Using perception surveys to assess safety system effectiveness

by Charles W. Bailey and Dan Petersen

In a previous article in PS (May 1984), we reviewed some experiments conducted using positive reinforcement in the railroad industry. While the only technique discussed in that article was the use of positive reinforcement, in actuality the experiments were only one phase of a much larger study, extending over a period of almost nine years.

The purpose of the entire study was to learn what works and what does not in safety programming in that industry. In reality what was learned has a much broader application than just the one industry. It is extremely useful to all industries. While there are many facets to the entire study, this article will confine itself to only one: the development of a survey instrument that has remarkable potential.

The study was conducted by a special study committee appointed by the Safety Section Steering Committee of the Association of American Railroads. It was partially funded through two contracts with the Federal Railroad Administration.

The study was conducted under a contract with the Office of Rails Safety Research with help and support from safety experts at the Ballistics Research Laboratory at Aberdeen Proving Grounds and The National Space Technology Laboratory, including statisticians from Computer Sciences Corporation. The Masters of Industrial Safety Program faculty, at the University of Minnesota—Duluth, provided assistance with survey data collection and evaluation and development of a training seminar to familiarize safety officers with results of the study.

Four Railroads made their properties available to the study committee for research:

Burlington Northern Railroad
Duluth, Missabe and Iron Range Railway
Illinois Central Gulf Railroad
Southern Railway System

Their participation and support were largely responsible for the success of the study.

History—need for study
The goal of this study was to discover methods of improving safety program effectiveness in the railroad industry. When the need for such a study began to be apparent in 1975, the railroad industry, along with a number of other industries in the United States, was experiencing a continuing increase in the number of employee injuries. The recently implemented OSHA Law was having
little or no effect in reducing accidents, even with the expenditure of large sums of money by the government and industry to achieve compliance. Clearly, the industry’s safety efforts were not as effective as they once had been.

One reason for this lack of progress, in the view of many safety officers, was a tendency to cling to certain practices and beliefs about safety which had little basis in fact. Research in the areas of psychology and sociology had clearly defined the role human behavior and motivation play in the workplace but few companies were applying these findings in their safety programs.

In their landmark study of the impact of OSHA on workers in the chemical, aerospace and textile industries, Northrup, Rowan and Perry of the Wharton School, University of Pennsylvania, concluded:

The rather substantial changes in corporate commitments and federal requirements in the safety and health field resulting from the law have not been matched by parallel changes in the attitudes and behavior of most workers. Many workers, particularly older, long service workers, may find themselves forced to accept more safety and health protection on the job than they feel is necessary or desirable.

There is ample evidence that regulation has fostered increased attention to safety and health issues and problems in all three industries examined here. There is little evidence, however, that this added attention has yet produced or is soon likely to produce a substantial improvement in the safety and health of most workers. This lack of significant impact exists because the added attention to safety and health generally has been focused primarily on the requirements of regulations instead of on the reduction of risks, in large part because the regulatory system focuses more on means than ends in the control of potential hazards.

To reverse a similar trend in the rail industry, the AAR Safety Section’s Steering Committee proposed to AAR management and to the FRA Office of Rail Safety Research that a joint study be conducted to examine present railroad safety programs, determine and quantify (if possible) the elements or activities which are associated with good safety performance, test elements which show promise in a railroad environment, and to make study results available to the industry for broader application.

Their recommendation was accepted and development of the study began in 1979.

Development of the study format
The contract between the Federal Railroad Administration and the Association of American Railroads was administered by the Ballistics Research Laboratory at Aberdeen Proving Grounds. This team of scientists had extensive experience with safety through their involvement in the development of safety systems for weapons and space programs.

The Aberdeen group proposed a study based on a “systems approach” to management of safety activities—an approach with which they were familiar and which they believed had worked well in the past.

The AAR Study Group, noting the need to go beyond present “traditional” and “procedural-engineered” models, explained its concept of learning more about the effects of behavior-based safety activities. Conceding that few such program activities then existed, Dr. Dan Petersen pointed out the need to remedy the situation through testing of promising “behavioral” approaches to determine their effectiveness.

In the end, both approaches became a part of the study. The staff at Aberdeen Proving Grounds would evaluate the effectiveness of the “procedural-engineered” approach while the AAR Study Group evaluated the “human-behavioral” approach.

To determine more precisely the factors involved in good safety performance, an in-depth survey of safety practices on two railroads was conducted. The participating road, designated Railroad I, had been a leader in safety performance for a number of years. Railroad II was near the bottom of the 18-railroad group in performance.

Questions on both the procedural-engineered and behavior-based approaches to safety were included in the survey.

Five levels of each railroad’s organization, from executives to hourly-rated employees, were surveyed using questionnaires in booklet form, specifically designed for each level, but including some questions common to all. This permitted tracking of responses to common questions from level to level to determine where gaps in communications might be affecting safety performance.

The questionnaires were administered to a stratified random sample of managers and employees at 5 levels in the organization, chosen to provide statistical validity of 95+% at each level.

Safety officers on the two study railroads received training in administering the survey and answering questions from those being asked to participate. Each person surveyed had to be approached, invited to participate, and provided with a place where they would have privacy while completing it. Participants were given an addressed, stamped envelope in which to mail the completed questionnaire to the National Space Technologies Laboratory. There, all data was processed by Computer Sciences Corporation’s staff. This procedure generated a response of better than 97% of the questionnaires distributed on each property.

Numbers on Each Railroad
Participating in Survey

<table>
<thead>
<tr>
<th>Level</th>
<th>RR I</th>
<th>RR II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Top Management</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>Middle Management</td>
<td>107</td>
<td>47</td>
</tr>
<tr>
<td>1st Line Supervisor</td>
<td>62</td>
<td>70</td>
</tr>
<tr>
<td>Employee</td>
<td>476</td>
<td>566</td>
</tr>
<tr>
<td>Total</td>
<td>669</td>
<td>712</td>
</tr>
</tbody>
</table>

Analysis of the data, led to development of an 18-category “profile” for each railroad’s safety efforts, based on responses to survey questions. The response to questions on Railroad I—the railroad with the best performance as measured by other program effectiveness criteria—was significantly more positive overall than that for Railroad II. It was apparent that the survey had been able to accurately identify differences in safety program effectiveness on the two railroads.

While the AAR Study Group was testing the human factor approach to safety programming, the Aberdeen Study Group attempted to find cor-
relations between procedural-engineered oriented safety program elements and various safety performance indicators developed from data supplied by the railroad industry. Data from 18 railroads was used to establish scores for each railroad in 12 subject areas.

1. Safety program content
2. Equipment and facilities resources
3. Monetary resources
4. Reviews, audits and inspections
5. Procedures development, review and modifications
6. Corrective actions
7. Accident reporting and analysis
8. Safety training
9. Motivational procedures
10. Hazard control technology
11. Safety authority
12. Program documentation

The Aberdeen Group’s hypothesis was that high scores in these areas would correlate generally with lower accident and injury rates. Instead, they found little correlation with these procedural-engineered factors. Their report of findings stated, in part:

*It was an unexpected result of this study that so little correlation was found to exist between actual safety performance and safety survey scores. The overall survey score has almost no correlation with train accident rates and cost indicators and is somewhat counter-indicative with respect to personal injury rates. The only two categories which correlated consistently and properly with accident rates were monetary resources and hazards control technology. Two categories, equipment and facilities resources and reviews, audits and inspections, had counter-intuitive correlations.*

This conclusion coincided with the findings of the AAR Study Group which found that responses to most pilot survey questions based on the presence or absence of these 12 criteria did not distinguish any significant differences between the two railroads studied. The results achieved by the Study Groups seemed to be saying:

1. The effectiveness of safety programs cannot be measured by the more traditional procedural-engineered criteria popularly thought to be factors in successful programs.
2. A better measure of safety program effectiveness is the response from the entire organization to questions about the quality of the management systems which have an effect on human behavior relating to safety.
3. The most successful safety programs are those which recognize and deal effectively with worker and supervisor behavior and attitudes which affect safety.

**Further refinement of the survey process**

The conclusion of the first phase of the project left many unanswered questions. Clearly, it had been proven that many of the things commonly thought to influence safety were having little or no effect. In its report, the Study Group cautioned that results could only be considered preliminary and recommended that they be used to develop an instrument and process which could be initiated by companies wishing to evaluate the overall effectiveness of their safety efforts.

Criteria were established by the AAR Study Group to guide development and evaluation of a second “verification” survey which they
proposed be conducted on two additional railroads. This second survey should:

1. Evaluate the organization’s perception of management systems which affect safety performance.
2. Ask the same questions of managers and employees at five different levels in the organization.
3. Be easily and economically administered, analyzed, and evaluated without using a main-frame computer.
4. Allow identification and comparisons of specific departments and divisions while maintaining the respondent’s anonymity.
5. Provide managers with data in a format which will allow definitive comparisons and decision-making.

Survey verification study
With the approval of the AAR Steering Committee and the FRA, the Study Group recruited two additional railroads on which to test the revised survey plan. The top safety officers of these roads and their staffs worked with the Study Group in development and planning of a survey to fulfill the requirements of the established criteria.

A total of 1815 completed questionnaires were received from the two railroads, allowing a 95+ % confidence level with respect to inferences concerning a whole railroad and a statistical validity of 90+ % at the region and division levels.

| Number on Each Railroad Participating in Survey |
|-------------|-------------|-----------|
| Level       | RR III      | RR IV     |
| Executive   | 14          | 1         |
| Top Management | 21     | 6         |
| Middle Management | 184 | 16   |
| 1st Line Supervisor | 372 | 23 |
| Employee    | 878         | 165       |
| Unspecified | 128         | 17        |
| Totals      | 1597        | 228       |

Description of Analysis Program
Questions which had elicited a statistically significant response on the pilot survey were assigned to 20 categories, each of which defined a specific management system affecting safety. Each question was assigned to one or more categories for the purpose of identifying response patterns.

Analysis programs were written in BASIC for the IBM PC-XT. They allow the user to specify search criteria based on any combination of responses by railroad, or organizational unit within a railroad. The program also permitted the creation (either on an ad hoc basis or permanently) of additional categories or grouping of questions which appear to be meaningful. Two output formats are available:

1. A display of all question categories and response values for a specified combination of search criteria. It includes an optional listing of response totals for each individual question. One aggregated report is created for each inquiry.

2. A display of multiple reports for specific organization units in three question categories without detailed individual question counts.

Results from either report format could then be tested for statistical significance.

Data in reports generated was converted to easily read bar charts which allowed comparisons between departments, divisions, regions and railroads for each of the 20 categories. Other graphic presentations of differences in response between levels in the organization from top executives to employee and areas of high and low positive response were also developed. This information was then presented to top management on both railroads to aid in analysis of safety program effectiveness.

Analysis of survey data by management
Comparison of data from the survey with other indicators of safety performance confirmed that units with the highest positive response to survey questions were generally those with the best performance as measured by other indicators of safety performance.

The analyses of survey data provided by the AAR Study Group proved to be extremely useful to managements on the participating railroads. For instance, on one railroad, the survey clearly identified the weaker management systems which were affecting safety performance as:

- "Recognition For Good Safety Performance"—with a low 48 % favorable response from hourly-rated employees.
- "Inspections"—only 51 % favorable response.
- "Supervisor Training"—49% favorable response.
- "Quality of Supervision"—50% favorable response.

This cluster clearly indicated a major problem in perception of supervisory performance, and suggested that management needed immediate action to train supervisors in all aspects of safety performance with an emphasis on observations of people and conditions, and on positive reinforcement techniques.

On the same railroad, the widest difference in perception of program effectiveness between hourly-rated employees and management was in these same four categories. Not only did they have a serious problem, but they were largely unaware of the problem even existed.

One of the regions on one study railroad was consistently lower in positive hourly-rated response than all of the other regions in nearly every category. Management on this region clearly had a severe credibility problem with employees in nearly every phase of their safety efforts.

On one railroad, one region’s scores on “recognition for good safety performance” were significantly lower than those for any other region. When hourly-rated employee perception is at the 40% level, it means that six out of ten workers believe they are not receiving the kind of recognition due them for the job they are doing.

On another railroad, two categories were significantly worse than all of the others indicating clearly the point of attack for future safety efforts.

The survey even documented a problem with injury reporting on one of the railroads where the survey results did not correlate with the accident record.

Use of survey data
Ongoing use of the data included:

1. One of the railroads is using the survey results to develop performance objectives for individual managers and safety committees. The top executive has decided to personally review the findings with safety committees which did not meet their previous year objectives.

2. Safety officers are making use of the data for their respective responsibility areas to call manager’s attention to possible problems which require corrective action. Their involvement in the survey process provided them with a more balanced picture of the management activities which affect safety performance.

3. Because the survey provides a
Comparison of Positive Responses by Category — 4 Railroads

Management Effectiveness

Safety Climate

Safety Program Specifics

Safety Related Programs

Human Factor Oriented Activities

Comparison — 4 Study Roads

“picture” of the attitudes and beliefs of the organization at a specific point in time, the data base generated can be compared with data from future surveys to determine the effects of new safety initiatives. One of the railroads is currently re-surveying in selected areas to determine the effects of the training program conducted as a part of the study.

4. Other railroads using the survey to measure safety program effectiveness will have access to a large and complete data base with which to compare their data. The AAR Study Group is recommending making a “survey package” of software and instructions available to assist member roads in this regard.

Data from the surveys on all four railroads has been compiled and appears in the above figures which show all 20 categories.

Conclusions

As a result of this nine-year study, these conclusions were reached:

1. The effectiveness of safety efforts cannot be measured by traditional (procedural-engineering) criteria.

2. The effectiveness of safety efforts can be measured with surveys of employee (hourly to executive) perceptions.

3. A perception survey can effectively identify the strengths and weaknesses of elements of a safety system.

4. A perception survey can effectively identify major discrepancies in perception of program elements between hourly rated employees and levels of management.

5. A perception survey can effectively identify improvements in and deterioration of safety system elements if administered periodically.

The above conclusions, based on the data, carry with them considerable importance to safety management thinking. In effect, the data strongly suggests that those twelve elements described as a “procedural engineering” or a systems approach are not closely related to safety performance or results.

The twelve elements are not a bad description of the elements of many safety programs in use in industry today. This data does not question the audit approach to safety performance improvement. It does question the validity of any audit approach made up of arbitrarily decided elements such as these twelve. It does question the validity of any so-called “packaged audit.”

It supports rather strongly the belief that a perception survey as described in this article (sometimes called a “culture survey”), properly constructed, is a better measure of safety performance, and a much better predictor of safety results.

For more information on the study discussed in this article, write for a booklet recently published describing the entire study: Safety Section, Association of American Railroads, 50 F Street N.W., Washington D.C. 20001.

Please indicate how helpful and informative you found this article by circling the appropriate number on the reader service card.

Very 108
Moderately 107
Not very 106