety efforts in the railroad industry: work practice observation (WPO): A pilot study. Final Report, Washington State Utilities and Transport Commission, Summer, WA.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Schein, E. H. (1992). *Organizational culture and leadership.* (2nd ed.). San Francisco: Jossey-Bass Publishers.

Scriven, M. (1991). *The evaluation thesaurus.* (4th ed.). Newbury Park, CA: Sage Publications.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference*. Boston: Houghton Mifflin.

Simard, M., & Marchand, A. (1994). The behaviour of first-line supervisors in accident prevention and effectiveness in occupational safety. *Safety Science 17*, 169-185.

Simard, M., & Marchand, A. (1997). Workgroups' propensity to comply with safety rules: the influence of macro-micro organizational factors. *Ergonomics*, *40*(2), 172-188.

Spigener, J. B., & Hodson, S. (1997). Are labor unions in danger of losing their leadership position in safety? Their resistance to behavior-based safety makes us wonder. *Professional Safety* (December), p. 37.

Stufflebeam, D.L., (1983). The CIPP model for program evaluation. In: Madaus, G., Scriven, M. and Stufflebeam, D. (eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston, MA: Kluwer-Nijhoff Publishing, 117-141.

Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, *22*(2), 183-209.

Sulzer-Azaroff, B., & Austin, J. (2000). Does BBS work? Behavior-based safety and injury reduction. *Professional Safety*, 19-24.

U.S. Government Accountability Office. (2009 Nov.). *Program evaluation: A variety of rigorous methods can help identify effective interventions.* Washington, DC: U.S. Government Accountability Office. Retrieved from http://www.gao.gov/new.items/d1030.pdf.

Walton, M. (1986). *The Deming management method*. New York: Perigee Books.

W. K. Kellogg Foundation. (2004). Logic model development guide. W. K. Kellogg Foundation. Retrieved from http://www.wkkf.org/default.aspx?tabid=101&CID=281&CatID=281&ItemID=2813669&NID=20&LanguageID=0

Weber, M. (1991). The social psychology of the world religions, In H. H. Gerth and C. W. Mills (eds.), *From Max Weber: Essays in sociology.* London: Routledge, 267–301, 280.

Yamaguchi, K. (1991). *Event history analysis*. Newbury Park: Sage.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users.* (3rd ed.). Thousand Oaks, CA: Sage.

Yin, R. K. (2003). *Case study research: Design and methods*. (3rd ed.). Thousand Oaks, CA: Sage.

Appendix A: Changing At-Risk Behavior (CAB) SASU: Individual Stories of Change

A.1 Executive Summary

This report[14] summarizes individual stories of change related to safety that occurred as a result of the CAB program. Formal and informal discussions were held over a 5-day time frame in the San Antonio, Laredo and Bastrop locations within the San Antonio Service Unit (SASU) of the Southern Region of the Union Pacific (UP) railroad. Engineers, conductors, and managers at these locations were asked about safety and how their attitudes and behaviors have changed as a result of the CAB process. The themes that emerged from those discussions are described, categorized, and summarized in this report.

Respondents suggest that the CAB activities have positive effects. Classes provided valuable safety refreshers by "keeping safety more on the front of the mind" and one person reported a shift in behavior as a result of attending a CAB class. Third-party CAB sampling was effective at changing behaviors and some stories highlight the value of having program participants conduct observations with their peers.

Workers and managers both described changes in practice. Workers report more rule compliance. They also report answering their cellular phones less in CRZ situations and discussing safety with their colleagues more. Managers report shifting toward a safety leadership style that is more cooperative and less authoritarian.

CAB is helping to develop a culture of safety. This is reflected in:

- Better relationships among workers, and between workers and managers.
- In more safety-related dialogue and communication.
- Internalization of CAB lessons and applying them to other situations.

CAB is also creating a culture where operating safely is better than operating expeditiously.

Workers are paying more attention to risk, creating safer workspaces, and communicating in new ways about safety. This is different from the atmosphere before CAB where "you were actually just across the cab but seemed like you were 2–3 miles apart."

Managers are changing their safety leadership style for the better by providing leadership that is less authoritarian and takes more of a coaching approach. Their respect of the CAB program keeps them out of the way of workers doing their job safely.

---

[14] This report would not have been possible without the thoughtful and thorough support of Michael Byars and Kelvin Phillips, SASU CAB Facilitators who organized the interviews.

A.2 Objectives

Toward the conclusion of the evaluation period of CAB, anecdotal accounts suggested that changes in approaches to safety were occurring at the *individual* level. Other interviews conducted as part of the CAB evaluation focused on program implementation and outcomes, not individual change. Because a qualitative perspective could generate a picture of this kind of change, the CAB evaluation included data collection and analysis to learn as much as possible about how the CAB program had changed safety attitudes and safety practices for individuals.

Since the demonstration pilot with road workers was implemented in a dispersed work setting there was minimal visibility into the change process because it took place on the road away from operations and management. Individual detailed stories about changes occurring "out there" were needed to provide a more detailed evaluation picture.

A.3 Method

Engineers, conductors, and managers were selected to participate in the interviews through purposive and snowball sampling (Crano and Brewer, 2002) (not to be confused with CAB "sampling," i.e., CSA feedback sessions). CAB facilitators in SASU purposefully identified trainmen and managers who they recognized as having experienced a change relating to safety over the last few years. During the interviews, the interviewer also asked participants if they could identify someone else who had experienced change (snowball sampling). These individuals were subsequently contacted and interviewed. The objective in selecting the interviewees was to ensure a roughly equal number of engineers and conductors and an equal numbers of senior (15+ yr) and junior (5–10 yr) workers. Eight engineers, six conductors, and two managers participated (Table 1).

Table A1. Respondents and Years of Experience in Current Position

Respondents	No.	No. of Yrs in Job
Engineers	8	4 ~ 38
Conductors	6	2 ~ 40
Managers	2	10 ~ 15

The semistructured interviews with an interview guide (Patton, 1987, 2002) obtained the stories of change. This technique is useful in situations where natural conversation might be more fruitful than direct interviewing (Patton, 1987). An interview guide consists of an *a priori* set of questions or topics that serve to guide the interviewer but do not direct the order or selection of questions asked. This ensures the flexibility to create a guided discussion. Rapport is quickly built in a semistructured approach and the guide helps the discussant remain topically focused.

The questions focused on the worker's "personal experience with safety" and how on- and off-the-job experience might have changed over the last few years. Especially important was whether or not the worker had made specific changes in behavior (i.e., "becoming more safe") and if they could relate that change to CAB. Discussions with the managers followed a similar protocol but focused on decisions they made as managers and how safety played into those decisions.

Interviewers audio-recorded and loaded these discussions into a qualitative data analysis software package (Atlas.ti). Audio files were coded for themes and meanings (Miles & Huberman, 1994; Patton 2002).

A.4 Findings

Themes identified in the discussions are identified using the CSA logic model discussed in Evaluation Plan on page 33 (see Figure 1).

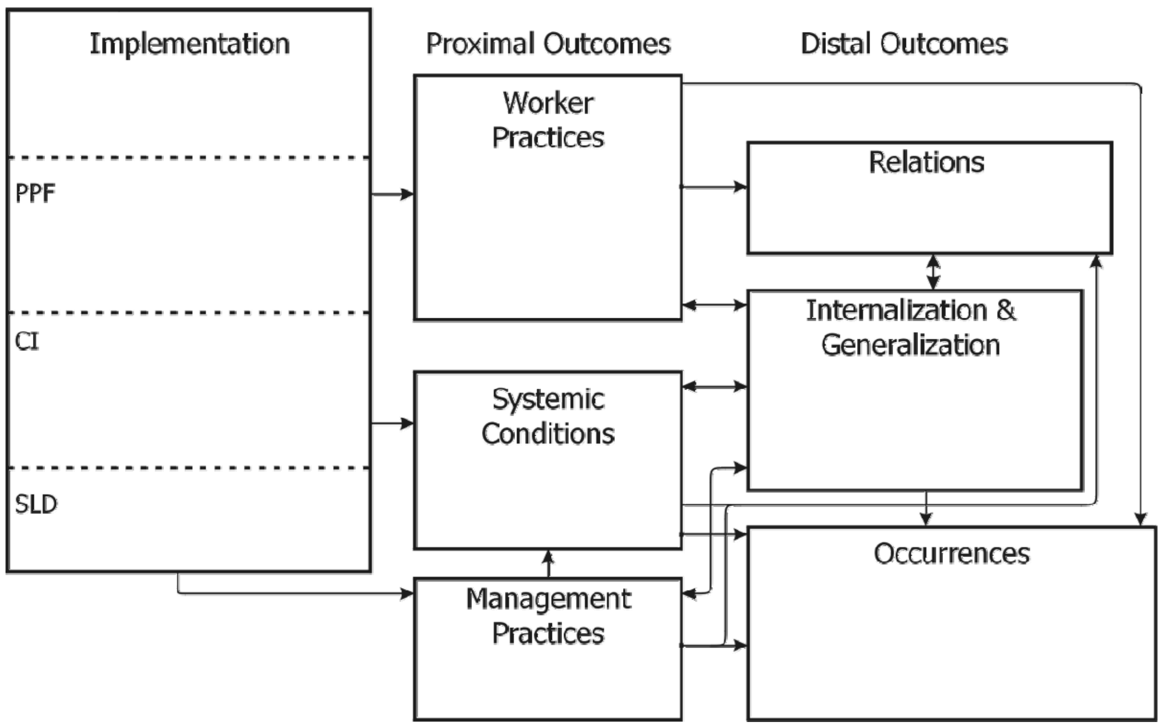


Figure 1. Logic Model for CAB

The logic model is designed with a left-to-right orientation so that the implementation of CAB leads to proximal outcomes, which in turn leads to distal outcomes. The arrows in the model indicate the effect each prior factor has on subsequent factors in the model. In this left-to-right orientation the right two columns are outcomes put in motion by the implementation in the left column. Because they occur as a direct result of the implementation, the middle column outcomes are proximal. In the right column, distal outcomes are mediated by the proximal outcomes. In some cases, these distal outcomes affect change on proximal outcomes as a feedback loop (see arrows between boxes).

Proximal outcomes include safer worker and manager practices and systemic improvements. Distal outcome include the nature of relations, internalization, and generalization of the CAB principles, and reduced occurrences. The internalization and generalization of CAB lessons has feedback relationships with all three proximal outcomes and culture/relations.

Themes identified in the interviews are organized using this logic model. Each section of this report begins with a description of the factor followed by supporting quotes and interpretations where necessary. Words in italics in the quotes represent emphasis in the respondent's tone of voice. Interviewer questions are identified by parentheses in the quotes.

A.5 Implementation

Implementation is defined as all the tasks and activities of CAB and is displayed on the left side. Implementation is the catalyst for all the activities and proposed outcomes in the right two columns. Some people described a change in their safety-related behavior and said they could attribute that change to either CAB Classes or CAB Sampling (feedback sessions) activities (see Figure 2).

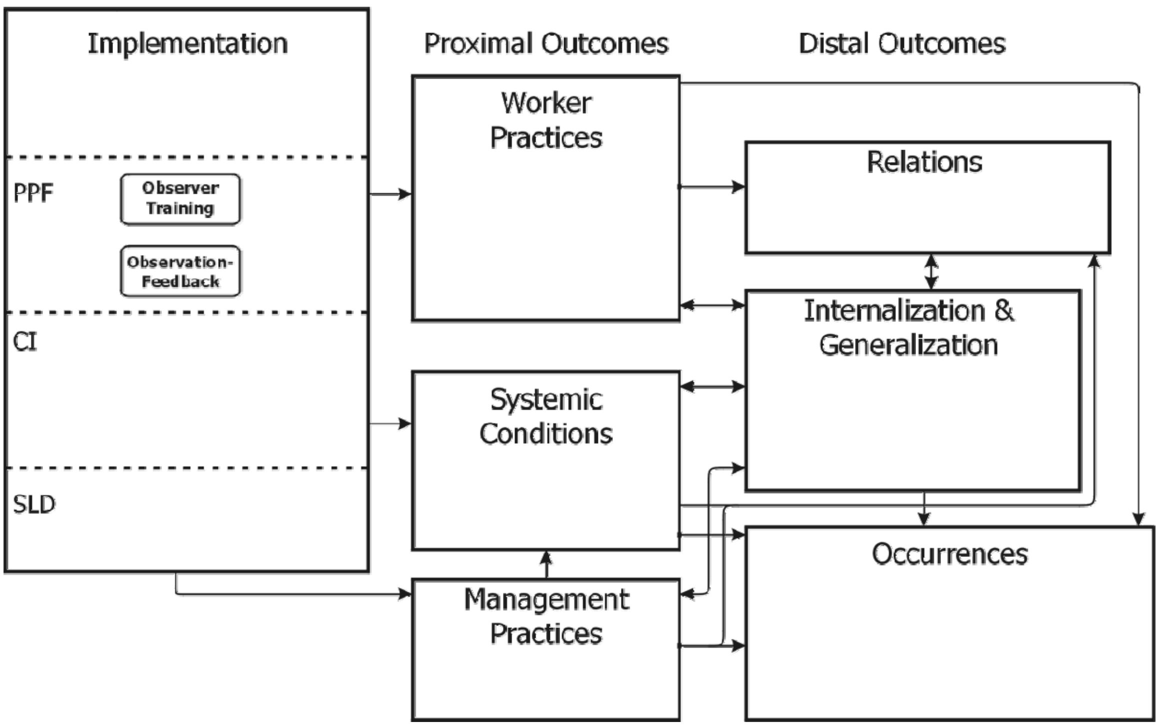


Figure 2. Logic Model, Implementation Factors in Change

A.5.1 CAB Training Class

A senior engineer attributed his personal change towards safety to the CAB training classes for conducting feedback sessions. He said the class was a turning point for him and suggested the class helped to refine his own awareness of risk.

> *Engineer – 34 yrs*
>
> I think it was immediately after the class [that I started being more aware]…I had asked a friend of mine about the CAB process and he said, "I went through it and it just teaches you about at-risk behavior." And I said "Whoa okay, explain." And he did. And I said, "Okay, I'll give it a good shot and sit through it." And, it was immediately after the class, I, you know, I thought, "this is a good thing, this is a good tool." It just fine-tunes what you already knew, what most common sense people already know, uh, those of us out

there who are safety conscious to begin with. Of course before the CAB class it was just if you got hurt, well you know…that's a bad thing, now it's a *really* bad thing. [13:3]

Managers also saw this change. One manager, with more than 30 years of experience on the railroad (14 of them as a manager) believed the classes were affecting "old heads"—the long-time workers.

Manager – 14 yrs

The guys are not scared to confront the old heads or the more experienced guys. And, the way I've always preached to everybody is that you know, you have to [confront them]. The young guys may know the rules, they may not know the application of the rule, but they, a lot of times, know the rules better than the old heads per se. Nowadays it's getting a little less like that because we're having rules classes every year. And, these CAB classes are really driving the old heads into believing "I know I've done this for 30 years, but you know, maybe I was doing it wrong, or maybe it is unsafe to do it." [10:3]

Most people did not pinpoint their change to a specific component of CAB. They talked about "acting safe" and that the classes simply added to their knowledge. When some mentioned the CAB class, they saw it as part of a slow layering of safety-related knowledge—e.g., the more times they hear about safety and see images of accidents, the more their "already safe" behavior increases.

A.5.2 Observation and Feedback

Peer-to-peer observation and feedback activities of CAB, referred to by CAB as "sampling," influenced people differently. One engineer said he now calls out his speed after being "caught" not doing this during a CAB feedback session. Another relatively junior engineer used self-sampling "for a while" to help him identify where his own personal safety-related behavior could use improvement. A few workers mentioned affecting others by their efforts.

An example of how one worker can impact another is where an engineer with about 15 years of experience facilitates change in a 30-plus year worker by pointing out the more senior worker's risky behavior. Although the engineer's description does not make it clear what kind of sampling he is doing, the context of the conversation containing this statement implies that he was sampling.

Engineer – 16 yrs

Just the other day I had an old head, 30 years plus (railroad experience), and I was like, "hey man…you got off the van, [and] when you were standing next to the locomotive you started putting on your glasses, you started putting in your plugs." Cause, you know he went to the rear of the van to get his PPE [but didn't put them on till he was standing by the locomotive]…and I go "What were you thinking". [And he said] "Well I know I gotta do it." And I go "Well you *gotta* do it man….Do you have grandchildren?" And he goes "You know I always talk about so-and-so I always brag about [them]." And I go "Exactly! Isn't the best thing in the world you ever heard was 'I love you dad' or 'I love you grandpa' or anything like that?" [He says] "Yea." [I say] "Well you ain't gonna hear that if you don't wear those plugs. Don't you wanna see them grow up?" [He says] "Yeah, of course I do." [I say] "Well you ain't gonna see them if you don't wear those glasses. I'm not telling you to wear them. I'm not a *manager* or anything like that. I'm here to help *you* out. You know what I mean?" I go, "You can't be anti-railroad, okay, when you

> work, 'the hell with the PPE'. You gotta realize it's for your benefit. We [samplers] are out here to make it safer for you." And he was like, "Okay, yeah, I see where you're coming from." So, do I see him now? Yeah, he even has those little things where his glasses are hanging [around his neck] and stuff like that. And he'll just wink at me like "I got em!" [8:3]

The above quote underscores how program participants often have the most program-based credibility. The quoted engineer knew the other worker personally and leveraged this relationship to convince him to change his behavior.

A.6 Proximal Outcomes – Worker Practices

Worker practices are normal workday behaviors that were targeted for change by the CAB process. Effects are proximal outcomes because they are impacts of implementation activities. Worker practices also affect distal outcomes (i.e., Culture/Relations, Internalization and Generalization, and Occurrences). The discussions for this report found workers and managers describing rule compliance more frequently (see Figure 3).

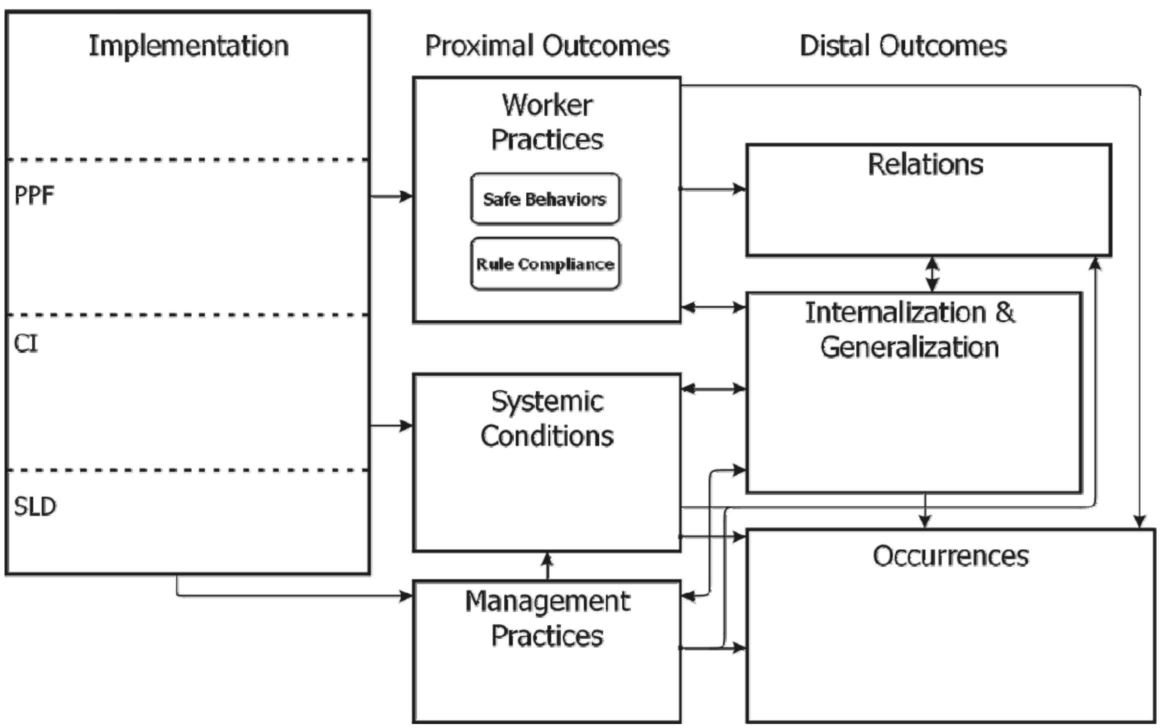


Figure 3. Logic Model, Worker Practices Outcomes

A.6.1 Rule Compliance

Workers knew about safety rules before CAB was implemented but they might not have followed them all the time. With CAB, rule compliance is more acceptable and is expected by both workers and managers who are more likely to follow the rules. A few workers talked about not answering their cell phone during a CRZ and a senior conductor reported calling out signals in a CRZ now more frequently than before CAB.

The following quote by an engineer with 9 years of experience illustrates this positive shift in rule compliance.

> *Engineer – 9 yrs*
>
> And of course, we all have to pass the rules exam. Everybody passes the rules exam. I mean we all know the rules, but do we follow the rules? …But, here these past few years we're starting to follow the rules. In the past 7–8 years it was never talked about. They didn't talk safety, they didn't preach safety. They preached: "we'll fire you when you break the rules, you'll get in trouble if you break the rules." Before this [CAB] it was, "we'll discipline you if you break the rules", so we all had to know the rules. [We] still know the rules, but now we follow them. [13:6]

This engineer describes a shift in perspective from focusing on *rule compliance* to focusing on *safety*. This same engineer talks about not answering his cell phone during critical times.

> *Engineer – 9 yrs*
>
> I think one of my habits now, you know, I don't pick up the phone when I'm in critical places, downtown cities. When I'm out in the country, oh sure I'll talk on the phone, what, I'm around cows? There might be a farmer once in a while. In major cities, I don't want to hear my phone. [13:5]

Although not answering cell phones has not been a long-standing rule it was a new rule brought in with the modified CRZ rules (Union Pacific, 2005). Several people implied this had always been a rule but that they were just now taking it more seriously.

One of the first items on the CAB data sheet for behaviors in a CRZ is "does the crew call out and acknowledge the signal aspect while in a CRZ." A senior conductor tells of this required communication as a catalyst for further change. He implies the CRZ communication is a direct result of CAB and that the CAB program requires this of him. His quote provides evidence that the rule might be followed more because of CAB.

> *Conductor – 40 yrs*
>
> In the old days… before CAB came around, you know the conductor's sitting on his side [of the cab] and he just did his thing. You know, maybe get occupied with whatever "schlitz list" or whatever he needed to do, look over his paperwork. And the engineer took care of his deal over there. So, you were actually just across the cab but seemed like you were 2–3 miles apart. You didn't know what he was doing. He didn't know what I was doing. CAB comes along, and this new process… at first it was hard—you know calling out signals, getting occupied with your red zones. It was all familiar territory. But where we were not very comfortable. Once we got comfortable with the program, with what we needed to do like calling out your signals…we were more alert. That's what it made me. [CAB] made me more conscious of my responsibility and also kinda influencing the engineer's responsibilities. Ah, he might have had a bad day yesterday. He might be thinking of something else. But, because of that red zone situation training, we seem to be more alert to our surroundings and situations, especially around red zones. [6:3]

CAB sampling and observation provided the impetus for the senior conductor to be more vocal in the cab (as expected under CRZ rules). This cross-cab dialogue makes him (and by extension, the entire crew) more alert.

A.7 Proximal Outcomes – Management Practices

Managers are changing their leadership style as a result of CAB. Following the trend with worker practices these changes are proximal outcomes resulting from CAB implementation. Changes in management practices also influence the distal outcomes. Respondents' changes in management practices are categorized as safety leadership (see Figure 4).

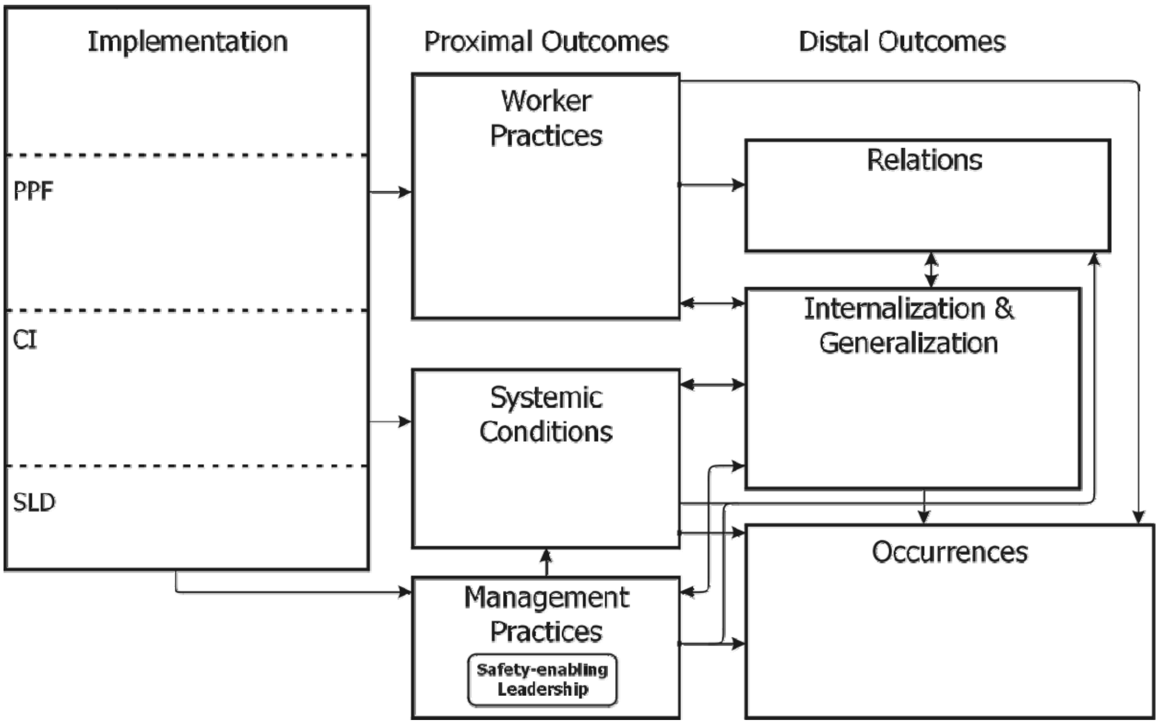


Figure 4. Logic Model, Management Practices Outcomes

A.7.1 Safety Leadership

Managers are moving towards more of a cooperative and less of an authoritarian leadership style. One manager describes his coach-like style of leadership.

> *Manager – 14 yrs*
>
> I try to manage [in a way so that] my guys can come in and tell me if there's an issue and I try to handle that issue. And, if I have an issue I tell them about it, and let them handle the issue. [10:10]

The coaching style of leadership is generating a positive response from employees. In addition a more cooperative labor–management environment is developing as a result. The manager in the above example implies a relationship between employees' willingness to raise risk-related issues and the leadership style of the manager—the more frequently managers ask employees for input, the more frequently employees are speaking up about potential safety-issues.

Manager – 14 yrs

> I see a whole lot more people coming in and saying "hey, you may take a look here and you may take a look there, and this might be happening, and that might be happening." …Before [CAB] they didn't want to have anything to do with it.…[Now] I think… a little more respect comes from the management towards the employee than there used to. [It] used…to be…"You work for me so do what I tell you to do." And, now it's "you work for me, what do you think will be the best way to get this done?" [10:9]

The manager involves workers in safety decisions and the employees feel more comfortable speaking up and engaging him. The manager is establishing an atmosphere for good safety behavior.

Another manager explains what it means to establish this atmosphere.

Manager – 10 yrs

> And it takes somebody actually doing the right thing so that it's known that it's okay to do the right thing. [16:2]

Managers also respect workers' commitment to safety more because they see that CAB training is making workers safer. One manager with about 14 years of experience put it this way.

Manager – 14 yrs

> It doesn't have to be as hands-on as it used to. You know, where if I told someone to go set out some cars in a certain track, before, I would go and make sure they did it safely. And now, I feel like I can delegate and say "hey go set these cars out over there" and I know, I almost know they're going to do it safely. …I definitely think that the training they've got with the CAB process and the mindset the CAB process preaches has made some people, who used to not think about it, think about it now. [10:2]

This manager has adjusted his perspective. Whereas in the past he might have felt he had to check up on his workers, he now feels that the workers are making safer choices. He attributes this change to the CAB process. Workers are receiving more beneficial training and the manager worries less about their safety. As a result, he gives them autonomy to do their work.

One engineer recognizes managers are more respectful of workers' focus on safety.

Engineer – 4 yrs

> [W]e have to have a job briefing to see what it is we're going to do. How we're going to tackle it. Well, by the time we get to the yard, they [our orders] might change. The manager's point of view was that I want you to go out there and just switch as much as you can, or make up these trains as fast as you can. You know, and that's the way they were used to it. We're going to go out there and do it fast. I think that now, like I said, with the change [since CAB], they [management] have come to accept, the majority of them, they have come to accept that we are gonna follow the rules. We're going to look out for ourselves. I think at the same time we're not only looking out for ourselves, but we're also looking out for the company. We don't want to get hurt. We don't want to cause no incidents. [6:13]

He recognizes that his managers have changed (since CAB) in that they have come to accept that workers will look out for themselves.

Another manager was asked for an example of a decision where safety and expedience might be the two choices. He mentioned "shoving cars" down a track for loading. The yard's layout gives the option of either going head-in to drop cars or backing them into a track. The most expedient method may be to back the train in and drop the cars so that the locomotive can pull out directly. Backing a train is not necessarily the safest way—the conductor must ride on the rear of the train and communicate with the engineer over the radio. When asked how his decision to move trains might have changed since CAB had been instituted[15], the manager thought the workers would be more vocal in their desire to do it more safely and might question his decision. Whether or not this pushback from a worker actually occurs is not the key point. The key point is that the manager said the belief that he *might* get pushback would impact his decision on how to move trains around the yard. The manager's decision-making was influenced by what he personally considered to be relevant, not necessarily the actions of the workers. This way of thinking is new and postdates CAB implementation.

A.8 Distal Outcomes – Relations

A culture of safety is a broad amalgamation of factors that represents an organization's commitment, style, and effectiveness in handling safety, including safety-related norms, competencies, values, attitudes, relations, and interaction patterns (Reason, 1997). As distal outcomes, culture indicators are products of practices and systemic conditions. In the logic model, culture is divided into reciprocally-related relations, and internalization and generalization. As the culture becomes more safety-oriented, people internalize safety and then create a stronger culture of safety. Respondents described more cooperation in the locomotive cab and improved relations between labor and management (see Figure 5).

[15] This was a junior manager with no experience in the railroads before CAB to compare to, so the statement must be taken with some caution. It is not clear if this kind of pushback would be a result of CAB. But, there is sufficient evidence from workers' comments to support the validity of this comment.

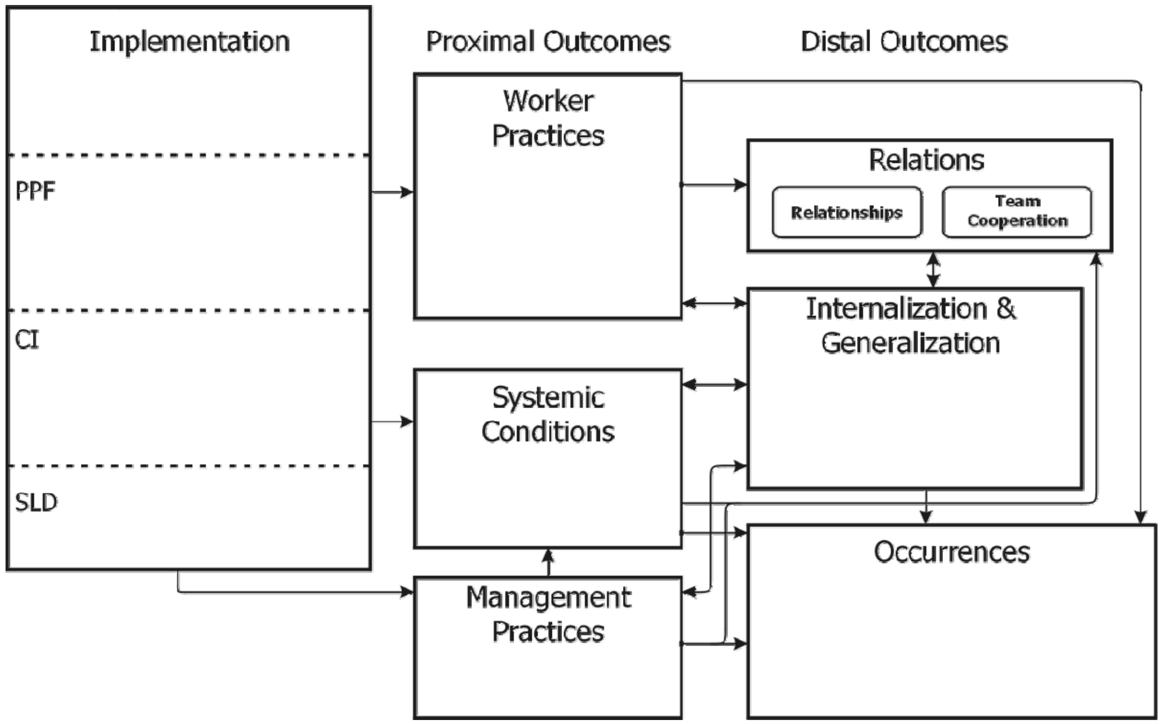


Figure 5. Logic Model, Relations Outcomes

A.8.1 Team Cooperation in the Cab

For some people, the in-cab experience since the start of CAB seems different with more of a team atmosphere in the cab. There is also more communication between the engineer and conductors and this has brought about a sense that they are "in this together."

> *Engineer – 12 yrs*
>
> Just communications over the slows that are out there. Obstructions, limits. You know, where our limits are—crossings that are not operable, mandatory directives that they give us over the radio.… Before it kinda was, "Well if you got it you got it" you know what I mean? But now it's like "did you get that, well here's where it is."…So, it's put communication back, and responsibility back on both crewmembers. And so, in that turn, you've got a guy watching out for you and you're also watching out for him. It's not so...individualized. [11:1]

This engineer says that past communication was minimal and that safety was left in the hands of the person who was directly responsible for the action. When directives or signals appeared, the conductor would *expect* the engineer to have seen it – and would not necessarily follow up. Since the start of CAB, both share the responsibility of making sure the other person is aware of the situation at hand and is taking the appropriate action. They are also both responsible for the train. This produces a feeling of shared responsibility and cooperation.

A senior conductor talked about how improved communication in the cab is making him feel like they are operating in a safer manner.

Conductor – 40 yrs

> We're approaching an approach signal. I'll look at it. Sometimes we're in a curve or something maybe he'll see it first. But most of the time somebody is going to call it out. Sometimes he'll call it out ahead of time, you know "approach" and you'll kinda [respond] "Oh? Okay, let's get it." You know, that kind of communication. Before, two people riding in the same cab [might] look at a signal. I looked at it. He looked at it. And yet I was thinking, "He's running, he should've looked at it." But, not anymore. That's what I like about this. I…feel more comfortable knowing that my train is now under control because we are communicating. It's no longer *he* is in control, but now it's *we* are in control [6:4]

A.8.2 Improved Labor–Management Relationships

Relationships between workers and managers are improving. Managers are helping create an atmosphere where safety is taken more seriously. As a result, workers are empowered to cooperate more with management and more respect is reported between the two groups.

Conductor – 40 yrs

> At first I was very, very belligerent, because you know…the FTX. I somehow didn't trust the [CAB] program. I never did trust them. Because that was "management." But, now that I see that management has their job, it's a line, they gotta do so many FTX's a month, they gotta do this, they gotta do that. But, now with this program I'm more prepared to say, well okay, you wanna come over onto my side of the line. I'm prepared now, it doesn't bother me now. [6:6]

This conductor understands and respects management's role in safety now that he has a safety program that belongs to workers. Because he was doing "his part" by participating in CAB, he understood that a manager also had to do "*his* part." CAB has legitimacy with the workers because it is not "owned" by management. They feel "safe" to talk openly about safety issues with other workers without fear they will be "written up." This even extends to their willingness to have frank conversations with managers about safety, as seen in an example [10:5] discussed later in Safety-Related Dialogue and Communication on page 176 [16].

Managers are shifting their leadership style and taking more interest in worker development. This has been a top-down process with upper management leading the way. One manager says that it had to be a top-down process because of prevalent feelings.

Manager – 10 yrs

> We need to be there for our employees. It's not our job to fire anybody, or to send anybody home. It's our job to help our employees. And that started from the top on down to the bottom management.... And, I think it had to start from the top and filter down. That's such a huge change on our structure as managers because it wasn't like that before. And now, … we really do care about our employees. [16:1]

---

[16] This example may be related to the particular manager. There is not the impression from these discussions that all managers are 1) approachable or 2) engaged with this program.

As an example of safety leadership, feedback from this manager indicates a change of perspective with managers taking more interest in their workers' well-being.

Workers feel they are now doing things more safely. Managers also are recognizing this, and also that workers are taking more interest in their employees. Both parties are more respectful of the other's role in safety.

A.9 Distal Outcomes – Internalization and Generalization

Internalization and generalization (I&G) is a broad category for a variety of outcomes from personal representations and expressions of organizational culture. For instance, lessons learned from CAB can be reflected in changes in working practices unanticipated by the intervention. I&G can also be a result of the CAB training class, from being sampled, or through on-the-job practices. Because I&G is personal, it is also an area where interviewers expected to hear most stories of personal change. If CAB was successful, workers will have internalized and generalized the lessons of CAB and may therefore have personal stories of change that include practices not specifically targeted by CAB. I&G might be illustrated as "I'm taking this on as my behavior and it has nothing directly to do with CAB."

I&G works reciprocally with the proximal outcomes. As workers and managers internalize CAB, there will be more of an influence from CAB on their practices and vice versa.

In addition, I&G promotes reciprocal relationships with workers and managers. They interact positively and by association have an impact on each other.

The goal of CAB–CRZ is to improve vigilance and attention to procedures under constraining signals. Whether or not the activity is CRZ-related, some respondents mention the broader impact from CAB, such as acting more safely at home and the increase in safety discussions at work. Internalization of the CAB principles enables people to understand how CAB works and to generalize its principles to many other domains (see Figure 6).

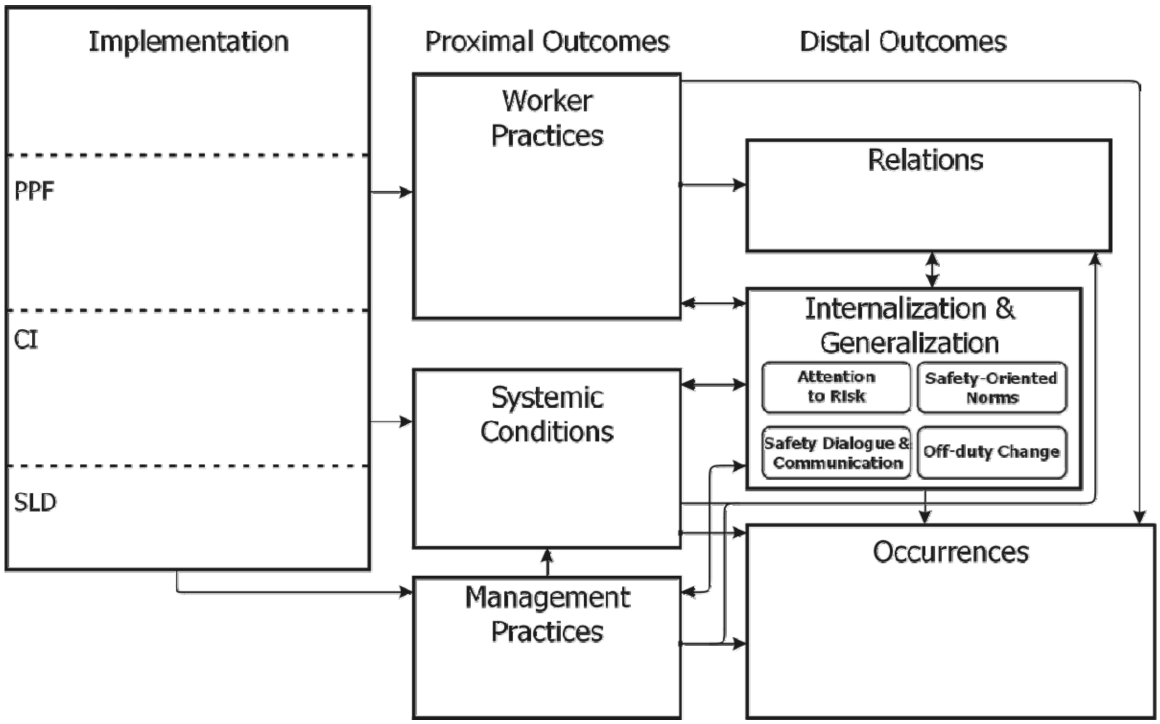


Figure 6. Logic Model, Internalization and Generalization Outcomes

A.9.1 Attention to Risk-Related Behavior and Conditions

Respondents report taking fewer risks and being more aware of risky conditions. Some examples of safe behaviors that are replacing risky behaviors include a conductor with 40 years of experience now carrying his own microphone in case a cab does not have one, passing on information about track conditions, and getting on and off the train in inclement weather.

In the following quote an engineer describes how his crew takes greater precautions in debarking their trains.

> *Engineer – 34 yrs*
>
> Many times during adverse weather conditions that we travel in, whether it be extreme desert in the deadly heat in Del Rio or blizzard cold icy conditions in Alpine. We cover it all sometimes in one single trip. And many times we've discussed about getting on and off with…all of our luggage and water bottles and things we have…to get on and off. And we'll discuss it before hand: "you step down, I'll hand this down, we'll use this side…look at the step, look at the slag." And, really take a serious conscious effort at where we're going to get off and what location. And even so far as tying up some of these trains in side tracks, especially around Alpine, Altuna, Lennox, 3 to 5 different tie up locations, we'll pick our spot. Not only for comfort for the van getting to us, but how are we going to get off here. [This is different than from before CAB when] just everybody not even talk about it. Sure, we were aware of it. But, there'd be many times where someone would slip, and drop something, [or] the conductor would get off way ahead of the engineer. Now it seems like there's more, now, especially the younger guys that'd gone through this class. Some of the older guys are resistant. But, the younger guys they're really conscious. [They'll say] "hey watch out, there's a cactus here, there's a big

> hole. When you get off, don't get off on the south side." You know, we pass that information along, and even so much as from one train to the next, one trip to the next. Somebody'll be in the depot and say "hey be sure when you stop in Lennox there's a washout". They'll pass this information for weeks at a time sometimes...[before CAB] they talked about it, but it definitely was not as prevalent as it is now. [13:1]

The engineer apparently could not identify the precise CAB component that stimulated this activity, but it is new since CAB implementation. The information-sharing and making choices about where to stop the train are evidence of workers generalizing CAB principles to other parts of their work. These behaviors are not part of CRZ, which CAB targeted for road workers. These behaviors are other parts of their work that are nonetheless influenced by a new internalized perspective on safety.

The following engineer's comments indicate an attention to risk that is being internalized and generalized to his everyday work. This engineer quickly points out that while he was always attentive, CAB has shown him that he needs to pay even more attention.

> *Engineer – 15 yrs*
>
> You start getting into a comfort zone after years of experience. And under the CAB process it reminds you… that all this time I've been doing this I haven't actually been attentive to the small details. Which is, like I said, crossing through tracks, or paying attention to the road when you're in Cab Red Zone. Especially, when you come up to an approach and you have to stop and you have a restriction up ahead…. Things happen because … you get too comfortable. And with the CAB process it has taught me that under Cab Red Zone you need to pay a little more attention. I'm not saying that you don't pay attention the whole time, but you need to pay *more* attention. [15:3]

When a 10-year veteran engineer appears to be taking on an environmental issue – missing conductor microphones –it shows how CAB is causing him to internalize his attention to risk.

> *Engineer – 4 yrs*
>
> One thing that I do like to point out is, as we're dealing with the CAB process, you're more aware of the little details that you really didn't pay attention to [previously]. For example… let's say we're in the cab and there was no microphone on the [conductor's] side….[T]he conductor used to get up and come over to the side of the engineer and he'll talk on the radio. At that point, as an engineer, I see that is in some ways actually minimizing my visibility of the track. It interferes with my visibility, at the same time I feel it is a safety issue because at any point the train can jerk and he can bounce off and he can go forward and fall down the stairs into the nose of the cab or he can fall on my side, causing a major accident….But, we've also seen the company being more responsive and providing replacement equipment when we report it. [6:9]

By focusing on safety in the cab through the CAB process, this engineer is more alert to other safety concerns. He has internalized CAB lessons and is generalizing them to other safety-related issues.

A.9.2 Safety-Related Dialogue and Communication

Safety-related dialogue and communication involves a broad array of communications throughout road and switching operations. When asked to describe safety-related change almost all respondents mentioned a change in the character or the amount of safety-related verbal communication. Talking about how to be safe is more acceptable than it was before CAB. Workers are no longer expected to figure out how to be safe by themselves.

The following comment from a senior engineer illustrates a culture where people are seeking out safety information from someone they know to be knowledgeable about the CAB sampling processes. This was not the norm before CAB.

> *Engineer – 16 yrs*
>
> Yeah, safety. The topic comes up a lot more now [since CAB]. You know, I'll sit in the siding and [they'll say] "Hey, I saw you [3rd party] sampling. What did you come across? What's the most popular thing you've seen?" And I'll tell them, "the most common thing is this, this, this. And this is what we need to change, this is what we're working on". Stuff like that. So, people know that I'm in "the process" so a lot of people approach me as far as questions and stuff like that. To see where we're going and what's it all about. [8:5]

In this next quote, an engineer with four years at the controls, but more than 10 years on the railroad, says it differently.

> *Conductor – 4 yrs*
>
> [W]e were so used to just…talk the minimum, the minimum possible. And now we understand that the more you talk, the more communicative you are, the more you are observant. [6:10]

One manager reflected on an incident where communication and dialogue in the crew area occurred in an unusual way:

> *Manager – 14 yrs*
>
> Well, I've had a couple of times where I don't think it would have happened before [CAB]…. But, I've had crews come in [to the terminal] and I mean the engineer and conductor are about to fight. The engineer's an old head and the conductor's a new guy that may be promoted. In one of the instances… they were arguing about the right way to do something. Where before it would have just been "hey you know you need to do this" and the old head would have said "I'm not doing it like that, this is the way I do it." And, I mean it turned in to probably not the best scenario because they were, you know, they were about to 'come to blows'. But, it all ended up getting defused and the engineer eventually agreed that you know, maybe that wasn't "the smartest thing for me to do" and I don't think that ever would have happened 3 years ago. I think 3 years ago they would have, either nothing would have been said to each other or they'd have went off the property and gotten in a fistfight. You know? And I attribute all of it to the CAB process because they had, the young guy had enough gumption to stand up to the old head and say this is the safe way to do it, irregardless of whether or not that's the case and had enough gumption to stand his ground because he knew that he had backing whether it be with the CAB class or me or the other managers that's out here. He knew that the most important thing was to do it safe. Period. And I don't think 3 years ago that was the case. [10:5]

This conductor was trying to respond safely, and the engineer, an old head, was being recalcitrant. Even though the dialog escalated into a loud affair, the conductor felt comfortable enough with his choice to bring it into the (public) crew room to discuss with a manager. This airing of concerns is not something that would have happened before CAB. Furthermore, this was a younger worker confronting an old head.

It is difficult to say what made this conductor feel that he could be successful in bringing his concerns to the manager in a public setting. Nevertheless, the manager makes it clear that this kind of open-door, public communication about safety is new since CAB.

This is a vivid example of worker empowerment. The conductor saw things were wrong and CAB may have given him a sense of control to make choices and decisions about safety. For him to not only say something to the older engineer, but to confront him in the face of pushback and to do it in a public space are all indications of a culture that supports empowering workers to make the best safety-related choices. In the end, the engineer conceded that he might have been wrong and that there "probably was a safer way." The result was a more informed engineer, a more empowered conductor, and a safer environment for both.

The discussion then continued with the manager:

> *Manager – 14 yrs*
>
> (It sounds like there's a venue for dialogue that has opened up?) A little bit. (And what I'm hearing is that 3 years ago it just wasn't there between engineers and conductors?) No, you didn't say nothing. [10:6]

While the evaluator was on an orientation ride in SASU, the engineer and conductor were observed talking about a safety-related issue. The engineer wondered why a certain marker along the rails was spaced differently in this yard compared to where he previously worked. The conductor asked why the engineer was just now bringing up this question now since he had been in this new location for a few years. The engineer's reply was he never felt there was a place for this kind of conversation before.

A.9.3 Safety-Oriented Norms

Safety-oriented norms are how people behave and how one is expected to behave in relation to safety in the workplace. Since discussions focused on individual stories of change related to safety, respondents provided few stories of safety-oriented norms.

One engineer did talk about the culture shift occurring across the yards.

> *Engineer – 16 yrs*
>
> The good thing I can say about what's going on right now is everybody right now is being more safe, or being more safety-conscious, if you will. It's not because they're afraid of the managers anymore. It's because they're saying "Hey, this is the way we're doing it with CAB. This is what I was taught in the CAB [class]. This is what we promote." You know? Before, it was being afraid of getting written up. Now it's just more, "Hey you know what? Let me do it *the right way*." Not being afraid of a manager coming after them and disciplining them. So, it's more what the program is instilling in their minds and what they're hearing in the field. [8:7]

A.9.4 Change in At-Risk Behaviors Off-Duty

Workers and managers are CABs as a result of CAB. Almost half of the people stated the lessons they are learning on the job are transferring to their lives off work. The following comments illustrate this change.

> *Manager – 15 yrs*
>
> I tell ya, the CAB class has made me start doing something that I never did before –I put safety glasses on when I weed eat or cut the grass. And, it's strictly just the CAB class that's done it. You know, and it doesn't have anything to do with the railroad but it's kinda the mindset that everybody's going to, you know. I gotta think about what I'm doing before I do it. Period. It doesn't matter if it's at work or at home or… driving down the road. [10:4]
>
> *Engineer – 15 yrs*
>
> Even at home, [I] go cut the grass…safety glasses. You know, I wouldn't think about that at home [before CAB]. At risk behavior, you know, you take chances, not to take chances. So, in that sense you know what? It's not worth it. I'd just rather put the glasses on and go out there and weed-eat. You know, never did that [before CAB]! [15:9]

A.10 Summary of Findings

Most of the respondents say CAB training classes and CAB sampling positively changed personal behavior. Classes are mostly seen as "more of a good thing" and are valuable in helping to keep safety on the minds of workers. The most compelling stories of change are from samplers about their interactions with other workers. This makes sense because the selection of class participants focused on those engaged in the CAB process as samplers.

As a result of CAB, safety practices for workers and managers appear to be improving. Responses from the interviews suggest that many workers are now more rule-compliant: there is less cell phone usage during CRZ and speeds and signals are called out more. Many managers are apparently seeing workers as more safety oriented and adjusting their leadership style towards more collaboration using safety-enabling approaches.

Many respondents mention that the safety culture has improved in the service unit and that relationships within the locomotive cab and between managers and workers are improving, especially around safety. Workers are problem-solving ways to avoid risks and discussing safety issues with their fellow workers more frequently, even when those discussions are contentious. Workers are also generalizing CAB knowledge to their off-duty activities like mowing the lawn and working around eye-hazards.

A theme not necessarily captured in the logic model is the program's peer-to-peer nature. Worker buy-in seems to be influenced by the fact that this is not a management program. Having another engineer or conductor out there looking over their shoulder gives the program more capital. An engineer with about 10 years experience said, "Guys out here don't seem to take it as personal when you're talking with them about safety." To the extent that the workers and the management know this, it can positively influence the buy-in and, in turn, their practice.

A.11 Strengths and Limitations

Since the evaluator for these interviews was not overly familiar with either the CAB program or the culture of the railroads in general, the inductive analysis is more valid than if a well-seasoned railroad person had performed the evaluation. Conscious or unconscious bias may have led participants to provide responses a more informed evaluator sought. In addition, and while the evaluator was sensitized to the interview topics, the semistructured interview and inductive analyses approach allowed the stories and analyses to evolve from the participants (Patton, 2002).

Given the important findings in this report, there are concerns over their interpretation. One concern is that the group of trainmen interviewed is small (n = 16). Each group was purposefully selected because they were thought to have experienced some sort of change in their safety behavior as a result of involvement in CAB. While these interviews indicate that there have been cases of personal change due to CAB, the breadth or amount of change throughout the entire SASU population cannot be estimated using this method. Even though there are some personal truths in the stories themselves, it should be remembered that these are also the stories of a "select," and not a representative, group of people. They are included in this report as "stories of success" (Brinkerhoff, 2003) of what change looks like for some who have benefited most from CAB. As such they should be recognized only as *individual* images of change not as stories that everyone experienced.

A.12 Conclusions and Implications

The semistructured interview discussion format provided valuable stories of individual change in relaxed, low-stress discussions. While limited in their generalizability, the discussions provide insight into changes in safety behavior for interviewed railroad workers. These findings suggest CAB is responsible for *positive change*.

CAB third party sampling has apparently been effective. The stories of transformation center primarily on samplers and describe the impact they had on someone else's practice. Committed samplers who "own" the process and have relationships with those they sample are probably more effective. This is not an unexpected outcome of a participatory process where individuals who run the program (volunteer samplers) become those most committed to making it effective (Cousins, 2003; Zimmerman, 2000).

Internalization and Generalization is central to the process of change for CAB. Communication as a mediator within this program should not be undersold. Most participants mention a change in communication as an outcome of CAB. People are talking about safety differently compared to how they communicated before CAB was implemented. CAB provides both a new venue for dialogue and the impetus for further communication around safety. Placing a peer "safety expert" in the locomotive cab or in the yards seems to have stimulated discussion of safety. This apparently has led crews to accept talking about safety as a smart thing to do. The dialogue is a catalyst for paying attention to risks and being more alert overall both in and out of the cab. The information being shared by these conversations gives workers the tools necessary to operate more safely. In this way, CAB implementation directly increases Internalization and

Generalization, rather than it being a distal outcome as depicted in the logic model on 164.

Workers expressed feelings of empowerment over their own safety. They reportedly feel more comfortable with their safety-related choices and are more likely to confront a manager who requires them to perform an unsafe procedure when a safer way is more feasible. This empowerment is supported by the fact that this is not a "management" program, yet it has support "up the chain of command." Empowerment is a key element in any participatory program where those taking some interest in making the program work develop a sense of ownership. Ownership promotes more internalization and generalization and subsequently expands the program's impact (Cousins, 2003; Cousins & Leithwood, 1986; Zimmerman, 2000). The peer-to-peer nature of this program provides credibility and seems to increase worker buy-in. If CAB had been a management-run program it probably would not have been given as much credibility among the workers.

For future demonstration projects this method of data collection and analyses provides valuable information. While safety culture was expected to change it was not clear what that change would look like. These discussions provide contextually valid stories of change (Creswell, 2003).

Participatory processes are fundamental to CAB and should be increased. Individual workers performing sampling and feedback activities have the most leverage with their fellow workers. Not only do they have the most contextual validity, they also become "evangelists" for the program because they have a vested interest in seeing it succeed.

Providing workers with opportunities to communicate amongst themselves about safety and the CAB process could capitalize on the effect of existing CAB activities. Communication among workers is already seen as a valuable outcome of CAB. Providing more organized opportunities to communicate about safety as a part of CAB may enhance this outcome.

A.13 References (Appendix A)

Brinkerhoff, R. O. (2003). *The success case method*. San Francisco: Berrett-Koehler Publishers.

Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation, Vol. 9. Part One: Perspectives, (pp 245–266). Dordrecht, The Netherlands: Kluwer Academic.*

Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research, 56,* 3(331–364).

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Creswell, J. W. (2003). *Research DESIGN: Qualitative, quantitative, and mixed method approaches.* (2nd ed.). Thousand Oaks, CA: Sage.

Good practice guide on cognitive and individual risk factors, RS/232, Issue 1 draft 1c (June 2007). London: Rail Safety and Standards Board.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (1987). *How to use qualitative methods in evaluation.* Newbury Park, CA: Sage.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Reason, J. (1997). *Managing the risks of organizational accidents.* Aldershot, England: Ashgate.

Zimmerman, M. A. (2000). Empowerment theory: Psychological, organizational and community levels of analysis. In J. Rappaport & E. Seidman (Eds.), *Handbook of Community Psychology* (pp.43–63). *New York: Kluwer Academic.*

Appendix B: Analysis Block Rates versus Gap Times

B.1 Abstract

Frequently evaluations of safety interventions use block rates for data, where each data point is the accident rate for a fixed block of time (e.g., injuries per 200,000 worker-hours for each month). Means or correlations are then calculated for the time blocks to assess if safety has changed over time or between groups. Gap times are a viable alternative to traditional block rates for the analysis of safety data. With gap times, each data point is the gap time or amount of risk exposure a site faces between two adjacent safety events, for example, the worker-hours between injuries. Unlike block rates, gap times avoid quantizing the data, lack arbitrary parameters that affect the results, and provide appropriate weights to the data points. The primary challenge with the analysis of gap times is adjusting for typically non-normal data distributions. However, this can also be an issue with block rates and for gap times can be handled through standard survival analysis techniques.

B.2 Program Evaluation and Safety Events

Numerous challenges exist in conducting a program evaluation of an intervention's impact on bottom-line safety performance. The baseline rates of safety events, such as accidents or injuries, tend to be relatively low, resulting in sparse data. To minimize both costs and safety risks associated with an experimental intervention, the intervention is often tried on a relatively small scale at one or a few pilot sites, further restricting the amount of available data. Safety-event data is inherently highly random (Reason, 1997) resulting in large variances that may thwart establishment of a statistically significant result.

Nonetheless, bottom-line safety performance is the ultimate variable of interest, thus evaluators are obligated to assess the impact of the safety intervention on safety events. Ethically, evaluators must determine the safety impact of an intervention as soon as possible and thus with as little data as possible. If the intervention is found to improve safety, then it may be expanded to prevent safety events beyond the pilot sites. If the intervention actually degrades safety, then it needs to be shut down promptly before it increases risk. Program evaluators thus need the most statistically powerful data-analysis techniques possible for handling safety-event data.

This paper reviews the common "block rate" approach of analyzing safety data, and illustrates its statistical weaknesses for program evaluation. The paper suggests gap times (Cook and Lawless, 2010) as an alternative and describes how it avoids these weaknesses.

B.3 Block Rates

B.3.1 Calculation and Analysis

Often, safety data in industry are expressed as *block rates*, or the rate of safety events for fixed-sized blocks of time (e.g., Anderson and Barkan, 2004). The rate is the number of safety events normalized by dividing it by some indication of the amount of exposure.

For example, injury data for the railroad industry is often expressed as the number of injuries per 200,000 worker-hours per month (e.g., Federal Railroad Administration (FRA), 2006). Specifically, to calculate the block rate for a month, an evaluator divides the number of injuries in that month by the number of worker-hours in that month, and multiplies the result by 200,000. In this case, the injuries are normalized by worker-hours and the block is a month. The constant 200,000 is included to provide a convenient scale for the rates.

For statistical analysis of safety interventions, each event rate for each block of time is a single data point, and the sample size equals the number of blocks (e.g., Cooper, Philips, Sutherland, and Makin, 1994; Ricci, 2003). Analyses compare the rates for multiple blocks to determine if safety performance has changed significantly (e.g., Hale, Guldenmund, van Loenhout, and Oh, 2008; O'Toole, 2002; Hickman and Geller, 2003; Krause, Seymour, and Sloat, 1999, Waters and Duncan, 2001). For example, an evaluator may compare the mean monthly injury rates at a site before and after a safety intervention using a t-test, where the estimated standard error is derived from the standard deviation of the monthly injury rates. The sample size is equal to the number of months in the evaluation period.

Alternatively, correlation analysis of block rates may be used to assess the presence of a safety trend. For example, an evaluator may calculate the correlation between the monthly injury rate and the number of months from a starting point. If safety is improving, there should be a negative correlation, with injury rates tending to decrease as the number of months increase. Both correlations and mean comparison may be used on the same block rates in a process control chart (St. Clair, 1995) to detect changes in a safety process.

B.3.2 Problems

There are four problems with using block rates as data points in statistical analyses:

- Data Quantization
- Invalidation of Parametric Assumptions
- Sensitivity to Arbitrary Parameters
- Imbalanced Weighting of Data Points

B.3.2.1 Data Quantization

Safety performance is a continuous variable, but block rates can effectively quantize the data into discrete levels. This effect is greatest when the duration of the block is at or below the average period between events. For example, when one-month blocks are used with a site averaging an injury per month or less. This quantization may potentially lower the statistical power of the analyses (Maguire, Pearson, and Wynn, 1952).

For example, consider a year of hypothetical raw injury data in Table 29, following an intervention that began on January 16.

Table 29. Hypothetical Raw Injury Data

Injury	Date & Time	"Days Since Last Injury"
Injury 1	01/01 00:00	
Injury 2	01/16 13:20	16
Injury 3	02/02 20:08	17
Injury 4	02/22 01:03	19
Injury 5	03/15 09:10	21
Injury 6	04/08 02:11	24
Injury 7	05/04 10:25	26
Injury 8	06/02 16:55	29
Injury 9	07/05 05:28	33
Injury 10	08/10 08:44	36
Injury 11	09/19 12:23	40
Injury 12	11/03 03:05	45
Injury 13	12/22 16:46	50

As indicated by the "Days Since Last Injury" sign at the entrance to the site, these data represent a perfectly monotonic improvement in safety, with injury frequency decreasing smoothly as time progresses, which Figure 27 illustrates graphically by plotting the injuries on a timeline. An evaluator would expect such a perfect monotonic relation between time and injury to produce correlation at or very close to 1.0[17].

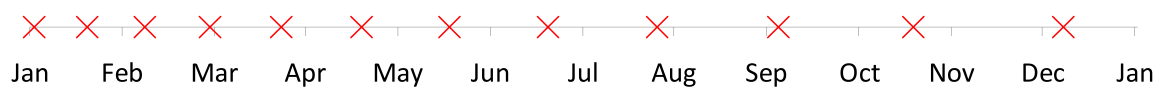


Figure 27. Timeline of Hypothetical Injuries

However, converting these data to monthly injury rates fails to show this smooth progression, as shown in Table 30 and Figure 28. For this example, worker-hours are constant at 1360 per day.

The data are quantized into only a few possible values, essentially "rounding" the data to values corresponding to zero, one, or two injuries per month. An evaluator may get a month with three or even more injuries, but these would be rare for the average injury frequency, and only modestly reduce quantization. Some reduction from strict quantization comes from variations in the worker-hours per month (in this example, variation is due to only the variable number of days per month). In most cases, this has a

[17] This paper uses an unrealistic monotonic relation for these examples to make the artifacts of block rate graphically apparent.

small impact on the quantization. The effect is to take a continuous ratio variable, injury rate, and force it into discrete categories, which discards information (Alt, King, and Signorino, 2000). All events within a block are treated as equivalent irrespective of when they occur within the block.

Table 30. Monthly Injury Rates

Month	Injuries	Worker-Hours	Block Rate*
January	2	42160	9.49
February	2	38080	10.50
March	1	42160	4.74
April	1	40800	4.90
May	1	42160	4.74
June	1	40800	4.90
July	1	42160	4.74
August	1	42160	4.74
September	1	40800	4.90
October	0	42160	.00
November	1	40800	4.90
December	1	42160	4.74
January	0	42160	.00
*Injuries / Worker-hours × 200,000			

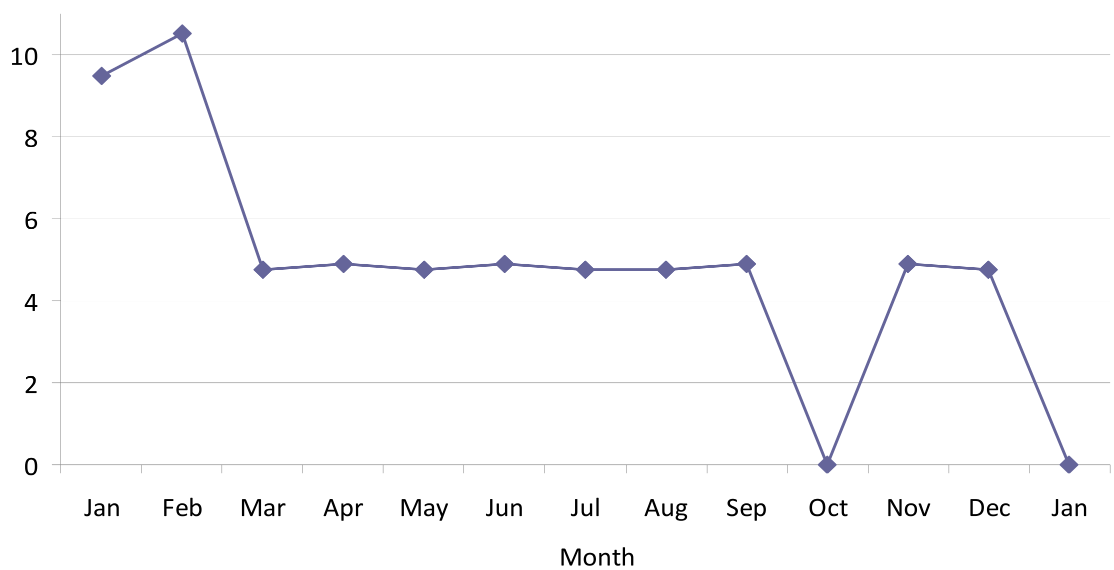


Figure 28. Monthly Injury Rates

The improving trend is still detectable in Figure 28, but appears weaker than it actually is, featuring a couple reversals in the trend that are absent in the raw data. The correlation between the block rate and month is –0.736, well below the 1.0 an evaluator may expect.[18] With actual data, this weakening of the apparent relation may make the difference between achieving a significant result or not. The quantization has added variance or noise to the data, reducing statistical power. This means an evaluation may fail to statistically detect an actual impact. Confidence intervals on the descriptive statistics are wider, so findings are less precise. To compensate, evaluations need to gather more data over a longer period, which adds expense and increases the threat of the intervention being canceled before the evaluation is completed.

An evaluator may reduce quantization by using larger time blocks. For example, an evaluator may use bimonthly injury rates rather than monthly, as Table 31 and Figure 29 show. This reduces quantization, with data values potentially assuming a greater number of possible values (i.e., those corresponding to 0 through 4 or more injuries per two-month block). The result is a considerably smoother representation of the relation of injuries with time, with a correlation of –0.824. However, it still does not capture the perfectly monotonic relation between injuries and time we have in this example. More seriously, the sample size has been halved, which offsets the gain in statistical power afforded by the higher correlation coefficient.

Table 31. Bimonthly Injury Rates

Month	Injuries	Worker-Hours	Block Rate
Jan-Feb	4	80240	9.97
Mar-Apr	2	82960	4.82
May-Jun	2	82960	4.82
Jul-Aug	2	84320	4.74
Sep-Oct	1	82960	2.41
Nov-Dec	2	82960	4.82
Jan-Feb	0	81600	.00

[18] The result is statistically significant ($z = –3.00$, $p = 0.00144$) for these hypothetical data, but only because of the underlying monotonic relationship that is very improbable for real data.

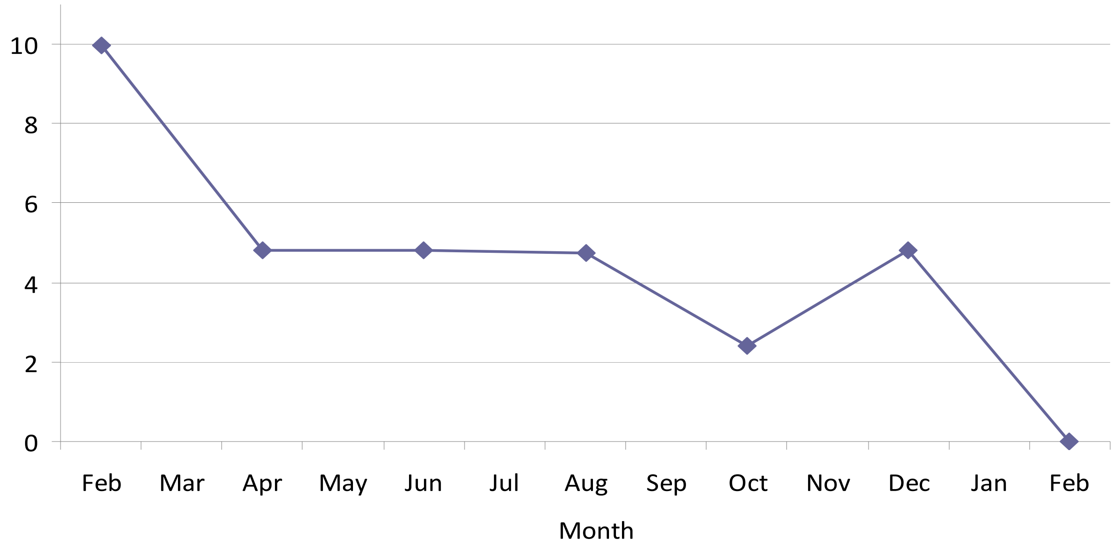


Figure 29. Bimonthly Injury Rates

B.3.2.2 Invalidation of Parametric Assumptions

Analysis of small sets of block rates is complicated by typically non-normal distributions. Indeed, it should be expected that safety events result from Poisson process, where the probability of each event is independent of the time since the previous event (Maguire, Pearson, and Wynn, 1952). This is consistent with Reason's (1996) "Swiss cheese" model of safety where for each moment of exposure one "spins the cheese," which represents a constant probability of an accident. If safety is a Poisson process, then the number of accidents per fixed block of time will have a Poisson distribution (Devore, 1982). Poisson distributions are characterized by population variances that equal the mean and increasing positive skewness as means decrease; skewness is greatest when the average number of events per block is small.

For sparse data, the normal distribution is not an adequate approximation of the Poisson distribution (Devore, 1982). Specifically, when there are typically fewer than dozens of events per block, the distribution of block-rate data from Poisson process may compromise several assumptions of conventional parametric analyses such as t-tests, analysis of variance (ANOVA), product moment correlation, and regression.

- *Normality of the Sampling Distribution*. T-tests and ANOVAs assume normally distributed sampling distributions. Consistent with the Central Limit Theorem, such sampling distributions may not be achieved when small sample sizes are combined with skewed data distributions, such as those with Poisson distributions (Hays, 1981).
- *Homogeneity of Variance*. Given the variance equals the mean in a Poisson distribution, if there are group differences in the mean safety performance, then there is an invalidation of the assumption of homogeneity of variance on which ANOVAs rely.

- *Homoscedasticity*. Regression analysis assumes constant variance of the residuals along the trend (i.e., homoscedasticity). Significance testing of the correlation furthermore requires that the residuals be normally distributed (Pedhazur, 1982). Neither of these assumptions are satisfied with Poisson distributions when the number of events per block are few. For correlations of such Poisson-distributed variables, the variance is correlated with the predicted values of the corresponding regression line, and the residuals are skewed.

Nonparametric statistical analysis may be used instead of parametric analyses to avoid these issues, but these typically have lower statistical power than their parametric counterparts (Hays, 1981). The data may also be transformed to achieve an approximately normal distribution; specifically, the square root transform makes Poisson distributions more normal-like and improves the homogeneity of variance (Winer, Brown, and Michels, 1991). However, the transformed distributions may still not be adequately normal when the events per block are very low. In any case, transformations complicate reporting the magnitude of the results. For example, the reversed-transform of the means that are inferentially compared will not equal the actual means of the groups.

A log-linear regression (e.g., as performed by Spangenberg, Mikkelsen, Dyreborg, and Baarts, 2002) is perhaps the best technique to address the non-normal characteristics of block rates (strictly speaking, log-linear regression uses frequencies of events per block, rather than block rates, but exposure may be included in the regression model as an "offset"). However, such analyses do not address the other problems presented by block rates, and thus an alternative to block rate is required.

B.3.2.3 Sensitivity to Arbitrary Parameters

To convert the raw injury data to block rates, an evaluator must select two parameters: the size of the time blocks and the relative start point of the blocks. It is culturally reasonable to use year-long or month-long blocks and start with the first of each month or year. However, these parameter choices are arbitrary from both a safety and statistical perspective. Nonetheless, they can affect the size of effects an evaluator observes.

For example, suppose for the raw data in Table 29 the first injury happened to occur on January 15 rather than January 1, but otherwise, the smooth improvements in safety carried forward (see Table 32).

Table 32. Hypothetical Raw Injury Data, Alternative Date of First Injury

Injury	Date & Time	"Days Since Last Injury"
Injury 1	01/15 00:00	
Injury 2	01/30 13:20	16
Injury 3	02/16 20:08	17
Injury 4	03/08 01:03	19
Injury 5	03/29 09:10	21
Injury 6	04/22 02:11	24
Injury 7	05/18 10:25	26
Injury 8	06/16 16:55	29
Injury 9	07/19 05:28	33
Injury 10	08/24 08:44	36
Injury 11	10/03 12:23	40
Injury 12	11/17 03:05	45
Injury 13	01/05 16:46	50

This happens to shift around the injuries for each month, as Figure 30 shows. This arbitrary difference in the data changes the correlation between the block rate and time from –0.736 to –0.647. Similar differences in the results can occur if an evaluator arbitrarily chooses to shift the start of each block, for example aggregating injuries from the fifteenth to fourteenth of adjacent months, rather than the first through the end of each month. With actual 30-month data, differences in p-values have been observed to vary from 0.0103 to 0.0919 by shifting the start day for month-long blocks.

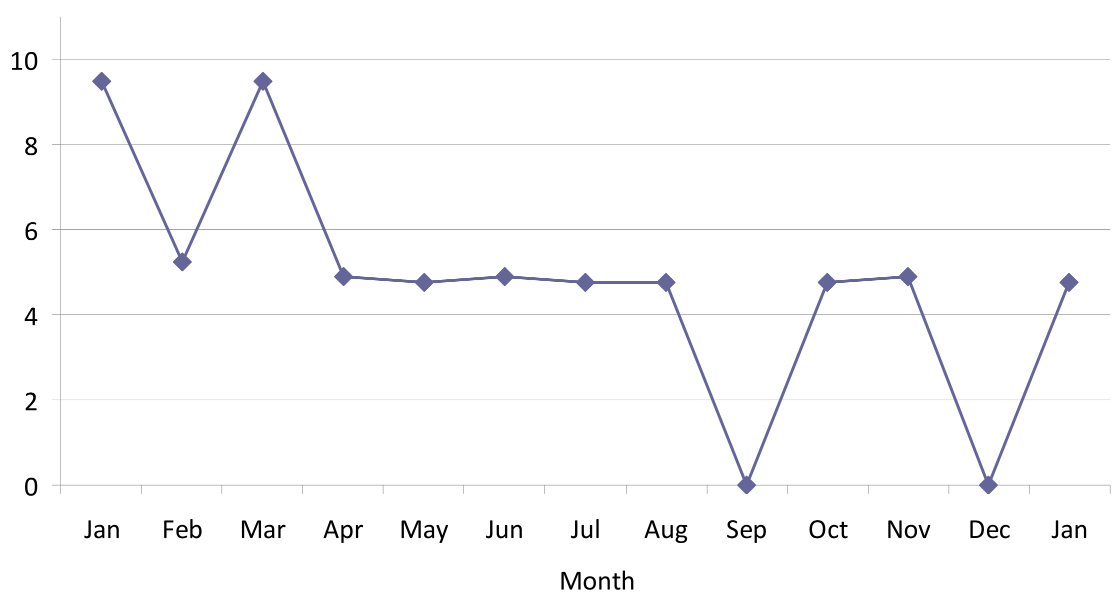


Figure 30. Monthly Injury Rates, First Injury on January 15

Block size represents another arbitrary parameter that nonetheless affects the results. In addition to monthly blocks, an evaluator may use blocks that are bimonthly, quarterly, annual, weekly, fortnightly, or any number of days. In general larger block sizes mean less quantization, increasing the strength of relations but with smaller sample sizes offsetting this. However, the offset is not mathematically perfect and different block sizes can result in different levels of significance for the same data.

Because of the sensitivity of block-rates to its parameters, an evaluator may fail to significantly detect actual change in safety performance simply because the evaluator chose an unfortunate block size and start point. To counteract this, an evaluator could “optimize” the start point and block size, exhaustively trying all values and choosing the ones that produce the highest level of significance. This, however, exposes the evaluator to the opposite problem of “over-fitting” the statistical analysis to random characteristics of the data, thus elevating the chance of a Type I error –appearing to detect an effect that is not in fact present.

B.3.2.4 Imbalanced Weighting of Data Points

Conventionally, blocks are defined by a fixed duration of time (e.g., a month) rather than a fixed duration of exposure (e.g., 50,000 elapsed worker hours). Statistics such as means assign equal weight to each data point, and therefore each safety level associated with each block has equal weight. However, arguably the weight of the safety occurring for each block should be proportional to the amount of work completed rather than the raw time. Normalization attempts to correct this by dividing the event frequency by the work completed in a block. Nonetheless, this practice can result in “paradoxical” results. For example, when the amount of work per block is not equal, the mean of the rates of the blocks does not necessarily equal the overall rate.

Consider, for example, the hypothetical data in Table 33, with various levels of worker hours for each month.

Table 33. Hypothetical Injury Data with Variable Work per Block

Month	**Injuries**	**Worker-Hours**	**Block Rate**
January	1	45,000	4.44
February	1	40,000	5.00
March	0	35,000	0.00
April	2	30,000	13.33
May	0	35,000	0.00
June	1	40,000	5.00
Totals	**5**	**225,000**	**27.78**

Over a six month duration, there were five injures and 225,000 worker-hours, for an overall rate of 4.44 injuries per 200,000 worker-hours (5 / 225,000 × 200,000). However, the average of the block rate values is 4.63 (27.78 / 6), an apparently higher rate. It is debatable if the 4.63 is a misrepresentation or merely a different measure of performance, but at the very least it may provoke some confusion when results are presented to stakeholders.

B.4 Gap Times

B.4.1 Definition

An alternative that avoids all these issues analysis of block rates is analysis of *gap times*, an approach to event data used in reliability engineering (Department of Defense, 1996). The data points are the gap times, or inter-arrival times, between events, rather than rates per block of time. Specifically, each data point is the amount of exposure present between two adjacent safety events. Gap "time" may be represented by clock or calendar time (e.g., number of days or hours between events), but more often it should be a more direct measure of risk exposure, such as worker-hours or vehicle-miles completed between events, which are themselves imperfectly related to time. In engineering applications, gap times are the collective amount of component operation between failures for a collection of components. In engineering, the definition of "component operation" varies depending on the component and the stress. For electronic components (e.g., disk drives), engineers commonly use hours of operation between failures (e.g., DigiKey, 2011). For analysis of metal fatigue failures, (such as on aircraft fuselages and landing gear), one uses the number of fatigue cycles (corresponding to the number of flights) between failures (e.g., FAA, 2011; see also Russell and Lee, 2005). Likewise, the definition of exposure for the gap times of safety events may vary. Generally, it should be whatever is conventionally used to normalize the data when block rates are used. For example, in the railroad industry, injuries are normalized by worker-hours, so the equivalent gap times would be in units of worker-hours between injuries. In contrast, FRA calculates the rate of train accidents per million train-miles (FRA, 2006), so the equivalent gap times would be units of train-miles between accidents.

Both exposure between safety events and component operation between failures express performance as a period rather than a rate, where a period is the arithmetic inverse of a rate. Just as hours between disk drive failure is the inverse of the failure rate per hour (Department of Defense, 1996), each gap time between safety events is the inverse of the event rate. For example, for any gap time of worker-hours between injuries, the injuries per 200,000 worker-hours is simply 200,000 divided by the gap time. That is, the more worker-hours that have elapsed between injuries the lower the rate per 200,000 worker-hours. This applies to statistics, such as means, in addition to individual data points.

Rather than parsing events into arbitrary time blocks to become data points for analysis, analysis of gap times uses the events themselves as the data points. The sample size is equal to the number of events rather than the number of blocks of an arbitrary size. Gap times preserve the continuous nature of the data lost by block rates. Where block-rate data points hover around only a few possible values for sparse data, gap times ranged in increments equal to the precision that the time of the event is known.

B.4.2 Calculation of Gap Times

Table 34 shows the calculation of gap time for the raw data in Table 29.

Table 34. Gap Time Calculation

Injury	Date & Time	Days Between Injuries	Work per Day	Gap Time
Injury 1	01/01 00:00			
Injury 2	01/16 13:20	15.56	1360	21156
Injury 3	02/02 20:08	17.28	1360	23506
Injury 4	02/22 01:03	19.20	1360	26118
Injury 5	03/15 09:10	21.34	1360	29020
Injury 6	04/08 02:11	23.71	1360	32244
Injury 7	05/04 10:25	26.34	1360	35827
Injury 8	06/02 16:55	29.27	1360	39808
Injury 9	07/05 05:28	32.52	1360	44231
Injury 10	08/10 08:44	36.14	1360	49146
Injury 11	09/19 12:23	40.15	1360	54606
Injury 12	11/03 03:05	44.61	1360	60674
Injury 13	12/22 16:46	49.57	1360	67415

For these example data, work was constant over time at 1360 worker-hours per day, with the site characterized by fully staffed shift work seven days per week. Thus, the total gap time (worker-hours between injuries, in this case) is the number of days between injuries times the work completed per day. This example assumes time of day is recorded for each injury. In practice, the safety event is sometimes only known to the nearest day, which is acceptable for many situations. However, if the time of day of the event is known, more precise gap times can be calculated by using the amount of work completed between each point of time within the days.

For actual safety data, the amount of work varies day to day, so the gap times should be the sum of the work for each day between events, rather than merely multiplying the days between injuries by a single "work per day" value. However, while the time of day of a safety event may be known, often the work conducted at a site is only available for an entire month, rather than each day or other finer units of time. Under these conditions, an evaluator may sum the each month's *average* daily work per day. Alternatively, an evaluator may use linear interpolation of each month's work, which is often more computationally convenient and arithmetically identical. Using average work per day rather than actual work per day represents a minor loss of precision as long as there are no large swings in work completed within each month.

B.4.3 Advantages over Block Rate

Gap times avoid all three issues associated with block rates.

- *Data Quantization.* Gap times are not subject to quantization of the data, and the additional noise it introduces into the data. The data is as precise as the date and time recorded for the event. Smooth relations remain smooth. For example, there is a 1.0 linear correlation between the gap times and time in Table 34. This can be seen in Figure 31. The result is the potential for gap-time analysis to provide statistically more powerful analyses of safety data.

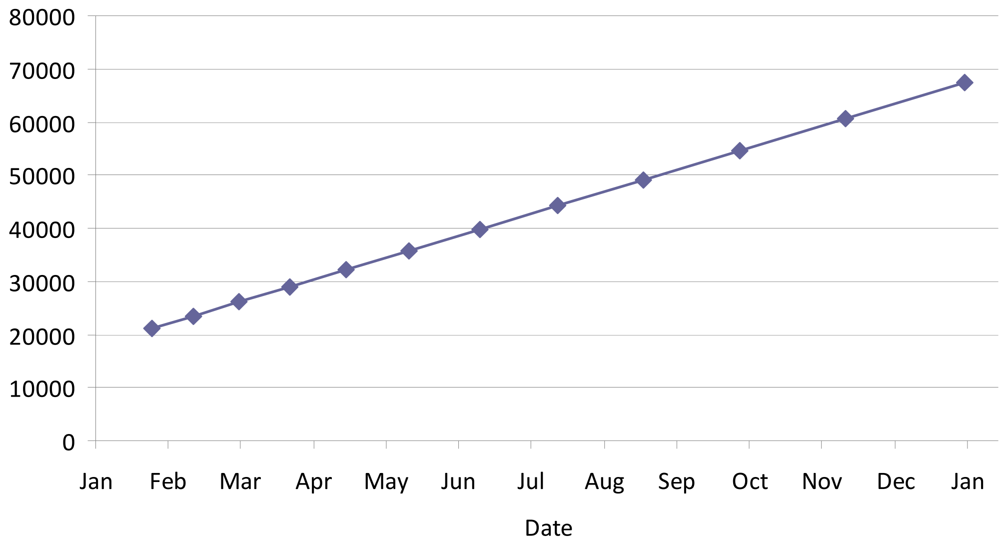


Figure 31. Gap Times over Time

- *Sensitivity to Arbitrary Parameters*. Gap times lack blocks, so there are no arbitrary parameters to affect the analyses.
- *Invalidation of Parametric Assumptions.* If safety events result from Poisson processes, the gap times are distributed exponentially. Exponential distributions of sparse data potentially violate the same assumptions as the Poisson distributions that characterize block rates. However, there are established fully developed statistical methods available for the analysis of exponential distributions that have been applied for problems in engineering (Nelson, 1982), medicine (Lee and Wang, 2003), and political science (Box-Steffensmeier and Jones, 1997). Key methods are provided in Analysis Methods below.
- *Imbalanced Weighting of Data Points.* Each data point is by definition weighted by the amount of work associated with it. The inverse of the mean of the gap times will always exactly equal the overall event rate irrespective of changes in the amount of work at a site, provided the evaluation period begins and ends on a safety event.

This is shown as follows. The event rate R for a period of time is the number of injuries n divided by the total work over the period W:

$$R = \frac{n}{W}$$

The total work is equal to the sum of work, w_i, completed between each event so:

$$R = \frac{n}{\sum w_i}$$

A gap time by definition is the work completed between each event, w_i. Thus the rate is the arithmetic inverse of the average gap times –the sum of the gap times divided by the number of events.

$$w_{avg} = \frac{1}{R} = \frac{\sum w_i}{n}$$

Lacking quantization, invalidation of distribution assumptions, arbitrary parameters, and imbalanced weighing, analysis of gap time has the potential to be a superior measure of safety performance over block rates.

B.4.4 Analysis Methods

Analysis of small sets of gap times is complicated by typically non-normal distributions. Indeed, if safety events are a Poisson process, where the probability of each event remains constant over the time between events, then the gap times will have an exponential distribution (Devore, 1982). Like Poisson processes for sparse events, exponential distributions are skewed and the variance varies with the mean, thus conventional parametric analyses may be unsuitable.

Gap times have an addition issue usually not found in block-rate data. Gap times typically include a "censored" datum. In a typical safety performance evaluation, the end time for the evaluation is unlikely to be precisely on the date of a safety event. Instead, the last recorded safety event may be days or weeks before end of the evaluation. It is not known when the next safety event will occur. However, it is known that the time until the first post-evaluation event must be greater than the time between the last event and the time of data acquisition. This renders final datum as "right-censored." The evaluator knows only the lower bound on the datum's value[19].

With conventional least-squares data analyses, a censored datum must be discarded since its precise value is not known. The sample size for a gap-time analysis is reduced from the number of safety events to the number of safety events minus one. Generally, excluding a single censored data point will have little effect on the findings. However, in cases where a safety intervention has had a substantial impact, the evaluation may end with a long streak of no safety events. It would be important to include that aspect of the data in the analysis.

Fortunately, there is *survival analysis* (Lee and Wang, 2003; Hosmer, Lemeshow, and May, 2008), a suite of analytical methods specifically designed to handle times until events and the non-normal distributions and censored data they tend to have. Key methods for evaluation include Cox regression and Weibull regression, which may be

---

[19] This is not an issue for block rates as long as the evaluation data begins and ends on the boundaries between two blocks.

directly applied to gap times to determine their relation to numerical and categorical variables, such as the presence and absence of a safety program (Cook and Lawless, 2010).

B.5 Conclusion

Gap times are a viable alternative to traditional block rates for the evaluation of safety data. Unlike block rates, gap times avoid quantizing the data, providing better statistical power. Both gap times and block rates tend to have non-normal distributions, but survival analysis techniques handle the non-normal distributions in gap times. Finally, gap times appropriately weigh the data points and lack arbitrary parameters that may affect the results.

The potential for improved statistical power through analysis of gap times is especially important for program evaluations of safety impacts, where data tend to be sparse, very noisy, but nonetheless scientifically and operationally important.

B.6 References (Appendix B)

Alt, J. E., King, G., & Signorino, C. S. (2000). Aggregation among binary, count, and duration models: Estimating the same quantities from different levels of data. *Political Analysis*, *9*(1), 21-44.

Anderson, R. T. & Barkan, C. P. L. (2004). Railroad accident rates for use in transportation risk analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *1863*, 88-98.

Box-Steffensmeier, J. M. & Jones, B. S. (1997). Time is of the essence: Event history models in political science. *American Journal of Political Science*, *41*(4), 1414-1461.

Cook, R. J. & Lawless, J. F. (2010). *The statistical analysis of recurrent events.* New York: Springer Science + Business Media.

Cooper, M. D., Phillips, R. A., Sutherland, V. J., & Makin, P. J. (1994) Reducing accidents using goal setting and feedback: A field study. *Journal of Occupational and Organizational Psychology, 67,* 217-240.

Department of Defense (1996). Handbook for reliability test methods, plans, and environments for engineering, development, qualification, and production (MIL-HDBK-781A). Washington, DC: U.S. Department of Defense.

Devore, J. L. (1982). *Probability and Statistics for Engineering and the Sciences*. (3^{rd} ed.) Pacific Grove, CA: Brooks/Cole Publishing.

Digikey Corporation, (2011). Digi-Key Catalog US2011, Thief Falls River, MN: Digikey Corporation. http://pdfcatalog.digikey.com/US2011/digikey.pdf downloaded August 18, 2011.

Federal Aviation Administration, (2011). *Emergency airworthiness directive (AD) 2011-08-51*. Washington, DC: Federal Aviation Administration. http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAD.nsf/0/77979606b6421

17286257869006afb63/$FILE/2011-08-51_Emergency.pdf, downloaded August 18, 2011.

Federal Railroad Administration (2006). *Railroad safety statistics 2005 annual report*. Washington, DC: Federal Railroad Administration. http://safetydata.fra.dot.gov/officeofsafety/ProcessFile.aspx?doc=bull2005-book.pdf, downloaded March 3, 2011.

Hale, A. R., Guldenmund, F. W., van Loenhout, P. L. C. H., & Oh, J. I. H. (2008). Evaluating safety management and culture interventions to improve safety: Effective intervention strategies & lessons learned. *Safety Science*, *48*(8), 1026-1035.

Hays, W. L. (1981). *Statistics* (3rd ed.) New York: CBS College Publishing.

Hickman, J. S. & Geller, E. S. (2003). Self-Management to Increase Safe Driving Among Short-Haul Truck Drivers. *Journal of Organizational Behavior Management*, *23*, 1-20.

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis.* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Krauss, T. R., Seymour, K. J., & Sloat, K. C. M., (1999). Long-term evaluation of a behavior-based method for improving safety performance: A meta-analysis of 73 interrupted time-series replications. *Safety Science*, *32*, 1-18.

Lee, E. T. & Wang, J. (2003). *Statistical Methods for Survival Data Analysis*. Hoboken, NJ: John Wiley & Sons.

Maguire, B.A., Pearson, E.S., & Wynn, A.H.A. (1952). The time intervals between industrial accidents. *Biometrika*, 39(1-2), 168-180.

Nelson, W. B. (1982). *Applied life data analysis*. Hoboken, NJ: John Wiley & Sons.

O'Toole, M. (2002). The relationship between employees' perceptions of safety and organizational culture. *Journal of Safety Research*, *33*, 231-243.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. (2nd ed.) Fort Worth, TX: Harcourt College Pub.

Russell, A, & Lee, K. L., (2005). *Structure-Property Relations in Nonferrous Metals*. Hoboken, NJ: John Wiley & Sons.

Reason, J., (1997). Managing the Risks of Organizational Accidents. Ashgate, Aldershot, UK.

Ricci, M. K. (2003). Evaluating behavior based safety efforts in the railroad industry: Work practice observation (WPO): A pilot study. Final Report, Washington State Utilities and Transport Commission, Summer, WA.

St. Clair, J. F. (1995). Safety Measurement: Towards Continuous Improvement. *Annual Quality Congress*, *49*. Cincinnati OH. May, 150-158.

Spangenberg, S., Mikkelsen, K.L., Kines, P., Dyreborg, J., & Baarts, C. (2002). The construction of the Øresund Link between Denmark and Sweden: The effect of a multifaceted safety campaign. *Safety Science*, *40*, 457-465.

Waters, R. M., & Duncan, M. (2001) Behavioral safety programs in the Department of Energy. *Proceedings of the 45th Annual Meeting of the Human Factors Society*, Minneapolis, MN. October.

Winer, B.J., Brown, D. R., & Michels, K. M. (1991). *Statistical Principles in Experimental Design* (3rd ed.). New York: McGraw-Hill.

Abbreviations and Acronyms

ANOVA	analysis of variance
BLET	Brotherhood of Locomotive Engineers and Trainmen
BNSF	Burlington Northern Santa Fe
BST	Behavioral Science Technology, Inc.
CAB	changing at-risk behavior
CI	continuous improvement
CRZ	cab red zone
CSA	clear signal for action
CTC	centralized traffic control
FELA	Federal Employer's Liability Act
FRA	Federal Railroad Administration
FTX	field training exercises
FWSU	Fort Worth Service Unit
HOSU	Houston Service Unit
KEYS	Kirby, East Yard, and SOSAN
LVSU	Livonia Service Unit
NTSB	National Transportation Safety Board
PPF	Peer-to-peer Feedback
SASU	San Antonio Service Unit
SLD	Safety Leadership Development
SOSAN	South San Antonio
TSC	total safety culture
UP	Union Pacific
UTU	United Transportation Union