

The Efficacy of Behavior-based Safety in the U.S. Railroad Industry: Evidence from Amtrak-Chicago

Joyce RANNEY

*Operator Performance and Safety Analysis Division,
Volpe National Transportation Systems Center, Cambridge, MA, USA*

Christopher NELSON

RAND Corporation, Pittsburgh, PA, USA

Michael COPLEN

*Human Factors Research & Development Program,
Federal Railroad Administration, Washington, DC, USA*

Abstract. This paper provides an implementation and impact evaluation of Clear Signal for Action (CSA), a railroad industry safety program that combines behavior-based safety with process improvement methods. The CSA project took place in two departments at Chicago's Amtrak terminal in 2001 to 2002. A corporate downsizing that coincided with much of the study period, as well as methodological issues, made results somewhat uncertain. However, a discernible increase in the prevalence of safe behavior, a reduction in injury rate variability, and improvement in management-worker relations and safety culture justify further research.

Keywords. Railroad, Behavior-based Safety, Process Improvement

1. Introduction

Behavior-based safety (BBS) approaches seek to provide safety managers with behavioral indicators to identify and address hazards before they eventuate into injuries. The Federal Railroad Administration (FRA) is interested in BBS because of the potential to reduce injuries through changes in underlying safety culture. Both the FRA and railroad safety professionals are interested in using observation data as an alternative indicator of safety as it provides plentiful data on at-risk behaviors *before* an event occurs, as opposed to injury data, which occur *after* the fact. Some BBS methods incorporate peer-to-peer feedback with the observations, thus simultaneously improving safety and safety culture.

However, traditional BBS approaches often place too little emphasis on the influence that upstream managers, systems, and practices have on at-risk behavior *and conditions*. As a result, BBS has received negative press from several unions. Thus, recent variants have added methods from process improvement, which pays attention to organizational policy and procedures, culture, and other systemic causes of injuries, as well as implementing corrective actions to address issues identified during observations. This paper provides an implementation and impact evaluation of one such variant as an instrument of safety improvement in the U.S. railroad industry.¹ Clear Signal for Action

¹ Behavioral Science Technology, Inc. (BST) carried out the project. The Federal Railroad Administration (FRA) Human Factors R&D Program (HF/R&D) sponsored both the project and the evaluation.

(CSA), the FRA methodology in this area, seeks to promote safe behavior and reduce injuries through both behavioral observations and feedback, and process improvement. First, *behavioral observations and feedback* are based on a site-specific datasheet developed by examining past injury reports to identify common behaviors and conditions leading to injuries (Krause, 1997). Observers are trained in how to *identify* behaviors/conditions covered in the datasheet, *deliver* feedback in a positive manner, and *interview* the person observed to identify barriers to safe practices.² This method uses peer-to-peer observations and encourages communication about safety, thereby improving the safety culture. Second, CSA employs *process improvement* techniques. A Steering Committee uses trend analysis of the observation-feedback data to identify high-exposure situations that might be addressed through implementing corrective actions. Both mechanisms are overseen by the Steering Committee, which includes representatives of labor, management, and safety professionals.

While there is considerable evidence that the observation and feedback components of BBS work in manufacturing settings (see, e.g., Sulzer-Azaroff & Austin, 2000; Guastello, 1993; Kopelman, 1986),³ less is known about the effectiveness of process improvement. Similarly, little is known about the joint effectiveness of BBS and process improvement. Ricci (2005) provided evidence for the effectiveness of observation, *without* feedback, and corrective actions in the railroad industry. The study reported on in this paper provides an evaluation of a project that sought to join peer-to-peer observation *with* feedback and corrective actions.

2. Data and Methods

The study was conducted at Amtrak's Chicago terminal from May 2001 through September 2002, and focused on two in-facility work settings: Station Services and Mail and Express.⁴ Unfortunately, a company-wide downsizing began nine months after project initiation and continued through the end of the project, seriously affecting implementation, especially in Mail and Express. Because of the uncertainty about the quality of the data from that unit, this paper reports only on findings from Station Services. Station Services primarily services passengers getting on and off Amtrak trains arriving at and departing from Union Station. The department worked a two-shift/seven-day-a-week operation.

The study focused on both program implementation and impacts. The implementation analysis relied on three data collection methods: direct observation of project meetings, interviews and focus groups, and document analysis. Interviews and focus groups were used to assess the extent to which (a) the program was implemented, (b) the stakeholders seemed committed to the project, (c) the methodology was applicable in the railroad setting, and (d) the program was having an effect on safety beyond impacts measurable through behavioral and injury data. Researchers attended 15 of 21 Steering Committee meetings.

² It is important to note that previous studies indicate that training by itself has a positive impact on organizational productivity, whether or not those trained actually put the new skills into practice (Kopelman, 1986).

³ Kopelman (1986), for instance, reviewed 42 field studies on job-related feedback, and in all cases the results were positive. On the average, work behaviors improved by 47 percent.

⁴ This is a work in progress and has not been formally reviewed by RAND.

Interview and focus group respondents included workers, workers trained as observers, managers, union representatives, and safety professionals. Interviews and focus groups were administered according to a protocol, with a note-taker present. Respondents were interviewed separately for an hour; focus groups lasted for an hour and a half.

The study's impact evaluation focused on both the prevalence of safe behavior and conditions and injuries in the study sites. Analysis of the impact on the prevalence of safe behavior and conditions relied on observation data collected during the project. Traditional inter-rater reliability tests were not possible. However, diagnostic tests suggest that observed trends in percent safe were not due to changes in observer composition. Impact analysis included comparison of the mean percent safe during the project with the mean percent safe during a baseline period from the same organization approximately one year before the project. Trends in percent safe during the project were assessed using mixed effects regression models.

Injury data were collected and analyzed to assess the project's impact on the frequency of reportable injuries and illnesses. Since the frequency of injuries varies with work hours (exposure), injury data were normalized either through use of rates⁵ or through adjustments in statistical models data. The analysis of injury data used a control group time series design with 69 observations before, 12 during, and 13 after the project, for the period January 1996 through September 2004. The control group was constructed using injury rates for all Amtrak-Chicago departments except Station Services and Mail and Express.

3. Key Findings

Key findings relate to (a) the quality of implementation, (b) impacts on the prevalence of safe behaviors and conditions, (c) impacts on injury rates, and (d) other impacts.

The implementation analysis found that participants in the program conducted an average of just 36 observations a month and never exceeded 138 in a given month, well below the target of 240 per month⁶ and below the median rate for all organizations of similar size implementing Behavioral Science Technology, Inc.'s (BST) methodology. Yet, the quality of observations ranked at or above the 75th percentile for similar organizations, as evaluated with a rating scale developed by BST (2001, 2004). The number of corrective actions was also below project targets. Only three such actions, addressing lifting, pinch point injuries, and pre/post vehicle inspections, compared with the target of eight were produced.⁷ Moreover, uneven implementation compromised some aspects of the corrective actions.

The Steering Committee was effective early in the process, while it received external guidance, but it struggled with questions about management support and job uncertainty. Still, the group executed 428 observations and two observation-training courses, removed three barriers, and completed two newsletters.

⁵ Injury rates are injuries divided by hours worked, multiplied by 200,000, which yields the number of injuries per 100 FTE-years. Data were obtained from Amtrak's Automated Safety Information System.

⁶ There were 240 employees in Station Services and Mail and Express; each was to be observed once a month.

⁷ One corrective action was to be produced during each of the eight months for which sufficient data were available.

Examination of the observation data suggests a discernible increase in the prevalence of safe behavior, from 87 percent safe during the baseline period to 94 percent during the project.⁸ We also sought to determine whether there were any improvements *during* the project. A mixed effects regression found that the percent of observations coded as safe grew at an annualized rate of 10 percent (see Table 1).⁹ By contrast, analysis of observation data related to specific corrective actions found only mixed pre/post impacts on effectiveness of the actions in increasing the prevalence of safe behavior.

Table 1: Growth Estimates for Datasheet Items Not Subject to Corrective Actions

Parameter	All observations	Matched observations
Starting value	0.89	0.89
Annualized rate of increase	0.10	0.11
Number of observations	265	145
Number of datasheet items	32	17

Matched observations include datasheet items present in both the baseline and project periods.

Given that injury data were collected monthly, we used a time series design to investigate the project's impact on injury rates. Poisson regression was used to model monthly injury counts as a function of time, a dichotomous variable representing introduction of the project, and hours worked. We found no discernible differences in injury rates before, during, and after the project in Station Services. Variation ratio tests,¹⁰ however, revealed that the project did coincide with discernible or near-discernible reductions in injury rate *variability* in Station Services (see Table 2). The data were insufficient to test for changes in injury severity.

Table 2: Changes in Monthly Incident Rate Variation (Standard Deviations)

Period	Station Services		Chicago Comparison	
	Full Sample	Restricted Sample	Full Sample	Restricted Sample
Before project (B)	7.8	6.6	2.4	2.3
During project (D)	5.1	5.1	2.2	2.7
After project (A)	10.8	10.8	2.0	2.0
H ₀ : B/D = 1	<i>p</i> = 0.08	<i>p</i> = 0.32	<i>p</i> = 0.74	<i>p</i> = 0.86
H ₀ : D/A = 1	<i>p</i> = 0.02	<i>p</i> = 0.02	<i>p</i> = 0.73	<i>p</i> = 0.73
H ₀ : B/A = 1	<i>p</i> = 0.16	<i>p</i> = 0.06	<i>p</i> = 0.42	<i>p</i> = 0.54

In terms of non-quantitative indicators, interview data provided evidence that the project led to improvements in safety awareness, dialogue about safety hazards, and labor-

⁸ The difference was statistically significant. However, the distributional properties of the percent safe data suggest caution in interpreting the standard errors.

⁹ The statistical analysis modeled percent safe as a function of time, random effects for datasheet items, and an observation-specific disturbance term. The model was estimated using generalized estimating equations with robust standard errors (see, e.g., Zeger, Liang, & Albert, 1988).

¹⁰ Here the null hypothesis that the ratio of the variances during the two periods is equal to 1 was evaluated using the F-distribution (see Mendenhall, Wackerly, & Scheaffer, 1990).

management relations. However, such impacts were largely restricted to workers directly involved in the project.

4. Discussion and Conclusion

These findings provide evidence that the project had positive impacts on the safety culture and on the prevalence of safe behaviors and conditions. While the injury data suggest no impact on the level of risk, there were impacts on monthly variation.

However, several limitations to the study must be considered. First, implementation was hindered by layoffs of both safety employees and frontline workers, with more senior workers taking the jobs of less senior workers, and significant changes in reporting relationships. Workers were therefore engaged in unfamiliar tasks, which might have increased injuries and reduced program effects. Second, the downsizing and other factors precluded use of standard methods for assuring the reliability of the observation data. Thus, we should consider measure instability as a rival explanation for the reported increases in the prevalence of safe behavior. Third, because baseline data were collected approximately one year before the project, history effects may have acted as threats to internal validity. Finally, the project might have increased workers' willingness to report injuries, which could have led us to underestimate the program's effectiveness.

Given the implementation difficulties, it may be surprising that any impacts were found. A stronger and more sustained implementation, with more observations, might have further increased the prevalence of safe behavior, and stronger functioning by the Steering Committee might have more successfully removed systemic barriers to safe behavior. Indeed, our evidence suggests that corrective actions were less effective than peer observation and feedback. Quality concerns with the corrective actions might have been partially related to the lack of management support and limited prior experience with corrective actions. Managers may have been hampered by downsizing pressures and confusion about how they were to support the project. Had the downsizing not been going on, it is likely that barrier removal would have been effective and that more observations would have been conducted – both of which might have caused observable decreases in injury rates.

The finding that the project was associated with a reduction in monthly variation in injury rates is intriguing but more difficult to interpret. Reductions in variation might suggest a lessening of the "injury-attention" cycle, in which high rates are followed by increased management attention to safety, which reduces rates. Low rates, however, lead to complacency and lowered defenses, generating higher rates. Second, as Deming (2000) and others in the "statistical process control" tradition note, systems with low variability are "in control," and thus more amenable to improvements, since low residual variance makes it more likely that managers can detect – and learn from – changes in system-level policy. In any case, this finding deserves additional exploration.

In conclusion, CSA seeks to improve peer-to-peer interactions concerning safety; enhance cooperation between management and labor in addressing safety issues and, in doing so, prevent injuries; and positively affect safety culture and injury rates. Given the implementation difficulties, and loss of observations, it is perhaps surprising to find any positive impacts. The findings suggest that CSA, if implemented effectively, is a

promising strategy for safety improvement. These findings are, however, subject to considerable uncertainty.

Future evaluation should address the methodological issues raised here through pursuit of stronger comparison group designs (including, where possible, random assignment and multiple baseline designs). Moreover, to ensure measure stability researchers will have to develop inter-rater reliability procedures that are congruent with the railroad environment. Future efforts should also explore in greater detail the implementation barriers associated with such projects in dispersed, high-conflict work environments. While many of the difficulties noted here were related to Amtrak's downsizing, some conditions are likely to be found in other carriers and industries as well. Given the potential for the program not only to decrease injury rates but also to improve management-worker relations and safety culture, and to prevent future injuries, additional implementations and evaluations designed to reduce this uncertainty seem justified.

4. References

- Behavioral Science Technology, Inc. (2001). *BAPP® Technology: Implementation manual: Volume 2: Observation*. Ojai, CA: BST Proprietary Document.
- Behavioral Science Technology, Inc. (2004). [BST, Inc., client comment quality ratings]. Unpublished raw data.
- Deming, W.E. (2000). *Out of the crisis*. Cambridge: MIT Press.
- Guastello, S.J. (1993). Do we really know how well our occupational accident prevention programs work? *Safety Science*, 16, 445-463.
- Kopelman, R.E. (1986). *Managing productivity in organizations*. New York: McGraw-Hill.
- Krause, T.R. (1997). *The behavior-based safety process* (2nd ed.). New York: Van Nostrand Reinhold.
- Mendenhall, W., Wackerly, D., & Scheaffer, R. (1990). *Mathematical Statistics with Applications*. Belmont, CA: Duxbury.
- Ricci, M.K. (2005, January). Evaluating behavior-based safety efforts in the railroad industry. Washington, DC. Paper presented at Poster Session 235, Transportation Research Board Annual Meeting.
- Sulzer-Azaroff, B., & Austin, J. (2000, July). Does BBS work? Behavior-based safety & injury reduction. *Professional Safety*, pp. 19-24.
- Zeger, S., Liang, K., & Albert, P. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44, 1049-1060.