Behavioral Safety Research in Manufacturing Settings: A Review of the Literature

Angelica C. Grindle
Alyce M. Dickinson
William Boettcher

ABSTRACT. Occupational safety continues to have large humanitarian and economic repercussions. This is particularly true in the manufacturing industry which has had the highest injury and illness rates for the past three years. Historically, attention was focused on determining the factors that correlated with safety. However, such approaches have

Angelica C. Grindle is affiliated with the Department of Psychology, Western Michigan University, Kalamazoo, MI. Alyce M. Dickinson, PhD is affiliated with the Department of Psychology, Western Michigan University, Kalamazoo, MI. William Boettcher is affiliated with Triad Performance Technologies, Inc.

Address correspondence to Alyce M. Dickinson, PhD, Department of Psychology, Western Michigan University, Kalamazoo, MI 49008 (E-mail: alyce.dickinson@wmich.edu or acgrindle@aol.com).

The authors extend appreciation to Wayne Fuqua, PhD and John Austin, PhD for their editorial assistance and to Ms. Courtney Rich for her assistance in obtaining articles. The authors also thank Reviewer A and Thomas C. Mawhinney, PhD for their help and insightful review of the manuscript.

Journal of Organizational Behavior Management, Vol. 20(1) 2000

fallen out of favor and the focus has shifted to manipulating the environment-engineering and behavioral safety. This manuscript reviews 18 behavioral safety programs implemented in manufacturing settings according to (a) settings, (b) subjects, (c) experimental design, (d) dependent variables, (e) intervention effectiveness, (f) miscellaneous effects, (g) maintenance, (h) integrity and reliability, and (i) social validity. Suggestions for future researchers are discussed. *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-342-9678. E-mail address: <getinfo@haworthpressinc.com> Website: <http://www.haworthpressinc. com>]*

KEYWORDS. Behavioral safety, research, literature, review, manufacturing settings, industrial settings

Occupational safety continues to be a large problem in the American workplace. In 1996 (the most recent statistics available), there were 6.2 million occupational injuries and illnesses in the private sector. This is equivalent to 7.4 injuries and illnesses per 100 equivalent full-time workers.[1, 2] Of those 6.2 million, 5.8 million resulted in job transfer, loss of consciousness, work or motion restriction, lost-time from work, or medical treatment beyond first-aid (United States Department of Labor [USDL], 1997, Dec.). Moreover, there were 6,112 occupational fatalities in 1996. This is equivalent to 17 workers killed per day while on the job; approximately one work related death every 86 minutes (USDL, 1997, Aug.).

The costs associated with work related injuries and illnesses are one of this nation's largest avoidable expenditures (Loafman, 1996). Besides the inestimable costs to the affected employees and their families in terms of pain and suffering, the actual monetary expenditures for unintentional occupational injuries and illnesses (i.e., injuries and illnesses that do not include homicides and suicides) totaled $121.0 billion in 1996. This figure includes wage and productivity losses ($60.2 billion), medical costs ($19.0 billion), administrative expenses ($25.6 billion), employer costs ($11.3 billion), damage to motor vehicles ($1.6 billion), and fire losses ($3.3 billion) (National Safety Council, 1997).

In the past few years, workers in the manufacturing industry have been at particular risk for occupational injuries and illnesses. Although, historically, the construction industry has had the highest occupational injury and illness rates,[3] in 1994, the incident rate in the manufacturing sector exceeded that of the construction industry for the first time in 20 years (USDL, 1995, Dec.). This trend continued throughout 1995 and 1996.

According to the United States Department of Labor, Bureau of Labor Statistics (1996, Dec.), the majority of 1996 non-fatal occupational injuries and illnesses occurred in manufacturing settings. For 1996, there were 10.6 injuries and illnesses per 100 full-time workers in manufacturing settings, a prevalence rate well above the national average of 7.4 per every 100 full-time workers (USDL, 1997, Dec.). In addition, the manufacturing industry accounted for three-fifths of newly reported 1996 occupational illnesses and its rate of restricted-work-activity cases was more than 200% the national average (USDL, 1997, Dec.).

Clearly, given the legal, monetary, and humanitarian implications, occupational safety is an important societal problem to which attention should be given. Furthermore, particular attention should be given to combat the increasing prevalence of injuries and illnesses in manufacturing settings.

APPROACHES TO IMPROVING OCCUPATIONAL SAFETY

Correlational Approaches

Historically, attention was focused on the determination of factors that correlated with occupational safety. As noted by Fitch, Hermann, and Hopkins (1976) and Komaki, Barwick, and Scott, (1978), for example, researchers have examined certain personality characteristics that were assumed to make workers more ‘‘prone’’ to injuries and illnesses. These personality characteristics include type A personalities and external safety locus of control (Evans, Palsane, & Carrere, 1987; Jones & Wuebker, 1993). Similarly, organizational factors that had high positive correlations with occupational injury and illness rates were determined. These factors included organizations that had seasonal layoffs, garnished employees’ wages, or were located in areas with easy access to prostitutes (Fitch et al., 1976). However, it is important to state that correlation is *not* causation. That is, even when certain correlations between personality characteristics or organizational factors are found, that does not mean that these personality characteristics or organizational factors *caused* the injuries and illnesses. It simply means that they are *associated* with them. As Skane (1985) stated, ‘‘Correlation analysis does not establish cause and effect, only the existence of a relationship'' (p. 281). Moreover, while identification of factors that correlate with occupational injury and illness rates may be of interest to some, it would be difficult, if not impossible, to change these factors in a manner that could have a positive impact on safety (McAfee & Winn, 1989). Thus, such approaches have fallen out of favor with safety professionals and the focus has shifted to the manipulation of the environment.

Environmental Approaches

Currently, there are two main methods in which the environment has been manipulated in order to improve safety: engineering and behavioral interventions. As described by Fitch et al. (1976), engineering interventions focus on, '' . . . reducing or eliminating physical hazards in the environment,'' while behavioral interventions attempt to, '' . . . change the behavior of the worker so that the interaction with the hazardous environment occurs in a safe fashion'' (p. 618). Although Fitch et al. restrict their description of behavioral interventions to changing the behavior of the *worker*, it should certainly be noted that behavioral interventions need to adopt a systems approach that focuses on altering and rewarding the behavior of staff at all levels. Appropriate organizational and performance management contingencies must be implemented by executives, managers, and supervisors to support and maintain safety. Thus, the management behaviors of organizational officials must be targeted for change as well–only then can workers improve and sustain their safety efforts. This system approach is nicely illustrated in an article by Sulzer-Azaroff, Loafman, Merante and Hlavacek (1990). Not only did these safety professionals pinpoint safety targets for the workers, but they also pinpointed the behaviors of the supervisors, managers and Director of Safety that were necessary for change to occur.

Safety Engineering. Examples of safety engineering include devices such as railings, mechanical guards, personal protective equipment, and ergonomically designed tools and equipment. Although safety engineering has been successful in reducing hazards in the environment, as Loafman (1996) discusses, there are three major problems with this approach. First, it would be extremely expensive and labor intensive to identify and rectify every possible hazardous condition in the work environment, if that were even possible. Second, safety engineering may foster unsafe reliance on artificial safety controls.

That is, workers may not attend to naturally occurring stimuli that indicate that behaving in a certain manner may result in an adverse consequence (i.e., an injury). They may assume, despite cues to the contrary, that if a guard or railing is not in place, a situation is safe. For example, a worker may presume that if a guard is not on a machine, it is safe to put his or her arm in the machine. Lastly, workers often ignore or circumvent safety devices. As Sulzer-Azaroff, Harris, and Blake-McCann (1994) stated, ''Injury analyses are replete with instances in which, despite training to the contrary, victims have returned to earlier deleterious habits or failed to use, circumvented, or misused proper precautions'' (p. 321). An explanation for why workers will ignore or circumvent engineering devices can be achieved by analyzing the natural contingencies for safe and unsafe performance that typically exist in the workplace.

At a general level, as several behavioral researchers and practitioners have discussed, the natural contingencies for safe and unsafe performance in the workplace can be examined in terms of the type, certainty, and immediacy of the consequences (e.g., Geller, 1996; Loafman, 1996; McSween, 1995; Sulzer-Azaroff, 1998). A safe or unsafe performance will be supported if it results in a reinforcing consequence (presentation of a positive consequence or escape from an aversive one) but will not be supported if it results in a punishing consequence (presentation of an aversive consequence or removal of a reinforcing consequence). Furthermore, consequences that are certain (probable) are more effective at supporting behavior than those that are uncertain (improbable). Finally, immediate, or relatively immediate, consequences are more effective at supporting behavior than those that are temporally delayed. A consideration of the type, certainty, and immediacy of the consequences for safe and unsafe performance in the workplace may explain why a worker performs unsafely despite extant engineering devices. This type of analysis is provided below.

First, serious injuries and illnesses are rare events in the workplace. A worker may perform a task in an unsafe manner hundreds or thousands of times without injury. Thus, although injury may be a punishing consequence, it is improbable. Furthermore, the use of engineering devices usually results in many immediate aversive consequences. For example, mechanical guards may slow down productivity which will be particularly problematic if workers are paid on a piece-rate basis; safety glasses may be uncomfortable or ugly, or make it difficult to see; and cut-proof gloves may make it difficult to handle a blade. Taken together, the natural consequences may support and encourage unsafe behaviors because: (a) performing in a safe manner results in immediate, probable, negative consequences such as discomfort and increased effort or time; while (b) performing in an unsafe manner rarely results in an injury but does result in immediate, probable positive consequences such as savings in time and effort and avoidance of discomfort.

The National Safety Council reports that human behavior is the cause of 94% of all injuries and illnesses (Loafman, 1996). Therefore, although safety engineering is effective in reducing occupational injuries and illnesses, it is insufficient because it does not directly address behavior–what people actually do on the job *even when engineering devices are available* (Loafman, 1996). As Sulzer-Azaroff et al. (1994) stated, ''It is recognized that even given the optimal in job and environmental design, people often act in harmful ways'' (p. 321). Thus, while there will always be a need for engineered solutions, the other side of safety, human behavior, must not be ignored. The consequences of safe and unsafe performance must be investigated and changed to obtain improvements above and beyond those that can be achieved by engineering efforts. This is why a behavioral approach to safety is critical.

Behavior as the Focus of Change. Behavioral safety is a specialty area within the field of Organizational Behavior Management, a field based on the application of the principles of Applied Behavior Analysis. Attention is focused upon what workers do on the job (behavior) and the contingencies of reinforcement (antecedents and consequences) that support or discourage safe behavior. Because of the continuing prevalence of injuries and illnesses in manufacturing settings, the present paper will review behavioral safety studies conducted in those settings. Nonetheless, it should certainly be noted that such programs have been successfully implemented in many other industries and settings, such as mining (e.g., Fox, Hopkins, & Anger, 1987), transportation (e.g., Haynes, Pine, & Fitch, 1982), vehicle maintenance (e.g., Komaki, Heinzmann, & Lawson, 1980), construction (e.g., Mattila & HyÖdynmaa 1988), and laboratories (e.g., Sulzer-Azaroff, 1978). Readers are encouraged to review these articles as well, as they provide evidence of the universality of the effectiveness of these programs.

METHOD

Criteria for Inclusion in Review

Potential articles were obtained from a computer database literature search (Psychinfo and ABI-INFORM) and by searching the bibliographies of behavioral safety articles. Those articles meeting the following four criteria were included in this literature review.

Behavioral Approach

Articles were included if they involved the manipulation of an independent variable to have an impact on safety related behavior (e.g., knees bent when lifting) or the results of behaviors, conditions, (e.g., aisles clear of obstacles). Not included were (a) programs designed to improve injury and illness rates without changing behaviors or conditions, (b) studies that attempted to improve safety indirectly (e.g., by decreasing drug use), (c) studies that examined correlations between injury and illness rates and personality characteristics or organizational factors, (d) theoretical articles, and (e) review articles.

Manufacturing Setting

Articles were included if they were conducted in settings where products were manufactured, processed, or assembled for subsequent sale. Programs in all other types of settings were excluded.

Occupational Safety

Articles focusing on the improvement of occupational safety were included. Excluded were programs designed to increase the off-the-job seat-belt wearing or occupational health and wellness (e.g., exercise, nutrition, smoking cessation).

Adequacy of Report

Articles were included if they were empirical reports that (a) included a sufficient amount of information to evaluate the independent variables, (b) reported effectiveness data, and (c) utilized a research design that allowed evaluation of the independent variable's effectiveness. Research designs considered acceptable included between-group designs and within-subject/group designs such as multiple-baseline designs and reversal designs. Before-after (AB) comparisons were included only if they also had a control group or were replicated in multiple settings, which would increase their validity.

In total, 18 studies met the inclusion criteria and were reviewed (see Table 1).[4,5] The review is divided into nine sections: (a) settings, (b) subjects, (c) experimental design, (d) dependent variables, (e) intervention effectiveness, (f) miscellaneous effects, (g) maintenance, (h) integrity and reliability, and (i) social validity.

LITERATURE REVIEW

Settings

Behavioral safety interventions have been effective in a wide variety of manufacturing settings. These organizations manufactured, processed or distributed products such as cellophane film, laboratory equipment, pastry products, freight wagons, telecommunication products, heat exchangers, paper, automobile sheet metal, farm machinery, ships, and textiles. In addition, these studies, although primarily conducted in the United States (n = 12), have been conducted in Israel (n = 3), Finland (n = 2), and the United Kingdom (n = 1). Information about the management structure, union representation, existing state of safety, and turnover rates within these settings varied, although the information was not provided in many studies. There appear to be few, if any, limits to the generality of the safety interventions with respect to the type of manufacturing setting.

Subjects

The number of subjects exposed to the 18 safety interventions is shown in Figure 1. Interventions were relatively large scale, with a median of 78 participants. The demographic characteristics of these subjects varied. Eight studies did not report subject gender. In 6 of the 10 remaining studies, the majority of participants were male; in 2, the majority were female, and in 2 there were an approximately equal

numbers of males and females. Subject ages, when reported, ranged from 22 to 60 years. Although of considerable interest due to potential interactions with safety programs, only four studies indicated how subjects were paid: In two, subjects received piece-rate pay (Laitinen, Saari, Kivistö, & Rasa, 1998; Saari & Näsänen, 1989), and in the other two subjects were paid hourly (Hopkins, Conard, Dangel, Fitch, Smith, & Anger, 1986; Komaki, Collins, & Penn, 1982). Given the types of jobs in the other studies, it is reasonable to assume that most of the subjects were paid hourly or piece-rate as well. Although the educational level and seniority of subjects were rarely reported, the educational level of subjects when mentioned was high school and seniority varied.

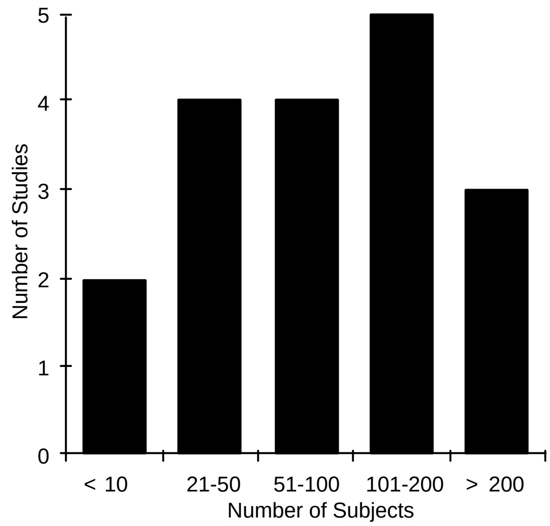
Reflecting the diversity of the settings, subjects held a variety of positions; they were assemblers, shipfitters, product testers, welders, gelcoaters, mill-workers, forklift operators, weavers, wrappers, machine-shop workers, evisceration workers, paint/sandblasters, sheet metal workers, and supervisors. Thus, as with the settings, there are few limits to the generality of the interventions in terms of subject characteristics or job tasks.

EXPERIMENTAL DESIGN

Figure 2 summarizes the experimental designs used. Four studies employed between-group designs (Cohen & Jensen, 1984; Ray et al., 1997; Zohar, 1980; Zohar, Cohen, & Azar, 1980).[4,5] Cohen and Jensen, however, were the only researchers who randomly assigned subjects to the experimental and control groups; the other three used existing groups, within or outside the organization, with similar demographic characteristics and job activities.

Nine studies used multiple-baseline designs where the introduction of one or more independent variables was staggered over time (Fellner & Sulzer-Azaroff, 1984, 1985; Hopkins et al., 1986; Komaki et al., 1978; Komaki et al., 1982; Reber & Wallin, 1984; Reber et al., 1990; Sulzer-Azaroff & DeSantamaria, 1980; Zohar & Fussfeld, 1981). These studies, with one exception, staggered the implementation across groups of subjects (e.g., departments, rooms, shifts, or supervisors). The exception, Fellner and Sulzer-Azaroff (1984), staggered the implementation across classes of behavior (i.e., practices versus conditions).

FIGURE 1. Number of subjects exposed to intervention per study.

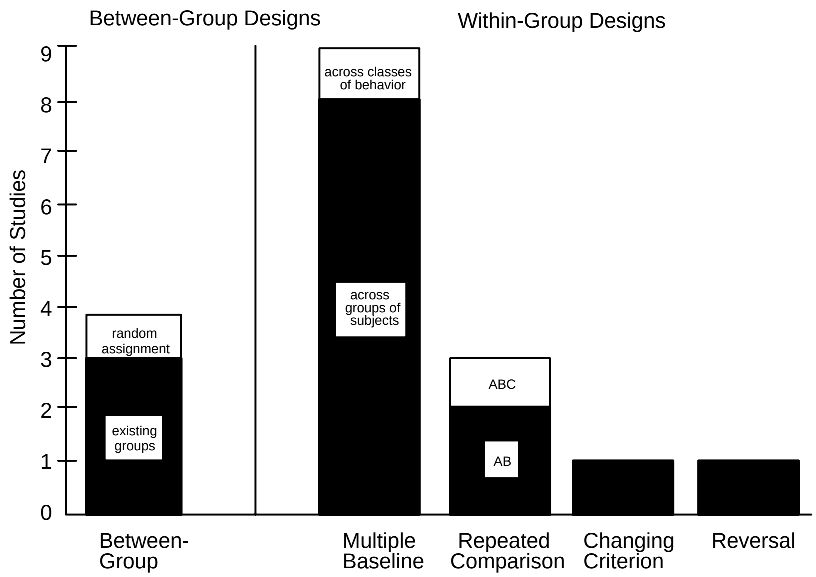


Note. M = 123 and Mdn = 78

Three other types of within-group designs have also been adopted. Cooper, Phillips, Sutherland, and Makin (1994), Laitinen et al. (1998) and Saari and Näsänen (1989) used comparison designs (AB or ABC) repeated in multiple departments, Sulzer-Azaroff, Loafman, Merante, and Hlavacek (1990) used a changing-criterion design, and Chhokar and Wallin (1984a) used a reversal design.

There are inherent problems with experimental control in occupational settings. Organizational officials, who are interested in the "bottom-line," may not be particularly interested in demonstrations of experimental control. Between-group designs may be particularly difficult to use because of the need for random assignment of subjects (Komaki, 1982). As exemplified by Ray et al. (1997), Zohar (1980), and Zohar et al. (1980), because of production, personnel, and/or union demands, organizational officials may require the use of extant groups and teams. Yet, without random assignment, the groups may differ from each other in significant ways (supervision, seniority, shift, current safety levels, job tasks, etc.) and improvements in the experimental group could reasonably be attributed to any one of the differ-

FIGURE 2. Types of experimental designs employed.



ences. In addition, organizational officials may simply be reluctant to implement a safety program for one group of employees, and not others. Given the political and professional problems associated with group designs and random assignment, future researchers should carefully consider the option of adopting a within-subject or within-group design.

Due to the paradigmatic preferences of behavioral researchers to use within-subject designs, combined, perhaps with the difficulties of using between-group designs, the majority of the researchers adopted within-subject designs. These designs eliminate the need for random assignment, which makes them a practical alternative to between-group designs. Because before-after (AB) comparisons do not eliminate alternative explanations, researchers who employ them may want to consider staggering the intervention across groups in a multiple-baseline fashion in order to better demonstrate experimental control. As before, however, organizational officials may not allow a safety program to be implemented for one group of workers and not others, even though the program would ultimately be implemented for all groups. In spite of this, staggered interventions may be an easier ‘‘sell’’ if the researcher explains that it is simply easier to initiate an intervention with a small number of employees rather than with all of them–especially if the organization is large. Organizational officials may readily agree to ‘‘start small’’ and then, over time, ‘‘roll the program out’’ to other units and departments; a practice that is quite common when organizations implement major initiatives.

Another alternative is to use a reversal design, but organizational officials may be reluctant to remove a program that is successful. Researchers must also consider the ethics of terminating, even temporarily, an effective program. Moreover, in some cases, it may not be possible to reverse, or completely withdraw, the intervention, which removes the possibility of proving its effectiveness. For example, safety training cannot be undone. And, although the formal aspects of the program may be removed, employees may surreptitiously use safety checklists if they found them valuable, or continue to support each other’s safe performances in less formal ways. Thus, it appears that the most reasonable design in organizational settings may be a multiple-baseline design. Regardless of what design is used, however, it is important for researchers to attempt to demonstrate experimental control.

DEPENDENT VARIABLES

Injury and Illness Rate

While the most obvious dependent variables in safety studies are injury and illness rates, as others have noted (Cohen & Jensen, 1984; Komaki et al., 1978; Sulzer-Azaroff et al., 1990), problems exist with their exclusive use. First, as McAfee and Winn (1989) stated, ‘‘ . . . accidents are rare events. Therefore, unless large samples of data are analyzed, such accidents may not be a sensitive dependent variable for measuring the success of a safety program’’ (p. 14). When there are not many accidents to begin with, decreases cannot be detected statistically and do not appear to be practically significant. Furthermore, injury and illness rates can vary depending upon seasonal production levels, which is particularly troublesome when using within-group designs because interventions may coincide with the seasonal fluctuations. Such fluctuations are sometimes controlled for by comparing injury and illness rates to those for the same time period in previous years (e.g., Fellner & Sulzer-Azaroff, 1984), but again, no statistical or practical differences may be found if rates are low to begin with.

Secondly, the reliability of injury and illness rates may be questioned because of changes in record keeping practices, and because employees may under-report the number and severity of injuries and illnesses (Chhokar & Wallin, 1984b). That is, observed decreases may not be actual decreases.

A third problem with injury and illness rates involves the inadvertent punishment of desirable behaviors. To illustrate, assume that valued rewards are provided to workers if reportable injuries or accidents have not occurred for a relatively long period of time. Employee reports of injuries toward the end of that interval are likely to be punished. Thus, *reporting* of injuries and illnesses may decrease but the actual injuries and illnesses may not. Certainly, such a decrease is not humanitarian as injuries could go untreated. Nor is such a decrease economically wise, as the long-term costs might far outweigh the short-term costs of immediate care (e.g., further injury, legal ramifications).

Practices and Conditions

Due to the above concerns, researchers have typically measured practices or conditions, alone or in conjunction with injuries and illnesses. A practice is a specific behavior (e.g., cutting in an outward motion), while a condition is the result of a set of behaviors (e.g., floor free of oil).

Safety targets. In order to determine the practices and conditions that would have the greatest impact on injuries and illnesses, 77% of the researchers conducted an initial assessment. Initial assessments usually included one or more of the following: (a) examination of injury and illness records; (b) review of company safety manuals, equipment handbooks, OSHA standards, and trade journals; (c) interviews of workers and supervisors; and (c) direct observation of workers. The specific methods of analysis, however, were not always provided or lacked detail. For purposes of replication, it would be helpful to have more information (readers should see Sulzer-Azaroff & Fellner, 1984, for a thorough treatment).

An interesting variation on the initial assessment was conducted by Hopkins et al. (1986). To determine the safety targets, researchers measured the amount of styrene in the air while employees performed their jobs in different ways. The safety targets were those practices that resulted in the lowest levels. Researchers also continued to measure and report the level of styrene in the air. In the remaining three studies (Zohar, 1980; Zohar et al., 1980; Zohar & Fussfeld, 1981), no formal initial assessments were conducted to determine the targets because the goal was to decrease hearing loss caused by noisy environments and thus the safety target was obvious–wearing earplugs.

The number of specific practices or conditions that were targeted in each study differed, as shown in Figure 3. Moreover, an extremely wide variety of safety targets were measured. The targets can be categorized into four main types: personal protective equipment (e.g., safety glasses, earplugs); material handling (e.g., cutting in outward motion); general safety (e.g., fire extinguishers in appropriate locations); and housekeeping (e.g., oil-free floor). Given the wide range of job types and settings, a detailed listing of the specific safety targets would be too cumbersome to provide here. Suffice it to say that one cannot help being impressed by the wide range of practices and conditions that have been targeted.

Final measures. Safety targets were combined into one overall safety score in almost all (88%) of the studies. Figure 4 displays the final safety measures that were used, and the number of studies using each measure. Figure 5 shows the ways in which the safety targets were combined into the overall scores. Only three studies (12%) used different measures: Sulzer-Azaroff and DeSantamaria (1980) used the frequency and type of hazards, Cohen and Jensen (1984) used the error rate (the percentage of incorrect behaviors), and Hopkins et al. (1986) measured the percent of 15-second observation intervals in which the target behaviors occurred.

Summary

Studies have used similar dependent variables. First, all primary measures have been based on behavior (e.g., practices, conditions). The rationale for the particular type of measure (e.g., percentage of safe practices vs. percentage of safe employees), however, was not provided. Such information would be beneficial. Second, although none of the studies used injury and illness rates as the primary dependent variable, presumably because of their inherent problems, nine did include them as secondary dependent variables (Cooper et al., 1994;

FIGURE 3. Number of safety targets per study.

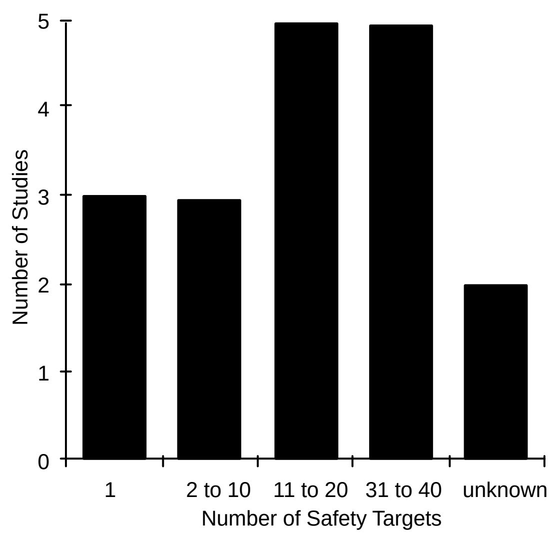


FIGURE 4. Number of measures per study.

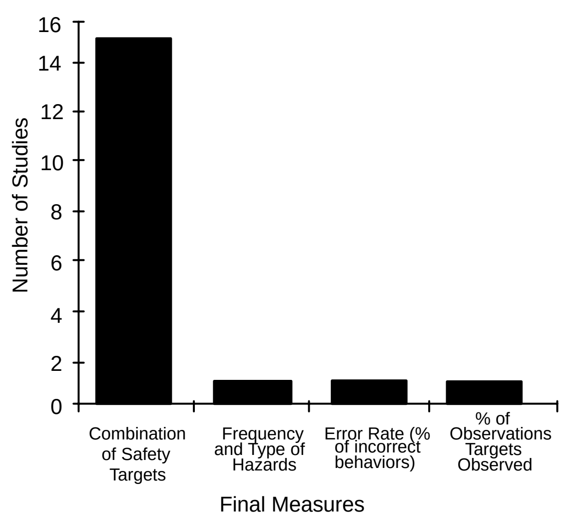
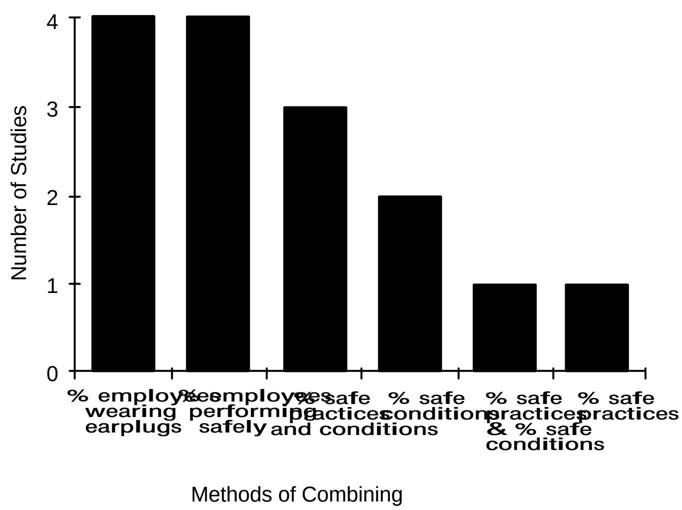


FIGURE 5. Methods of combining safety targets into overall scores.



Fellner & Sulzer-Azaroff, 1984, 1985; Komaki et al., 1978; Ray et al., 1997; Reber & Wallin, 1984; Reber et al., 1990; Saari & Näsänen, 1989; Sulzer-Azaroff et al., 1990). Their inclusion certainly is warranted, as the reduction of injuries and illnesses is the ultimate goal of safety programs. As noted earlier, Hopkins et al. (1986) adopted a very interesting measure. The purpose of this study was to decrease illness due to styrene exposure. Styrene-related illnesses, such as cancer, occur as a result of cumulative, repeated exposures, and symptoms may not present for months or years. Because of this, the rates of styrene-related illnesses are not useful dependent variables. While practices that led to the lowest levels of styrene in the air were the primary dependent variables, researchers also measured styrene air levels. This latter measure is particularly interesting because it provides an intermediate link between the behavioral practices and the later illnesses. Future researchers should consider using such measures, especially when the link between behavioral changes and their safety consequences are remote.

INTERVENTION EFFECTIVENESS

The 18 studies reviewed were categorized as (a) ‘‘singular’’ interventions (n = 4), (b) package programs (n = 6), or (c) component analyses (n = 8). The magnitude of effects of these interventions, in terms of percent improvement, is provided in Table 2. Data are presented for the percent improvement over baseline, and, for the component analyses, percent improvement over each preceding intervention. The percent improvement could not be determined for one package program (Hopkins et al., 1986) and one component analysis (Fellner & Sulzer-Azaroff, 1985), and thus these studies are not included on the table.

Singular Interventions

Researchers in two studies intervened with feedback (Fellner & Sulzer-Azaroff, 1984; Zohar et al., 1980) and two with token economies (Zohar, 1980; Zohar & Fussfeld, 1981). Percent improvements ranged from 9% to 157% over baseline. The same behavior, wearing ear plugs, was targeted in the three studies that resulted in the largest increases of 80%, 119%, and 157%. These stellar results may also be due to the fact that only one behavior was targeted.

Feedback. Fellner and Sulzer-Azaroff (1984) provided weekly graphic, numeric, and verbal feedback to employees. The percentage of safe practices and conditions improved 9% over baseline, and injury and illness rates decreased. Zohar et al. (1980) implemented a unique procedure to increase wearing ear plugs. They described the importance of wearing ear plugs and the hearing loss associated with excessive noise levels to employees in both the experimental and control groups. They then gave short-term hearing loss tests to employees in the experimental group before and after work two times during a one month period. Employees were instructed not to wear earplugs while they worked the first time they were tested and to wear them the second time. The test results were given to each individual employee and publicly posted. Earplug use increased significantly (85%-90%) for the experimental group and remained approximately 10% for the control group. Interestingly, in an attempt to increase earplug use in the control group, management subsequently implemented a disciplinary program. Employees were required to wear earplugs for gradually increasing lengths of time. If workers did not

TABLE 2. Summary of Magnitude of Effects of Behavioral Safety Systems

Singul Interventions			
			over baseline
Fellner & Sulzer-Azaroff (1984)			8.93%
Zohar (1980)			15.714%
Zohar et al. (1980)			118.75%
Zohar & Fussfeld (1981)			80.10%
Package Programs			
			over baseline
Cooper et al. (1994)			32.48%
Komaki et al. (1978)			32.40%
Laitinen et al. (1998)			56.14%
Sulzer-Azaroff & De Santamaria (1980)			59.40%
Sulzer-Azaroff et al. (1990)			37.05%
Component Analyses			
Chhokar & Wallin (1984)	over baseline	over training & goal-setting	over training, goal-setting, & weekly feedback
training & goal-setting	24.09%	–	–
training, goal-setting, & weekly feedback	45.04%	16.88%	–
training, goal-setting, & bi-weekly feedback	48.41%	19.60%	2.33%
Cohen & Jensen (1984)	over baseline		
training group	18.18%		
training & feedback group	22.86%		
Komaki et al. (1982)	over baseline	over antecedents	
antecedents	8.22%	–	
consequences	24.06%	14.86%	
Ray et al. (1996)	over baseline	over training	over training, & feedback
training	0.90%	–	–
training & feedback	12.66%	11.65%	–
training, feedback, & goal-setting	24.16%	23.05%	10.21%
Reber & Wallin (1984)	over baseline	over training	over training & goal-setting
training	13.91%	–	–
training & goal-setting	24.66%	9.44%	–
training, goal-setting, & feedback	53.7%	34.6%	23.0%
Reber et al. (1990)	over baseline	over training	over training & goal-setting
training	13.27%	–	–
training & goal-setting	24.30%	9.69%	–
training, goal-setting, & feedback	52.91%	34.75%	22.45%
Saari & Näsänen (1989)	over baseline	over feedback to supervisors	
feedback to supervisors	14.72%	–	
feedback to workers	37.19%	19.95%	

TABLE 3. Summary of the Intervention Components of the Package Programs

	Intervention Components					
Authors	Training	Feedback	Praise	Goal-Setting	Tangible Rewards	Partici-pation
Cooper et al. (1994)		×		×		
Hopkins et al. (1986)	×	×	×			
Komaki et al. (1978)	×	×	×	×		
Laitinen et al. (1998)	×	×				×
Sulzer-Azaroff & De Santamaria (1980)		×	×			
Sulzer-Azaroff et al. (1990)	×	×	×	×	×	

wear earplugs, they were taken off the job and docked pay. This program failed.

Token economy. In Zohar (1980) and Zohar and Fussfeld (1981), random tours were conducted daily and employees wearing earplugs were given tokens redeemable for merchandise. In the Zohar study, the tokens had a standard value. In the Zohar and Fussfeld study, the value of the tokens depended upon the percentage of employees wearing earplugs. Both procedures were highly effective. The percentage of employees wearing earplugs increased to 90% and 95%, respectively, representing increases over baseline of 157% and 80%.

Package Programs

In the six package programs, training, feedback, praise, goal-setting, and/or tangible rewards were combined. The components of these programs are summarized in Table 3. As can be seen from Table 2, the package programs have been very successful, with improvements ranging from 32% to 59% over baseline. The components that were functionally important cannot be isolated; nonetheless, the results of these studies are impressive. The six studies are detailed below.

Cooper et al.'s package (1994) consisted of participative goal-setting and weekly written and graphic feedback to workers. The percentage of safe behaviors and conditions improved significantly in 9 of 14 departments (ceiling effects prevented improvement in the remaining 5 departments). In addition, injury and illness rates decreased 21% overall, and accidents decreased 74%. Although there was an inverse relationship between the injury and illness rates and the percentages of safe behavior, it was not statistically significant. Komaki et al. (1978) implemented a similar program that combined training, assigned goal-setting, daily verbal praise, and graphic feedback that was provided four times a week. The percentage of safe practices and conditions increased dramatically when the program was implemented (32% over baseline) and decreased when it was withdrawn, demonstrating a functional relationship. Moreover, within a year, the incident rate of lost-time accidents decreased to less than 10 per million man-hours, and employees worked 280,000 hours without suffering a disabling injury. Whereas the organization had ranked last in safety within the corporation, with these gains, the organization ranked first. Sulzer-Azaroff et al.'s (1990) program was composed of training followed by weekly oral, written, and graphic feedback, praise, goal-setting, and monthly tangible rewards. Once again, the percentage of practices and conditions scored as safe increased dramatically (37% over baseline), and OSHA recordable and lost-time accidents decreased. Laitinen et al. (1998) implemented a comprehensive program based on participatory management, ergonomics, and behavioral safety. It included (a) training in ergonomics and housekeeping standards, (b) worker participation in the development of the program and generation of ideas for improvement, and (c) weekly graphic feedback to employees. The percentages of safe conditions increased 56% over baseline and the percentage of sick leave decreased from 12.8% to 9.9%.

While the preceding researchers intervened with groups of employees, Hopkins et al. (1986) intervened with individual workers. Workers were initially trained, and then observed twice a day. If workers were performing safely, they were praised, if not, they were prompted to perform safely. A majority of the safety targets improved and air levels of styrene declined 36%-57%. Sulzer-Azaroff and De-Santamaria's intervention (1980) targeted individual supervisors, rather than front-line staff. Written feedback, which included the number and location of hazards, and specific suggestions for improvement, were given to supervisors twice a week. Verbal and written praise were also provided. This program was highly successful: Both the frequency and type of hazardous conditions decreased dramatically

(59%). The researchers noted that, in most cases, supervisors had to rely on their workers to implement the suggestions, however, the ongoing behavior of the supervisors or the workers was not assessed.

Component Analyses

Training vs. training and feedback. Cohen and Jensen (1984) examined the effects of training alone versus training plus feedback on the performance of forklift drivers. In the first phase, they used a between group design: One group received training, a second received both training and feedback (daily verbal and posted), and the control group did not receive either. While training improved safety 18% over baseline levels, training plus feedback improved it more (23% over baseline), however, the difference was not statistically significant.

Training with feedback vs. goal-setting. Ray et al. (1997) examined the relative effectiveness of training, feedback, and goal-setting. In this study, baseline was followed by training. Feedback was then added in the form of a sign that displayed the average weekly safety index. In the final condition, assigned goal-setting was added. Training did not improve performance. Feedback improved performance 12% over training, and goal-setting improved performance 10% over training and feedback. Thus, goal-setting enhanced the effects of feedback. The performance of a control group did not change during this period of time. In addition, the injury record for the experimental group improved. The enhancing effects of goal-setting are consistent with those reported by Balcazar, Hopkins, and Suarez (1985/1986) in their classic review of feedback interventions.

Similar studies were conducted by Reber and Wallin (1984) and Reber et al. (1990), where training was provided in the first condition, assigned goal-setting added in the second, and feedback in the third. In both studies, goal-setting consisted of a sign posted with the goal on it. In addition, in Reber and Wallin, supervisors reminded workers of the goal weekly. In both studies, during the feedback phase, numeric scores were posted after each observation (approximately three times a week) and graphed weekly. The results of both programs were similar: (a) training improved performance 13% over baseline, (b) goal-setting improved safety 9%-10% over training, and (c) feedback improved safety 22%-23% over training and goal-setting. In neither study did workers consistently meet goals until feedback was provided. In addition to improvements in behaviors, OSHA recordable and lost-time rates showed marked decreases in both studies. Moreover, in the Reber and Wallin study, the inverse correlation between safe behaviors and injury and illness rates was statistically significant, unlike in the Cooper et al. (1994) study.

Training with goal-setting vs. feedback. Chhokar and Wallin (1984a) examined the relative effects of training with assigned goal-setting and graphic feedback. In addition, feedback frequency (once a week versus every two weeks) was examined. Training with goal-setting significantly improved performance 24% over baseline. Performance again improved when feedback was implemented (17% over training and goal-setting), and decreased when the feedback was withdrawn. No differences were found when the feedback was provided weekly or every other week. Similar to Reber and Wallin (1984) and Reber et al. (1990), workers did not consistently achieve their goals until feedback was provided. Feedback clearly enhanced the effectiveness of goal-setting in these studies, findings that are consistent with the general literature regarding goal-setting and feedback (Locke, Shaw, Saari, & Latham, 1981; Muchinsky, 1997).

Feedback vs. assigned goal-setting vs. participatory goal-setting. Fellner and Sulzer-Azaroff (1985) examined the relative effectiveness of participatory and assigned goal-setting. The study extended their previous one that examined the effects of feedback (Fellner & Sulzer-Azaroff, 1984). Feedback was provided during baseline. Publicly posted charts contained percentages of safe conditions and practices, a graph of the percentages, and specific hazards and their locations. In the assigned goal-setting phase, the foreman praised improvements, placed the goal on the graph, and informed the workers of the goal. During participatory goal-setting, the foreman praised improvements, asked for suggestions on how to improve areas that had not improved, and asked workers to suggest a goal. The goal was then placed on the graph. Results were highly variable across work areas. Overall, assigned goal-setting significantly increased the percentage of safe conditions but not safe practices. Even though workers met the goals significantly more often when they set them than when the supervisor assigned them (67% versus 47%, respectively), participatory goal-setting did not affect either conditions or practices. Thus, the results favored assigned over participatory goal-setting. Neither type of goal-setting affected injury or illness rates. Unlike the previous studies reviewed, goal-setting did not enhance the effects of feedback. Up-ward safety trends and ceiling effects in the feedback phases may have precluded further improvements, however.

Training and feedback to employees vs. to supervisors. Saari & Näsänen (1989) compared the effects of providing feedback to supervisors versus providing it directly to employees. In the first condition, all employees were trained, and the safety score was given to the foremen 1-3 times a week. In the second phase, graphic feedback was provided directly to employees 1-3 times a week in addition to the foremen. When foremen received the feedback performance improved 15% over baseline but improvements were limited to areas near the foremen's office. No improvements occurred in more remote areas. When feedback was provided directly to employees in addition to supervisors, safety improved another 20%. Significant decreases in injuries and illnesses occurred as well.

Antecedent vs. consequent. In 1980, Komaki, Heinzmann, and Lawson reported that feedback, as a consequence, improved safety more than the antecedent of training. When reviewing the results of that study, Komaki et al. (1982) stated:

> A close analysis of the above study, however, revealed an alternative explanation for the results. As in virtually all training programs, the information in the study was provided only once; safety was the subject of one training session but was not necessarily mentioned again. In contrast, during the feedback phase the graphs were updated three or more times a week and supervisors collected the information and provided feedback. Thus, the consequent control procedure's effectiveness may have been due to the greater frequency of stimuli changes and/or supervisor attention rather than the feedback per se. (p. 335)

Komaki et al. (1982) again examined the effects of antecedent versus consequent interventions, this time keeping supervisor involvement and stimulus changes constant across the two conditions. In the antecedent condition, employees were trained on safety rules, rules were posted, supervisors discussed the rules every week, and a new rule was highlighted 3 times a week. In the consequent condition, employees were trained on graph interpretation, supervisors discussed safety weekly, and graphic feedback was updated 3 times a week. During the antecedent phase, performance improved in only two of the four departments (an overall 8% improvement over baseline). When consequences were added, performance improved in all four departments, with an overall increase of 24% over baseline, and 15% over antecedents. As the authors stated, ''These results confirm that performance consequences such as feedback play a critical role in work motivation and that antecedents alone may not be effective in all cases, even if one can rely on fairly extensive supervisor involvement'' (p. 334).

Summary

Overall, behavioral safety interventions have been effective in improving safety, with consequent interventions proving more effective than antecedent interventions. That said, a few comments should be made that may be of use to future researchers. First, although there were a variety of interventions, the rationale for why a specific intervention was used was not specified in most studies. It appears the researchers simply employed a number of behavior change strategies that have been shown to be effective in other settings. Moreover, few researchers adequately tied the rationale to any concepts and principles of behavior analysis, in general, or to a behavioral analysis of safety, in particular.

Secondly, in some studies, although the interventions were successful at a general level, not all the safety targets were positively affected (Cohen & Jensen, 1984; Hopkins et al., 1986). For example, in Hopkins et al., wearing a respirator in the presence of styrene did not change, presumably because of the discomfort and inconvenience. Yet this is a critical behavior. In Cohen and Jensen study, forklift operators did not increase the extent to which they looked over their shoulders when driving in reverse. Again, failure to do so clearly has potentially hazardous outcomes. The authors stated that, '' . . . driving in reverse caused them to breathe in noxious fumes. Further, continuous looking over one's shoulder is an unnatural and uncomfortable posture to assume for prolonged periods'' (p. 132). Researchers should carefully examine the natural contingencies for the types of behaviors that are resistant to change, and modify their programs accordingly. Caution should also be taken when developing safety targets. It is not desirable to behaviorally engineer the breathing of noxious fumes or increase behaviors that could cause strain injuries. Engineering or behavioral alternatives may be available that would reduce the need for such behaviors. Behavioral and safety engineers should consult with employees who are likely to know much more about the job than they do and then work as a team to solve such problems.

Similarly, the effects of the interventions in the two Fellner and Sulzer-Azaroff studies (1984, 1985) were highly variable. The reasons for such variability need to be explored. In the 1984 study, improvements occurred only in those locations where employees paid greater attention to the feedback, and supervisors discussed the feedback more frequently. Fellner and Sulzer-Azaroff speculated that the effectiveness of feedback may be promoted by its examination and discussion, and research along these lines is warranted. It would also be of interest to determine if the type or frequency of feedback affects its power. In many of the studies feedback was provided in multiple forms (e.g., graphic, oral, and written) simultaneously, thus the relative effectiveness of feedback type cannot be ascertained. Furthermore, feedback was delivered daily, 3-4 times a week, weekly or bi-weekly. Although Chhokar and Wallin (1984a) found no differences in performance when feedback was provided weekly or bi-weekly, additional evaluations of feedback frequency would be of interest. In their review of the general effects of feedback, Balcazar et al. (1985/1986) stated that graphic feedback provided once a week was more effective than other types; however, their conclusion was based on across-study comparisons rather than experimental comparisons. Direct comparisons in the safety literature would certainly yield valuable information for safety practitioners as well as for others.

Thirdly, due to the possible ceiling effects in the Fellner and Sulzer-Azaroff (1985) study, the relative effectiveness of assigned versus participative goal-setting has yet to be established. Additionally, there are no objective guidelines with respect to how high a goal should be set. Most researchers mentioned setting ‘‘difficult yet attainable’’ goals, however, more specific and objective guidelines for setting safety goals in relation to baseline performance would be beneficial.

Finally, although some component analyses have been conducted, they are not sufficient. The component analyses suggest that (a) both feedback and goal-setting enhance training, (b) goal-setting enhances feedback, and (c) feedback enhances goal-setting. Thus, the most effective combination appears to be training, goal-setting and feedback. Additional research is needed, however. Moreover, few studies have investigated the effects of tangible rewards. Rather, the majority of studies have relied on feedback as a consequence. The rare exceptions were studies conducted by Sulzer-

Azaroff et al. (1990), Zohar (1980) and Zohar and Fussfeld (1981). This, in spite of Balcazar et al.'s (1985/1986) advice:

> If no system of functional, differential consequences exist, there is probably no point in establishing a feedback system. Effort would be better spent developing procedures for reinforcing wanted behaviors.
>
> If a feedback system is going to be established independently of careful considerations of the existence of functional, differential consequences . . . the evidence suggests that the best bets are to combine feedback that is graphically presented at least once a week with tangible rewards. Eighty percent of the studies with known effects that applied these characteristics were consistently effective regardless of whether goal setting procedures were additionally used. (p. 84)

The field would, thus, be well-served by experimental comparisons of feedback, goal-setting and tangible rewards.

MISCELLANEOUS EFFECTS

Injuries and Illnesses

Nine studies included measures of injury and illness rates (Cooper et al., 1994; Fellner & Sulzer-Azaroff, 1984, 1985; Komaki et al., 1978; Ray et al., 1997; Reber & Wallin, 1984; Reber et al., 1990; Saari & Näsänen, 1989; Sulzer-Azaroff et al., 1990). With the exception of Fellner and Sulzer-Azaroff (1985), all reported decreases in injury and illness rates following intervention. However, as discussed earlier, such results must be interpreted with caution. For example, Reber and Wallin described how record-keeping practices changed during the study, thus raising questions about the measure's reliability. Nevertheless, even though inherent difficulties exist with the use of injury and illness measures, researchers should still consider using them as ancillary measures while appropriately acknowledging their limitations.

Costs and Benefits

The costs of a safety program must not outweigh its benefits in order for it to be accepted initially and later maintained. Sulzer-Azaroff et al. (1990) reported the total cost savings of their intervention to be $55,000.00 and Cooper et al. (1994) reported that their program paid for itself. Reber and Wallin (1984) and Reber et al. (1990) reported ‘‘6 figure savings’’ and ‘‘substantial monetary savings,’’ respectively; however, the actual numbers were not provided. In the Fellner and Sulzer-Azaroff (1984, 1985) studies, most of the costs came from the development of the program, and because the running costs were low ($14.00 and $28.00 per week, respectively), substantial savings were assumed. Laitinen et al. (1998) reported the cost of their three-year program to be about $1.4 million. However, the program included many physical improvements that accounted for half of this expenditure. No data were provided on assumed savings. Zohar (1980) and Zohar and Fussfeld (1981) reported the costs of the tokens to be $15.00 and $10.00 per employee, respectively, which, as they pointed out, is less expensive than typical poster campaigns. Therefore, given that the average occupational injury or illness can run into the tens of thousands of dollars, substantial cost savings can be assumed.

There is an inherent obstacle in obtaining cost savings numbers. The measure compares the start-up and daily operation costs of the program from estimates of what the injuries and illnesses that *did not occur* would have cost. Given the problems previously discussed with injury and illness rate measures, any cost savings must be cautiously interpreted. Nevertheless, as stated, the costs of a safety program must not outweigh its benefits. Thus, cost/benefit analyses are an important ancillary measure and should be reported more frequently.

Productivity

Hopkins et al. (1986) obtained measures of productivity and time spent working before and after the intervention. When the program was first implemented, there was an initial decrease in productivity for 3 of the 4 workers. However, performance then rose above baseline levels. No changes were found in the time spent working. Similarly, Komaki et al. (1978) reported that there were no fluctuations in productivity due to their safety interventions. Sulzer-Azaroff and DeSantamaria (1980) did not obtain formal measures of productivity, but they anecdotally reported that productivity may have increased following the implementation of the intervention. The authors speculated that these productivity improvements may have occurred because of

(a) increased time on the job because injuries were reduced and (b) the ''safer'' arrangement of materials and tools. As discussed earlier, in applied settings decision-makers are usually not the researchers and it is certainly reasonable for them to be concerned about improving safety at the expense of productivity. Moreover, however unfortunate, safety professionals are often in conflict with production professionals regarding safety programs and expenditures. Thus, measures of productivity such as those used by Hopkins et al. are important to overcome such concerns, or at the very least, to determine the effects of a safety program on productivity. For these reasons, future researchers should obtain such measures when possible.

MAINTENANCE

Maintenance of the Program

Only 6 studies of the 18 studies mentioned whether the program was maintained in the organization after the study. Of those six, programs in four were maintained; programs in the other two were abandoned, however, similar programs were adopted in other units in the organization. The Sulzer-Azaroff and DeSantamaria (1980) program was kept in place with modifications (e.g., feedback was provided once a week instead of twice a week). The Komaki et al. (1978) program was maintained with modifications (e.g., graphs were posted once a week instead of four days a week) and expanded to include the other work shift after management saw the effects of the reversal phase. The Fellner and Sulzer-Azaroff (1984) feedback program was continued and later used as the site for the participatory versus assigned goal-setting study (e.g., Fellner & Sulzer-Azaroff, 1985). Similarly, the Sulzer-Azaroff et al. (1990) program was continued with other departments added. In two studies, Saari and Näsänen (1989) and Laitinen et al. (1998), the safety program was withdrawn from the units that participated in the study, perhaps to examine post-study/reversal effects, however similar programs were adopted in other units of the organization.

The reasons why these programs were continued and, presumably, not the others were not specified in most studies. Sulzer-Azaroff and her colleagues certainly deserve special recognition, however, because all of the Sulzer-Azaroff programs were maintained. It is likely that this maintenance is due to pre-program systems analyses. A prime example of this approach is provided in Sulzer-Azaroff et al. (1990). A comprehensive system analysis was conducted to insure that the program was compatible with other formal and informal systems within the organization. Not only were all of the program components based on this analysis, but specific pinpoints were identified and communicated to each key organizational official (supervisor, manager and director). Both researchers and practitioners are strongly encouraged to follow suit. While this approach is likely to improve maintenance, researchers should, nonetheless, examine the specific variables that are important in keeping a successful program in place. Organizational officials are not always data-driven: The decision to keep or disband a program may be made for reasons independent of the data, at least independent of the data that are collected as part of the safety program. Thus, determining the reasons why decision-makers choose to keep or remove a successful program would be beneficial in order to design programs that are maintained after the researchers leave the setting.

Maintenance of the Behavior Change

As indicated above, safety programs were maintained in only four studies. In two of the four, researchers reported whether performance was also maintained. Sulzer-Azaroff and DeSantamaria (1980) measured safety 3 days, 2 weeks, 6 weeks and 4 months after the experiment had been formally terminated. Safety remained excellent in all six participating departments. Komaki et al. (1978) reported that their results sustained and the injury rate continued to decline but neither the time period nor data were provided.

Six researchers conducted performance assessments after the program had been withdrawn. Five of the six assessments occurred six or fewer months after withdrawal (Cohen & Jensen, 1984; Laitinen et al., 1998; Zohar, 1980; Zohar et al., 1980; Zohar & Fussfeld, 1981). In all but one, program effects maintained. In the exception, Zohar and Fussfeld, the results maintained in all four departments after three months and in three of the four after six months. Saari and Näsänen (1989) were the only researchers to formally assess maintenance for longer periods of time. Performance remained stable in two units, A and B, for 22 and 13 months respectively. In addition, although no formal measures were obtained, Zohar anecdotally reported that most employees were still wearing earplugs one year after the program's termination.

In Sulzer-Azaroff and DeSantamaria (1980) and Komaki et al. (1978), the programs were maintained and the results maintained. This is, of course, good news, although more data of this kind are required to convincingly state that behavioral safety programs lead to sustained improvements. What is interesting, and more difficult to understand, however, is why performance maintained in sites where the safety programs were withdrawn. That is, by what mechanisms were the behaviors supported after program termination? In Zohar (1980), Zohar et al. (1980), and Zohar and Fussfeld (1981), rates of earplug use remained high at follow-up in spite of the fact that due to high turnover (60%, 65%, and 40%, respectively) large numbers of employees had never been exposed to the safety program. The authors provided three possible reasons. First, at the organizational level, new policies were implemented. Like punctuality and proscribed production levels, earplug use was made mandatory, and supervisors were responsible for maintaining it. Secondly, at the departmental level, new cultural norms regarding earplug use may have been established. Finally, at the employee level, the contrived program contingencies may have induced employees to wear ear plugs (in spite of the immediately punishing consequences), thus bringing employees in contact with the positive natural contingencies (reduction in noise and hearing loss) that maintained the behavior. This analysis, of course, is only relevant for workers who were initially exposed to the program. However, these workers may have prompted and praised ear plug use by newly hired workers. Saari and Näsänen (1989) speculated that the feedback (functioning as a reinforcer) on housekeeping outcomes (e.g., trash can empty) resulted in the outcomes themselves becoming conditioned reinforcers. Thus, once the graphs were no longer displayed, performance maintained because employees received feedback and reinforcement directly from the housekeeping outcomes. Laitinen et al. (1998) and Cohen and Jensen (1984) suggested that the performance levels may have maintained because of the participatory process that involved people from all levels of the organization in the development and execution of the programs. In addition, Cohen and Jensen hypothesized that maintenance may have been due to safer habits and new group norms sustained by peer modeling and management support.

It is important to demonstrate that behavioral interventions improve safety, but it is also important to demonstrate their effectiveness over long periods of time. In addition, the organizational, group, and individual mechanisms that maintain safety after programs are formally withdrawn should be determined. Only then will it be possible to *systematically* design safe ''cultures'' and develop new group ''norms.''

INTEGRITY AND RELIABILITY

Integrity of the Independent Variable

Intervention integrity measures were provided in only a few studies. Three examined the integrity of training by measuring employees' knowledge of the training material (Chhokar & Wallin, 1984a; Reber & Wallin, 1984; Reber et al., 1990). Performance tests indicated that training was successful. Subjects in the Ray et al. (1997) study demonstrated their knowledge during the training session, although no data were provided. Sulzer-Azaroff and DeSantamaria (1980) conducted sessions with supervisors to ensure that they understood the feedback form, however, no measures were taken. Sulzer-Azaroff et al. (1990) informally assessed integrity of the independent variables.

Reported problems typically involved the behaviors of supervisors and observers. For example, Komaki et al. (1978) and Cooper et al. (1994) indicated that supervisors did not deliver praise as planned. Similarly, Komaki et al. (1982) noted that supervisors attended fewer and fewer feedback meetings over time. In addition, they found that the employees who did not prefer the consequent condition over the antecedent condition did not understand the graphs. Cooper et al. also reported that observers did not gather data in some weeks. In Zohar et al. (1980), the observer for the control group initially inflated the data and a double-monitoring system had to be put into place. These problems suggest that researchers should take steps to ensure the integrity of independent variables. Special attention should be focused on those who have direct responsibility for implementation: training and monitoring are essential. Moreover, a behavioral safety system must be incorporated into the existing management structure for it to maintain once the researchers leave the organization. Without implementing internal controls to ensure integrity, a behavioral safety program may not maintain over time, regardless of its initial success.

Reliability of the Dependent Variable

Behavioral safety researchers did, for the most part, assess the reliability of the dependent variables, typically in terms of interobserver agreement. Four studies did not report reliability measures (Cooper et al., 1994; Laitinen et al., 1998; Ray et al., 1997; Zohar, 1980), and Sulzer-Azaroff et al. (1990) assessed reliability only informally. Zohar et al. (1980) reported initial reliability problems, however, interobserver agreement was high in the other studies, ranging from 83%-100%. In addition to measuring the reliability of their behavioral targets, Hopkins et al. (1986), commendably, assessed the reliability of the styrene air level measures and productivity measures.

SOCIAL VALIDITY

Social validity (the evaluation of the acceptability of a program by its consumers) is important not only to increase ‘‘buy-in’’ for programs but also to decrease resistance to it and, possibly, increase worker morale. The three main types of social validity include the acceptance of (a) the goals of the program, (b) the procedures employed, and (c) the outcomes of the program (Schwartz & Baer, 1991).

Goals

By its nature, the goals of behavioral safety programs are socially valid. That is, given the rates of occupational injuries and fatalities and the monetary costs involved, few would argue that the improvement of safety is not a worthy goal.

Procedures

A number of studies examined the second type of social validity–the acceptance of the procedures employed. During the course of the study, Chhokar and Wallin (1984a), Cooper et al. (1994), Reber and

Wallin (1984), and Reber et al. (1990) used a questionnaire to determine the acceptance of the program and found that it was acceptable to employees. Laitinen et al. (1998) conducted before and after assessments of the perceived physical and psychosocial (e.g., cooperation, solidarity, and support) working conditions. They found both improved significantly. Others examined the social validity of the procedures *after* the completion of the study. For example, Komaki et al. (1982) assessed employee preference for the antecedent versus the consequent condition, and found that 72% of the employees preferred the consequent condition. Interestingly, as mentioned earlier, they also found that those who preferred the antecedent condition also reported that they did not understand the feedback graphs. In Fellner and Sulzer-Azaroff (1985), employees, responding to a questionnaire, said that they were indifferent to the feedback and goal-setting condition. That notwithstanding, employees indicated that they preferred to set their own goals rather than to have supervisors assign them, even though participative-goal setting was not more effective than assigned goal-setting. Participants in Saari and Näsänen's (1989) study rated the interventions positively. Although Sulzer-Azaroff et al. (1990), Komaki et al. (1978), and Ray et al. (1997) did not formally assess employee acceptance, they anecdotally reported positive reviews of their programs. In fact, both Komaki et al. (1978) and Sulzer-Azaroff et al. (1990) stated that workers would cheer when the new data point was added to the graph. Similarly, Ray et al. reported favorable responses to feedback, and indicated that employees would suggest additional ways to improve safety during feedback sessions.

Outcomes

Sulzer-Azaroff et al. (1990) were the only researchers who reported the acceptance of outcomes, and they did so only anecdotally, noting that the safety director said, ''The program was fantastic. I never dreamed people would be so successful'' (p. 118). Laitinen et al. (1998) hinted at outcome acceptance when they discussed the initial difficulties in obtaining funding for the program: They reported that as the effects in the first department became known, funding was no longer a problem, a clear indication of management's acceptance of the outcomes. Outcome acceptance can also be inferred in those studies where the programs were maintained or expanded. Thus, although it appears that behavioral safety programs are viewed favorably in terms of their procedures, more pervasive use of formal and objective social validity measures of the procedures and the outcomes would be of interest. They would certainly be of value when marketing behavioral safety programs to new organizations.

CONCLUSION

Behavioral safety researchers have demonstrated the effectiveness of a variety of behavioral interventions in a wide variety of manufacturing settings with many different jobs. Although we have restricted our review to applications in manufacturing settings, our conclusions do not differ from those who have examined behavioral applications in other settings. For example, McAfee and Winn (1989) reviewed the results of 24 behavioral safety programs that were implemented between 1971 and 1987. Settings that were represented in their review included human service settings, coal mines, packaging forwarding facilities, city maintenance divisions, city refuse divisions, urban transit, textile weaving mills, and public safety departments. Upon finding this diversity, McAfee and Winn stated, ''In the 24 studies, 20 different job classifications are represented. Certainly, researchers can't be criticized for limiting their studies to only a narrow range of jobs or industries'' (p. 9). They concluded that:

> The major finding was that every study, without exception, found that incentives or feedback enhanced safety and/or reduced accidents in the work place, at least in the short term. Few literature reviews find such consistent results. Although this may be surprising to some, others might argue that this finding is simply further proof of the law of effect which contends that rewarded behavior tends to be repeated. (p. 15)

Perusal of recent studies conducted in settings other than manufacturing (e.g., Austin, Alvero, & Olson, 1998; Austin, Kessler, Riccobono, & Bailey, 1996; Laitinen & Ruohomaki, 1996; Sulzer-Azaroff, in press) and case studies in recent behavioral safety texts (Geller, 1996; McSween, 1995) also adds credence to the generality of our conclusions. Our selected review permitted us to highlight the success of behavioral interventions in an industry where risk of injury is high. It also permitted us to provide a more detailed analysis than a more comprehensive review would have permitted.

While the success rate of behavioral safety interventions is high, questions remain: Questions, that if answered, would lead to continued development of our work in safety. We do not yet conclusively know which independent variables are most important, nor how they relate to a behavioral analysis of safe performance. We encourage safety professionals to conduct functional analyses prior to intervention. In addition, we have not addressed ways to modify those behaviors that appear resistant to change when feedback and/or goal-setting fail. Finally, we need to identify the factors that lead to long-term program maintenance and performance change. This latter suggestion may well be our most important task. Given the humanitarian and economic importance of occupational safety, we encourage additional research that will help determine the most effective and efficient methods that will lead to long-term safety in organizations. We hope that this review will be a spring board to that end.

NOTES

1. An Occupational Injury is, "any injury such as a cut, fracture, sprain, amputation, etc., which results from a work accident or from an exposure involving a single incident in the work environment" (Occupational Safety and Health Administration [OSHA]).

2. An Occupational Illness is, "any abnormal condition or disorder, other than one resulting from an occupational injury, caused by exposure to environmental factors associated with employment. It includes acute and chronic illnesses or diseases which may be caused by inhalation, absorption, ingestion, or direct contact" (OSHA).

3. According to the Bureau of Labor Statistics (1997, Dec.), the incident rates represent, "the number of injuries and illnesses per 100 full-time workers and were calculated as: (N/EH) x 200,000, where N = number of injuries and illnesses; EH = total hours worked by all employees during the calendar year; 200,000 = base for 100 equivalent full-time workers (working 40 hours per week, 50 weeks per year.)" (Table 1.)

4. Two separate studies are discussed in the Cohen and Jensen (1984) manuscript. The current paper only reviews study one in Cohen and Jensen.

5. Two separate studies are discussed in the Zohar (1980) manuscript. The current paper only reviews study two in Zohar.

REFERENCES

Austin, J., Alvero, A. M., & Olson, R. (1998). Prompting patron safety belt use at a restaurant. *Journal of Applied Behavior Analysis, 31*, 655-657.

Austin, J., Kessler, M. L., Riccobono, J. E., & Bailey, J. S. (1996). Using feedback and reinforcement to improve the performance and safety of a roofing crew. *Journal of Organizational Behavior Management, 16*, 49-75.

Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985/1986). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7,* 65-89.

Chhokar, J. S., & Wallin, J. A. (1984a). A field study on the effect of feedback frequency on performance. *Journal of Applied Psychology, 69,* (3), 524-530.

Chhokar, J. S., & Wallin, J. A. (1984b). Improving safety through applied behavior analysis. *Journal of Safety Research, 15,* 141-151.

Cohen, H. H. & Jensen, R. C. (1984). Measuring the effectiveness of an industrial lift truck safety training program. *Journal of Safety Research, 15,* 125-135.

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,* 219-240.

Evans, G. W., Palsane, M. N. & Carrere, S. (1987). Type A behavior and occupational stress: A cross-cultural study of blue-collar workers. *Journal of Personality & Social Psychology, 52(5),* 1002-1007.

Fellner, D. J., & Sulzer-Azaroff, B. (1984). Increasing industrial safety practices and conditions through posted feedback. *Journal of Safety Research, 15,* 7-21.

Fellner, D. J., & Sulzer-Azaroff, B. (1985). Occupational Safety: Assessing the impact of adding assigned or participative goal-setting. *Journal of Organizational Behavior Management, 7(1/2),* 3-24.

Fitch, H. G., Hermann, J., & Hopkins, B. L. (1976). Safe and unsafe behavior and its modification. *Journal of Occupational Medicine, 18(9),* 618-622.

Fox, D. K., Hopkins, B. L., & Anger, W. K. (1987). The long-term effects of a token economy on safety performance in open-pit mining. *Journal of Applied Behavior Analysis, 20,* 215-224.

Geller, S. E. (1996). *Working safe: How to help people to actively care for health and safety.* Radnor, PA: Chilton.

Haynes, R. S., Pine, R. C., Fitch, H. G. (1982). Reducing accidents rates with organizational behavior modification. *Academy of Management Journal, 25*(2), 407-416.

Hopkins, B. L, Conard, R. J., Dangel, R. F., Fitch, H. G., Smith, M. J., & Anger, W. K. (1986). Behavioral Technology for Reducing Occupational Exposures to Styrene. *Journal of Applied Behavior Analysis, 19*(1), 3-11.

Jones, J. W. & Wuebker, L. J. (1993). Safety locus of control and employees' accidents. *Journal of Business and Psychology, 7*(4), 449-457.

Komaki, J. L. (1982). The case for the single case: Making judicious decisions about alternatives. In L. W. Frederiksen (Ed.), *Handbook of organizational behavior management* (pp. 145-176). New York: John Wiley.

Komaki, J., Barwick, K. D., & Scott, L. R. (1978). A Behavioral Approach to Occupational Safety: Pinpointing and Reinforcing Safe Performance in a Food Manufacturing Plant. *Journal of Applied Psychology, 63*(4), 434-445.

Komaki, J., Collins, R. L., & Penn, P. (1982). The Role of Performance Antecedents and Consequences in Work Motivation. *Journal of Applied Psychology, 67*(3), 334-340.

Komaki, J., Heinzmann, A. T., & Lawson, L. (1980). Effect of training and feedback:

component analysis of a behavioral safety program. *Journal of Applied Psychology*, *65*, 261-270.

Laitenen, H., & Ruohomaki, I. (1996). The effects of feedback and goal setting on safety performance at two construction sites. *Safety Science*, *24*, 61-73.

Laitinen, H., Saari, J., Kivistö, M., & Rasa, P. (1998). Improving physical and psychosocial working conditions through a participatory ergonomic process: A before-after study at an engineering workshop. *International Journal of Industrial Ergonomics*, *21*, 35-45.

Loafman, B. (1996). Rescue from the safety plateau. *Performance Management Magazine*, *14*(3), 3-10.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin*, *90*, 125-152.

Mattila, M. & HyÖdynmaa, M. (1988). Promoting job safety in building: An experiment on the behavior analysis approach. *Journal of Occupational Accidents*, *9*, 255-267.

McAfee, R. B. & Winn, A. R. (1989). The use of incentives/feedback to enhance work place safety: A critique of the literature. *Journal of Safety Research*, *20*(1), 7-19.

McSween, T. E. (1995). *The values-based safety process: Improving your safety culture with a behavioral approach*. New York: Van Nostrand Reinhold.

Muchinsky, P. M. (1997). *Psychology applied to work* (5th edition). New York: Brooks/Cole.

National Safety Council. (1997). *Accident Facts, 1996 Edition*. Itasca, IL: Author.

Occupational Safety and Health Administration. *Instructions for OSHA no. 200*. Washington D. C.: Author.

Ray, P. S., Bishop, P. A., Wang, M. Q. (1997). Efficacy of the components of a behavioral safety program. *International Journal of Industrial Ergonomics*, *19*, 19-29.

Reber, R. A., & Wallin, J. A. (1984). The effects of training, goal-setting, and knowledge of results on safe behavior: a component analysis. *Academy of Management Journal*, *27*(3), 544-560.

Reber, R. A., Wallin, J. A., & Chhokar, J. S. (1990). Improving safety performance with goal-setting and feedback. *Human Performance*, *3*(1), 51-61.

Saari, J. & Näsänen, M. (1989). The effect of positive feedback on industrial housekeeping and accidents: a long term study at a shipyard. *International Journal of Industrial Ergonomics*, *4*, 201-211.

Schwartz, I. S. & Baer, D. M. (1991). Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis*, *24*, 189-204.

Skane, D. H. (1985). *Elementary Statistics*. Reading, MA: Addison-Wesley.

Sulzer-Azaroff, B. (1978). Behavioral ecology and accident prevention *Journal of Organizational Behavior Management*, *2*, 11-44.

Sulzer-Azaroff, B. (1990). Strategies for maintaining change over time. *Behaviour Change*, *7*(1), 3-15.

Sulzer-Azaroff, B. (1998). *Who killed my daddy? A behavioral safety fable*. Cambridge, MA: Cambridge Center for Behavioral Studies.

Sulzer-Azaroff, B. (in press). Behavioral safety for the twenty-first century. In H.

Laitinen and J. Saari (Eds.), *People and work*. Finnish Department of Occupational Health: Division of Occupational Safety.

Sulzer-Azaroff, B. & DeSantamaria, M. C. (1980). Industrial Safety Hazard Reduction through Performance Feedback. *Journal of Applied Behavior Analysis*, *13*(2), 287-295.

Sulzer-Azaroff, B. & Fellner, D. J. (1984). Searching for performance targets in the behavior analysis of occupational safety: An assessment strategy. *Journal of Organizational Behavior Management*, *6*(2), 53-65.

Sulzer-Azaroff, B., Harris, T. C., & Blake-McCann, K. (1994). Beyond training: organizational performance management techniques. *Occupational Medicine 9*(2), 321-339.

Sulzer-Azaroff, B., Loafman, B., Merante, R. J., & Hlavacek, A. C. (1990). Improving Occupational Safety in a Large Industrial Plant: A Systematic Replication. *Journal of Organizational Behavior Management*, *11*(1), 99-120.

United States Department of Labor, Bureau of Labor Statistics. (1995, December). *Workplace injuries and illnesses in 1994.* Technical Report #USDL-95-508.

United States Department of Labor, Bureau of Labor Statistics. (1997, August). *National census of fatal occupational injuries, 1996.* Technical Report #USDL-97-266.

United States Department of Labor, Bureau of Labor Statistics. (1997, December). *Workplace injuries and illnesses in 1996.* Technical Report #USDL-97-453.

Zohar, D. (1980). Promoting the use of personal protective equipment by behavior modification techniques. *Journal of Safety Research*, *12*(2), 78-85.

Zohar, D., Cohen, A., & Azar, N. (1980). Promoting increased use of ear protectors in noise through information feedback. *Human Factors*, *22*(1), 69-79.

Zohar, D., & Fussfeld, N. (1981). Modifying earplug wearing behavior by behavior modification techniques: an empirical evaluation. *Journal of Organizational Behavior Management*, *3*(2), 41-52.

TABLE 1. Magnitude of Effects of Behavioral Safety Interventions

Author	Product	Intervention components	Results: Safety Dependent Variables	Results: Productivity	Results: Cost/Benefit	Maintenance: Performance	Maintenance: Program	Social Validity*: Procedures	Social Validity*: Outcome
			Singular Interventions						
Fellner & Sulzer-Azaroff (1984)	paper napkins & tissues	graphic, numeric, & verbal feedback	Practices & conditions significantly improved. Injuries & illnesses decreased.		assumed		Yes-used in 1985 study		I
Zohar (1980)	textiles	token economy-token value standard	Earplug wearing rates significantly improved.		assumed	Yes			
Zohar et al. (1980)	metal fabrication	individual & posted feedback on short-term hearing loss	Earplug wearing rates significantly improved.			Yes			
Zohar & Fussfeld (1981)	textiles	token economy-token value based on group performance	Earplug wearing rates significantly improved.		assumed	Yes for 3/4 depts.			
			Package Programs						
Cooper et al. (1994)	cellophane film	participative goal-setting; written & graphic feedback	Practices & conditions significantly improved in 9 of 14 departments. Injuries & illnesses decreased.		paid for itself			X	
Hopkins et al. (1986)	laboratory equipment	training; verbal praise; verbal prompts	Practices improved. Styrene air levels decreased.	initial decrease followed by improvements					
Komaki et al. (1978)	pastry products	training; assigned goal-setting; graphic feedback; verbal praise	Practices & conditions improved when implemented and decreased when withdrawn. Injuries & illnesses decreased	no detrimental effect		Yes	Yes with modifications and expanded	A	I
Laitinen et al. (1998)	freight wagons	graphic feedback; ergonomic training; participatory process	Conditions significantly improved. Sick leave rates decreased.		unknown savings	Yes	No, but adopted in other units	X	I
Sulzer-Azaroff & De Santamaria (1980)	custom fabricated products	written feedback; verbal & written praise	Hazard frequency decreased.	anecdotally improved		Yes	Yes with modifications		I
Sulzer-Azaroff et al. (1990)	telecommunication products	training; oral, written, & graphic feedback; praise; goal-setting; tangible rewards	Practices & conditions improved. Injuries & illnesses decreased.		$55,000 savings		Yes and expanded	A	A

*X = assessed; A = anecdotally; I = inferred

TABLE 1 (continued)

Author	Product	Intervention components	Results: Safety Dependent Variables Component Analyses	Productivity	Cost/ Benefit	Maintenance: Perfor- mance	Maintenance: Program	Social Validity*: Proce- dures	Social Validity*: Out- come
Chhokar & Wallin (1984)	heat exchangers	training & assigned goal-setting vs. training, goal-setting and graphic feedback (weekly vs. every 2 weeks)	Training, goal-setting & feedback was more effective than training & goal-setting. No differences in performance as a result of feedback frequency.					X	
Cohen & Jensen (1984)	distribution warehouse	training vs. training & feedback (verbal & posted)	Training with feedback was more effective than training alone.			Yes			
Fellner & Sulzer-Azaroff (1985)	paper napkins & towels	assigned goal-setting vs. participatory goal-setting	Assigned goal-setting was more effective than participative goal-setting. No decreases in injuries beyond those achieved by previous feedback study.		assumed			X	
Komaki et al. (1982)	poultry	antecedents vs. consequences	Consequent intervention was more effective than antecedent intervention					X	
Ray et al. (1997)	automobile sheet metal	training vs. training & feedback (numeric) vs. training, feedback, & assigned goal-setting	Training with feedback was more effective than training alone. Training with feedback & goal-setting was more effective than training & feedback. Injuries & illnesses decreased.					A	
Reber & Wallin (1984)	farm machinery	training vs. training & goal-setting vs. training, goal-setting, & feedback (numeric & graphic)	Training with goal-setting was more effective than training alone. Training with goal-setting & feedback was more effective than training and goal-setting. Injuries & illnesses decreased.		6 figure savings			X	
Reber et al. (1990)	farm machinery	training vs. training & goal-setting vs. training, goal-setting & feedback (numeric & graphic)	Training with goal-setting was more effective than training alone. Training with goal-setting & feedback was more effective than training & goal-setting. Injuries & illnesses decreased.		"sub-stantial"			X	
Saari & Näsänen (1989)	ships	feedback to supervisors vs. feedback to workers	Feedback to workers was more effective than feedback to supervisors. Injuries & illnesses decreased.			Yes	No, but adopted in other units	X	I

*: = assessed; = anecdotally; I = inferred

numbers of males and females. Subject ages, when reported, ranged from 22 to 60 years. Although of considerable interest due to potential interactions with safety programs, only four studies indicated how subjects were paid: In two, subjects received piece-rate pay (Laitinen, Saari, Kivistö, & Rasa, 1998; Saari & Näsänen, 1989), and in the other two subjects were paid hourly (Hopkins, Conard, Dangel, Fitch, Smith, & Anger, 1986; Komaki, Collins, & Penn, 1982). Given the types of jobs in the other studies, it is reasonable to assume that most of the subjects were paid hourly or piece-rate as well. Although the educational level and seniority of subjects were rarely reported, the educational level of subjects when mentioned was high school and seniority varied.

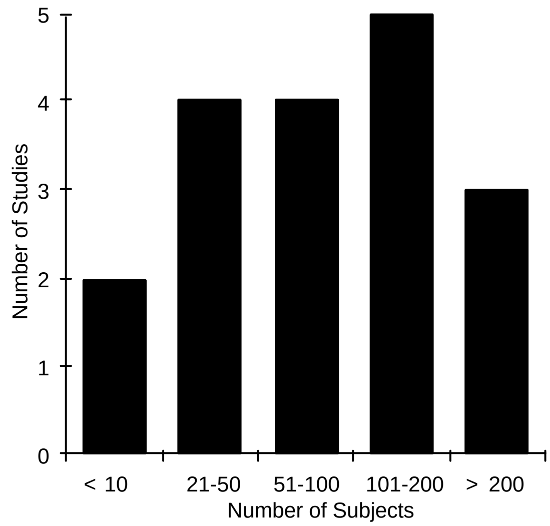
Reflecting the diversity of the settings, subjects held a variety of positions; they were assemblers, shipfitters, product testers, welders, gelcoaters, mill-workers, forklift operators, weavers, wrappers, machine-shop workers, evisceration workers, paint/sandblasters, sheet metal workers, and supervisors. Thus, as with the settings, there are few limits to the generality of the interventions in terms of subject characteristics or job tasks.

EXPERIMENTAL DESIGN

Figure 2 summarizes the experimental designs used. Four studies employed between-group designs (Cohen & Jensen, 1984; Ray et al., 1997; Zohar, 1980; Zohar, Cohen, & Azar, 1980).[4,5] Cohen and Jensen, however, were the only researchers who randomly assigned subjects to the experimental and control groups; the other three used existing groups, within or outside the organization, with similar demographic characteristics and job activities.

Nine studies used multiple-baseline designs where the introduction of one or more independent variables was staggered over time (Fellner & Sulzer-Azaroff, 1984, 1985; Hopkins et al., 1986; Komaki et al., 1978; Komaki et al., 1982; Reber & Wallin, 1984; Reber et al., 1990; Sulzer-Azaroff & DeSantamaria, 1980; Zohar & Fussfeld, 1981). These studies, with one exception, staggered the implementation across groups of subjects (e.g., departments, rooms, shifts, or supervisors). The exception, Fellner and Sulzer-Azaroff (1984), staggered the implementation across classes of behavior (i.e., practices versus conditions).

FIGURE 1. Number of subjects exposed to intervention per study.

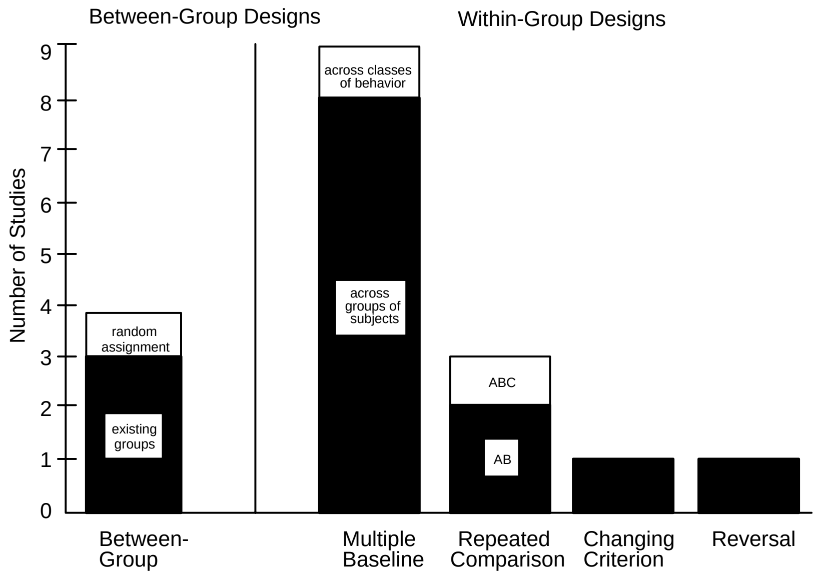


Note. M = 123 and Mdn = 78

Three other types of within-group designs have also been adopted. Cooper, Phillips, Sutherland, and Makin (1994), Laitinen et al. (1998) and Saari and Näsänen (1989) used comparison designs (AB or ABC) repeated in multiple departments, Sulzer-Azaroff, Loafman, Merante, and Hlavacek (1990) used a changing-criterion design, and Chhokar and Wallin (1984a) used a reversal design.

There are inherent problems with experimental control in occupational settings. Organizational officials, who are interested in the "bottom-line," may not be particularly interested in demonstrations of experimental control. Between-group designs may be particularly difficult to use because of the need for random assignment of subjects (Komaki, 1982). As exemplified by Ray et al. (1997), Zohar (1980), and Zohar et al. (1980), because of production, personnel, and/or union demands, organizational officials may require the use of extant groups and teams. Yet, without random assignment, the groups may differ from each other in significant ways (supervision, seniority, shift, current safety levels, job tasks, etc.) and improvements in the experimental group could reasonably be attributed to any one of the differ-

FIGURE 2. Types of experimental designs employed.



ences. In addition, organizational officials may simply be reluctant to implement a safety program for one group of employees, and not others. Given the political and professional problems associated with group designs and random assignment, future researchers should carefully consider the option of adopting a within-subject or within-group design.

Due to the paradigmatic preferences of behavioral researchers to use within-subject designs, combined, perhaps with the difficulties of using between-group designs, the majority of the researchers adopted within-subject designs. These designs eliminate the need for random assignment, which makes them a practical alternative to between-group designs. Because before-after (AB) comparisons do not eliminate alternative explanations, researchers who employ them may want to consider staggering the intervention across groups in a multiple-baseline fashion in order to better demonstrate experimental control. As before, however, organizational officials may not allow a safety program to be implemented for one group of workers and not others, even though the program would ultimately be implemented for all groups. In spite of this, staggered interventions may be an easier ‘‘sell’’ if the researcher explains that it is simply easier to initiate an intervention with a small number of employees rather than with all of them–especially if the organization is large. Organizational officials may readily agree to ‘‘start small’’ and then, over time, ‘‘roll the program out’’ to other units and departments; a practice that is quite common when organizations implement major initiatives.

Another alternative is to use a reversal design, but organizational officials may be reluctant to remove a program that is successful. Researchers must also consider the ethics of terminating, even temporarily, an effective program. Moreover, in some cases, it may not be possible to reverse, or completely withdraw, the intervention, which removes the possibility of proving its effectiveness. For example, safety training cannot be undone. And, although the formal aspects of the program may be removed, employees may surreptitiously use safety checklists if they found them valuable, or continue to support each other’s safe performances in less formal ways. Thus, it appears that the most reasonable design in organizational settings may be a multiple-baseline design. Regardless of what design is used, however, it is important for researchers to attempt to demonstrate experimental control.

DEPENDENT VARIABLES

Injury and Illness Rate

While the most obvious dependent variables in safety studies are injury and illness rates, as others have noted (Cohen & Jensen, 1984; Komaki et al., 1978; Sulzer-Azaroff et al., 1990), problems exist with their exclusive use. First, as McAfee and Winn (1989) stated, ‘‘ . . . accidents are rare events. Therefore, unless large samples of data are analyzed, such accidents may not be a sensitive dependent variable for measuring the success of a safety program’’ (p. 14). When there are not many accidents to begin with, decreases cannot be detected statistically and do not appear to be practically significant. Furthermore, injury and illness rates can vary depending upon seasonal production levels, which is particularly troublesome when using within-group designs because interventions may coincide with the seasonal fluctuations. Such fluctuations are sometimes controlled for by comparing injury and illness rates to those for the same time period in previous years (e.g., Fellner & Sulzer-Azaroff, 1984), but again, no statistical or practical differences may be found if rates are low to begin with.

Secondly, the reliability of injury and illness rates may be questioned because of changes in record keeping practices, and because employees may under-report the number and severity of injuries and illnesses (Chhokar & Wallin, 1984b). That is, observed decreases may not be actual decreases.

A third problem with injury and illness rates involves the inadvertent punishment of desirable behaviors. To illustrate, assume that valued rewards are provided to workers if reportable injuries or accidents have not occurred for a relatively long period of time. Employee reports of injuries toward the end of that interval are likely to be punished. Thus, *reporting* of injuries and illnesses may decrease but the actual injuries and illnesses may not. Certainly, such a decrease is not humanitarian as injuries could go untreated. Nor is such a decrease economically wise, as the long-term costs might far outweigh the short-term costs of immediate care (e.g., further injury, legal ramifications).

Practices and Conditions

Due to the above concerns, researchers have typically measured practices or conditions, alone or in conjunction with injuries and illnesses. A practice is a specific behavior (e.g., cutting in an outward motion), while a condition is the result of a set of behaviors (e.g., floor free of oil).

Safety targets. In order to determine the practices and conditions that would have the greatest impact on injuries and illnesses, 77% of the researchers conducted an initial assessment. Initial assessments usually included one or more of the following: (a) examination of injury and illness records; (b) review of company safety manuals, equipment handbooks, OSHA standards, and trade journals; (c) interviews of workers and supervisors; and (c) direct observation of workers. The specific methods of analysis, however, were not always provided or lacked detail. For purposes of replication, it would be helpful to have more information (readers should see Sulzer-Azaroff & Fellner, 1984, for a thorough treatment).

An interesting variation on the initial assessment was conducted by Hopkins et al. (1986). To determine the safety targets, researchers measured the amount of styrene in the air while employees performed their jobs in different ways. The safety targets were those practices that resulted in the lowest levels. Researchers also continued to measure and report the level of styrene in the air. In the remaining three studies (Zohar, 1980; Zohar et al., 1980; Zohar & Fussfeld, 1981), no formal initial assessments were conducted to determine the targets because the goal was to decrease hearing loss caused by noisy environments and thus the safety target was obvious–wearing earplugs.

The number of specific practices or conditions that were targeted in each study differed, as shown in Figure 3. Moreover, an extremely wide variety of safety targets were measured. The targets can be categorized into four main types: personal protective equipment (e.g., safety glasses, earplugs); material handling (e.g., cutting in outward motion); general safety (e.g., fire extinguishers in appropriate locations); and housekeeping (e.g., oil-free floor). Given the wide range of job types and settings, a detailed listing of the specific safety targets would be too cumbersome to provide here. Suffice it to say that one cannot help being impressed by the wide range of practices and conditions that have been targeted.

Final measures. Safety targets were combined into one overall safety score in almost all (88%) of the studies. Figure 4 displays the final safety measures that were used, and the number of studies using each measure. Figure 5 shows the ways in which the safety targets were combined into the overall scores. Only three studies (12%) used different measures: Sulzer-Azaroff and DeSantamaria (1980) used the frequency and type of hazards, Cohen and Jensen (1984) used the error rate (the percentage of incorrect behaviors), and Hopkins et al. (1986) measured the percent of 15-second observation intervals in which the target behaviors occurred.

Summary

Studies have used similar dependent variables. First, all primary measures have been based on behavior (e.g., practices, conditions). The rationale for the particular type of measure (e.g., percentage of safe practices vs. percentage of safe employees), however, was not provided. Such information would be beneficial. Second, although none of the studies used injury and illness rates as the primary dependent variable, presumably because of their inherent problems, nine did include them as secondary dependent variables (Cooper et al., 1994;

FIGURE 3. Number of safety targets per study.

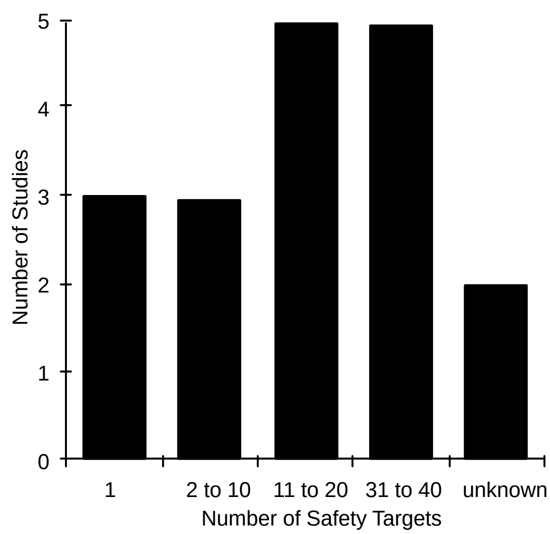


FIGURE 4. Number of measures per study.

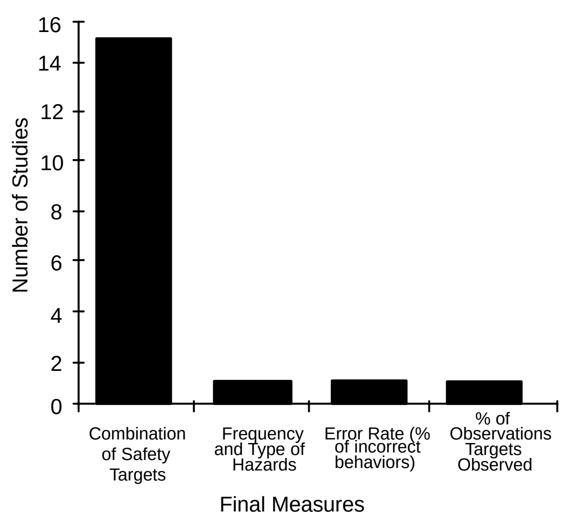
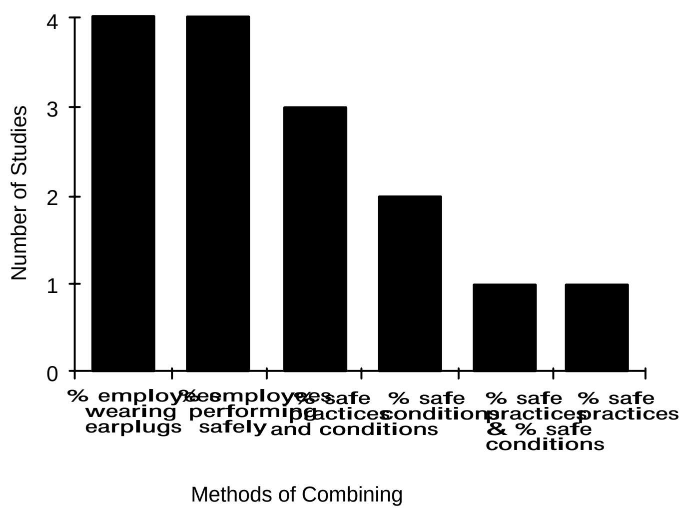


FIGURE 5. Methods of combining safety targets into overall scores.



Fellner & Sulzer-Azaroff, 1984, 1985; Komaki et al., 1978; Ray et al., 1997; Reber & Wallin, 1984; Reber et al., 1990; Saari & Näsänen, 1989; Sulzer-Azaroff et al., 1990). Their inclusion certainly is warranted, as the reduction of injuries and illnesses is the ultimate goal of safety programs. As noted earlier, Hopkins et al. (1986) adopted a very interesting measure. The purpose of this study was to decrease illness due to styrene exposure. Styrene-related illnesses, such as cancer, occur as a result of cumulative, repeated exposures, and symptoms may not present for months or years. Because of this, the rates of styrene-related illnesses are not useful dependent variables. While practices that led to the lowest levels of styrene in the air were the primary dependent variables, researchers also measured styrene air levels. This latter measure is particularly interesting because it provides an intermediate link between the behavioral practices and the later illnesses. Future researchers should consider using such measures, especially when the link between behavioral changes and their safety consequences are remote.

INTERVENTION EFFECTIVENESS

The 18 studies reviewed were categorized as (a) ‘‘singular’’ interventions (n = 4), (b) package programs (n = 6), or (c) component analyses (n = 8). The magnitude of effects of these interventions, in terms of percent improvement, is provided in Table 2. Data are presented for the percent improvement over baseline, and, for the component analyses, percent improvement over each preceding intervention. The percent improvement could not be determined for one package program (Hopkins et al., 1986) and one component analysis (Fellner & Sulzer-Azaroff, 1985), and thus these studies are not included on the table.

Singular Interventions

Researchers in two studies intervened with feedback (Fellner & Sulzer-Azaroff, 1984; Zohar et al., 1980) and two with token economies (Zohar, 1980; Zohar & Fussfeld, 1981). Percent improvements ranged from 9% to 157% over baseline. The same behavior, wearing ear plugs, was targeted in the three studies that resulted in the largest increases of 80%, 119%, and 157%. These stellar results may also be due to the fact that only one behavior was targeted.

Feedback. Fellner and Sulzer-Azaroff (1984) provided weekly graphic, numeric, and verbal feedback to employees. The percentage of safe practices and conditions improved 9% over baseline, and injury and illness rates decreased. Zohar et al. (1980) implemented a unique procedure to increase wearing ear plugs. They described the importance of wearing ear plugs and the hearing loss associated with excessive noise levels to employees in both the experimental and control groups. They then gave short-term hearing loss tests to employees in the experimental group before and after work two times during a one month period. Employees were instructed not to wear earplugs while they worked the first time they were tested and to wear them the second time. The test results were given to each individual employee and publicly posted. Earplug use increased significantly (85%-90%) for the experimental group and remained approximately 10% for the control group. Interestingly, in an attempt to increase earplug use in the control group, management subsequently implemented a disciplinary program. Employees were required to wear earplugs for gradually increasing lengths of time. If workers did not

TABLE 2. Summary of Magnitude of Effects of Behavioral Safety Systems

Singul Interventions			
			over baseline
Fellner & Sulzer-Azaroff (1984)			8.93%
Zohar (1980)			15.74%
Zohar et al. (1980)			118.75%
Zohar & Fussfeld (1981)			80.10%
Package Programs			
			over baseline
Cooper et al. (1994)			32.48%
Komaki et al. (1978)			32.40%
Laitinen et al. (1998)			56.14%
Sulzer-Azaroff & De Santamaria (1980)			59.40%
Sulzer-Azaroff et al. (1990)			37.05%
Component Analyses			
Chhokar & Wallin (1984)	over baseline	over training & goal-setting	over training, goal-setting, & weekly feedback
training & goal-setting	24.09%	–	–
training, goal-setting, & weekly feedback	45.04%	16.88%	–
training, goal-setting, & bi-weekly feedback	48.41%	19.60%	2.33%
Cohen & Jensen (1984)	over baseline		
training group	18.18%		
training & feedback group	22.86%		
Komaki et al. (1982)	over baseline	over antecedents	
antecedents	8.22%	–	
consequences	24.06%	14.86%	
Ray et al. (1996)	over baseline	over training	over training, & feedback
training	0.90%	–	–
training & feedback	12.66%	11.65%	–
training, feedback, & goal-setting	24.16%	23.05%	10.21%
Reber & Wallin (1984)	over baseline	over training	over training & goal-setting
training	13.91%	–	–
training & goal-setting	24.66%	9.44%	–
training, goal-setting, & feedback	53.7%	34.6%	23.0%
Reber et al. (1990)	over baseline	over training	over training & goal-setting
training	13.27%	–	–
training & goal-setting	24.30%	9.69%	–
training, goal-setting, & feedback	52.91%	34.75%	22.45%
Saari & Näsänen (1989)	over baseline	over feedback to supervisors	
feedback to supervisors	14.72%	–	
feedback to workers	37.19%	19.95%	

TABLE 3. Summary of the Intervention Components of the Package Programs

	Intervention Components					
Authors	Training	Feedback	Praise	Goal-Setting	Tangible Rewards	Partici-pation
Cooper et al. (1994)		x		x		
Hopkins et al. (1986)	x	x	x			
Komaki et al. (1978)	x	x	x	x		
Laitinen et al. (1998)	x	x				x
Sulzer-Azaroff & DeSantamaria (1980)		x	x			
Sulzer-Azaroff et al. (1990)	x	x	x	x	x	

wear earplugs, they were taken off the job and docked pay. This program failed.

Token economy. In Zohar (1980) and Zohar and Fussfeld (1981), random tours were conducted daily and employees wearing earplugs were given tokens redeemable for merchandise. In the Zohar study, the tokens had a standard value. In the Zohar and Fussfeld study, the value of the tokens depended upon the percentage of employees wearing earplugs. Both procedures were highly effective. The percentage of employees wearing earplugs increased to 90% and 95%, respectively, representing increases over baseline of 157% and 80%.

Package Programs

In the six package programs, training, feedback, praise, goal-setting, and/or tangible rewards were combined. The components of these programs are summarized in Table 3. As can be seen from Table 2, the package programs have been very successful, with improvements ranging from 32% to 59% over baseline. The components that were functionally important cannot be isolated; nonetheless, the results of these studies are impressive. The six studies are detailed below.

Cooper et al.'s package (1994) consisted of participative goal-setting and weekly written and graphic feedback to workers. The percentage of safe behaviors and conditions improved significantly in 9 of 14 departments (ceiling effects prevented improvement in the remaining 5 departments). In addition, injury and illness rates decreased 21% overall, and accidents decreased 74%. Although there was an inverse relationship between the injury and illness rates and the percentages of safe behavior, it was not statistically significant. Komaki et al. (1978) implemented a similar program that combined training, assigned goal-setting, daily verbal praise, and graphic feedback that was provided four times a week. The percentage of safe practices and conditions increased dramatically when the program was implemented (32% over baseline) and decreased when it was withdrawn, demonstrating a functional relationship. Moreover, within a year, the incident rate of lost-time accidents decreased to less than 10 per million man-hours, and employees worked 280,000 hours without suffering a disabling injury. Whereas the organization had ranked last in safety within the corporation, with these gains, the organization ranked first. Sulzer-Azaroff et al.'s (1990) program was composed of training followed by weekly oral, written, and graphic feedback, praise, goal-setting, and monthly tangible rewards. Once again, the percentage of practices and conditions scored as safe increased dramatically (37% over baseline), and OSHA recordable and lost-time accidents decreased. Laitinen et al. (1998) implemented a comprehensive program based on participatory management, ergonomics, and behavioral safety. It included (a) training in ergonomics and housekeeping standards, (b) worker participation in the development of the program and generation of ideas for improvement, and (c) weekly graphic feedback to employees. The percentages of safe conditions increased 56% over baseline and the percentage of sick leave decreased from 12.8% to 9.9%.

While the preceding researchers intervened with groups of employees, Hopkins et al. (1986) intervened with individual workers. Workers were initially trained, and then observed twice a day. If workers were performing safely, they were praised, if not, they were prompted to perform safely. A majority of the safety targets improved and air levels of styrene declined 36%-57%. Sulzer-Azaroff and De-Santamaria's intervention (1980) targeted individual supervisors, rather than front-line staff. Written feedback, which included the number and location of hazards, and specific suggestions for improvement, were given to supervisors twice a week. Verbal and written praise were also provided. This program was highly successful: Both the frequency and type of hazardous conditions decreased dramatically

(59%). The researchers noted that, in most cases, supervisors had to rely on their workers to implement the suggestions, however, the ongoing behavior of the supervisors or the workers was not assessed.

Component Analyses

Training vs. training and feedback. Cohen and Jensen (1984) examined the effects of training alone versus training plus feedback on the performance of forklift drivers. In the first phase, they used a between group design: One group received training, a second received both training and feedback (daily verbal and posted), and the control group did not receive either. While training improved safety 18% over baseline levels, training plus feedback improved it more (23% over baseline), however, the difference was not statistically significant.

Training with feedback vs. goal-setting. Ray et al. (1997) examined the relative effectiveness of training, feedback, and goal-setting. In this study, baseline was followed by training. Feedback was then added in the form of a sign that displayed the average weekly safety index. In the final condition, assigned goal-setting was added. Training did not improve performance. Feedback improved performance 12% over training, and goal-setting improved performance 10% over training and feedback. Thus, goal-setting enhanced the effects of feedback. The performance of a control group did not change during this period of time. In addition, the injury record for the experimental group improved. The enhancing effects of goal-setting are consistent with those reported by Balcazar, Hopkins, and Suarez (1985/1986) in their classic review of feedback interventions.

Similar studies were conducted by Reber and Wallin (1984) and Reber et al. (1990), where training was provided in the first condition, assigned goal-setting added in the second, and feedback in the third. In both studies, goal-setting consisted of a sign posted with the goal on it. In addition, in Reber and Wallin, supervisors reminded workers of the goal weekly. In both studies, during the feedback phase, numeric scores were posted after each observation (approximately three times a week) and graphed weekly. The results of both programs were similar: (a) training improved performance 13% over baseline, (b) goal-setting improved safety 9%-10% over training, and (c) feedback improved safety 22%-23% over training and goal-setting. In neither study did workers consistently meet goals until feedback was provided. In addition to improvements in behaviors, OSHA recordable and lost-time rates showed marked decreases in both studies. Moreover, in the Reber and Wallin study, the inverse correlation between safe behaviors and injury and illness rates was statistically significant, unlike in the Cooper et al. (1994) study.

Training with goal-setting vs. feedback. Chhokar and Wallin (1984a) examined the relative effects of training with assigned goal-setting and graphic feedback. In addition, feedback frequency (once a week versus every two weeks) was examined. Training with goal-setting significantly improved performance 24% over baseline. Performance again improved when feedback was implemented (17% over training and goal-setting), and decreased when the feedback was withdrawn. No differences were found when the feedback was provided weekly or every other week. Similar to Reber and Wallin (1984) and Reber et al. (1990), workers did not consistently achieve their goals until feedback was provided. Feedback clearly enhanced the effectiveness of goal-setting in these studies, findings that are consistent with the general literature regarding goal-setting and feedback (Locke, Shaw, Saari, & Latham, 1981; Muchinsky, 1997).

Feedback vs. assigned goal-setting vs. participatory goal-setting. Fellner and Sulzer-Azaroff (1985) examined the relative effectiveness of participatory and assigned goal-setting. The study extended their previous one that examined the effects of feedback (Fellner & Sulzer-Azaroff, 1984). Feedback was provided during baseline. Publicly posted charts contained percentages of safe conditions and practices, a graph of the percentages, and specific hazards and their locations. In the assigned goal-setting phase, the foreman praised improvements, placed the goal on the graph, and informed the workers of the goal. During participatory goal-setting, the foreman praised improvements, asked for suggestions on how to improve areas that had not improved, and asked workers to suggest a goal. The goal was then placed on the graph. Results were highly variable across work areas. Overall, assigned goal-setting significantly increased the percentage of safe conditions but not safe practices. Even though workers met the goals significantly more often when they set them than when the supervisor assigned them (67% versus 47%, respectively), participatory goal-setting did not affect either conditions or practices. Thus, the results favored assigned over participatory goal-setting. Neither type of goal-setting affected injury or illness rates. Unlike the previous studies reviewed, goal-setting did not enhance the effects of feedback. Up-ward safety trends and ceiling effects in the feedback phases may have precluded further improvements, however.

Training and feedback to employees vs. to supervisors. Saari & Näsänen (1989) compared the effects of providing feedback to supervisors versus providing it directly to employees. In the first condition, all employees were trained, and the safety score was given to the foremen 1-3 times a week. In the second phase, graphic feedback was provided directly to employees 1-3 times a week in addition to the foremen. When foremen received the feedback performance improved 15% over baseline but improvements were limited to areas near the foremen's office. No improvements occurred in more remote areas. When feedback was provided directly to employees in addition to supervisors, safety improved another 20%. Significant decreases in injuries and illnesses occurred as well.

Antecedent vs. consequent. In 1980, Komaki, Heinzmann, and Lawson reported that feedback, as a consequence, improved safety more than the antecedent of training. When reviewing the results of that study, Komaki et al. (1982) stated:

> A close analysis of the above study, however, revealed an alternative explanation for the results. As in virtually all training programs, the information in the study was provided only once; safety was the subject of one training session but was not necessarily mentioned again. In contrast, during the feedback phase the graphs were updated three or more times a week and supervisors collected the information and provided feedback. Thus, the consequent control procedure's effectiveness may have been due to the greater frequency of stimuli changes and/or supervisor attention rather than the feedback per se. (p. 335)

Komaki et al. (1982) again examined the effects of antecedent versus consequent interventions, this time keeping supervisor involvement and stimulus changes constant across the two conditions. In the antecedent condition, employees were trained on safety rules, rules were posted, supervisors discussed the rules every week, and a new rule was highlighted 3 times a week. In the consequent condition, employees were trained on graph interpretation, supervisors discussed safety weekly, and graphic feedback was updated 3 times a week. During the antecedent phase, performance improved in only two of the four departments (an overall 8% improvement over baseline). When consequences were added, performance improved in all four departments, with an overall increase of 24% over baseline, and 15% over antecedents. As the authors stated, ‘‘These results confirm that performance consequences such as feedback play a critical role in work motivation and that antecedents alone may not be effective in all cases, even if one can rely on fairly extensive supervisor involvement’’ (p. 334).

Summary

Overall, behavioral safety interventions have been effective in improving safety, with consequent interventions proving more effective than antecedent interventions. That said, a few comments should be made that may be of use to future researchers. First, although there were a variety of interventions, the rationale for why a specific intervention was used was not specified in most studies. It appears the researchers simply employed a number of behavior change strategies that have been shown to be effective in other settings. Moreover, few researchers adequately tied the rationale to any concepts and principles of behavior analysis, in general, or to a behavioral analysis of safety, in particular.

Secondly, in some studies, although the interventions were successful at a general level, not all the safety targets were positively affected (Cohen & Jensen, 1984; Hopkins et al., 1986). For example, in Hopkins et al., wearing a respirator in the presence of styrene did not change, presumably because of the discomfort and inconvenience. Yet this is a critical behavior. In Cohen and Jensen study, forklift operators did not increase the extent to which they looked over their shoulders when driving in reverse. Again, failure to do so clearly has potentially hazardous outcomes. The authors stated that, ‘‘ . . . driving in reverse caused them to breathe in noxious fumes. Further, continuous looking over one’s shoulder is an unnatural and uncomfortable posture to assume for prolonged periods’’ (p. 132). Researchers should carefully examine the natural contingencies for the types of behaviors that are resistant to change, and modify their programs accordingly. Caution should also be taken when developing safety targets. It is not desirable to behaviorally engineer the breathing of noxious fumes or increase behaviors that could cause strain injuries. Engineering or behavioral alternatives may be available that would reduce the need for such behaviors. Behavioral and safety engineers should consult with employees who are likely to know much more about the job than they do and then work as a team to solve such problems.

Similarly, the effects of the interventions in the two Fellner and Sulzer-Azaroff studies (1984, 1985) were highly variable. The reasons for such variability need to be explored. In the 1984 study, improvements occurred only in those locations where employees paid greater attention to the feedback, and supervisors discussed the feedback more frequently. Fellner and Sulzer-Azaroff speculated that the effectiveness of feedback may be promoted by its examination and discussion, and research along these lines is warranted. It would also be of interest to determine if the type or frequency of feedback affects its power. In many of the studies feedback was provided in multiple forms (e.g., graphic, oral, and written) simultaneously, thus the relative effectiveness of feedback type cannot be ascertained. Furthermore, feedback was delivered daily, 3-4 times a week, weekly or bi-weekly. Although Chhokar and Wallin (1984a) found no differences in performance when feedback was provided weekly or bi-weekly, additional evaluations of feedback frequency would be of interest. In their review of the general effects of feedback, Balcazar et al. (1985/1986) stated that graphic feedback provided once a week was more effective than other types; however, their conclusion was based on across-study comparisons rather than experimental comparisons. Direct comparisons in the safety literature would certainly yield valuable information for safety practitioners as well as for others.

Thirdly, due to the possible ceiling effects in the Fellner and Sulzer-Azaroff (1985) study, the relative effectiveness of assigned versus participative goal-setting has yet to be established. Additionally, there are no objective guidelines with respect to how high a goal should be set. Most researchers mentioned setting ‘‘difficult yet attainable’’ goals, however, more specific and objective guidelines for setting safety goals in relation to baseline performance would be beneficial.

Finally, although some component analyses have been conducted, they are not sufficient. The component analyses suggest that (a) both feedback and goal-setting enhance training, (b) goal-setting enhances feedback, and (c) feedback enhances goal-setting. Thus, the most effective combination appears to be training, goal-setting and feedback. Additional research is needed, however. Moreover, few studies have investigated the effects of tangible rewards. Rather, the majority of studies have relied on feedback as a consequence. The rare exceptions were studies conducted by Sulzer-

Azaroff et al. (1990), Zohar (1980) and Zohar and Fussfeld (1981). This, in spite of Balcazar et al.'s (1985/1986) advice:

> If no system of functional, differential consequences exist, there is probably no point in establishing a feedback system. Effort would be better spent developing procedures for reinforcing wanted behaviors.
>
> If a feedback system is going to be established independently of careful considerations of the existence of functional, differential consequences . . . the evidence suggests that the best bets are to combine feedback that is graphically presented at least once a week with tangible rewards. Eighty percent of the studies with known effects that applied these characteristics were consistently effective regardless of whether goal setting procedures were additionally used. (p. 84)

The field would, thus, be well-served by experimental comparisons of feedback, goal-setting and tangible rewards.

MISCELLANEOUS EFFECTS

Injuries and Illnesses

Nine studies included measures of injury and illness rates (Cooper et al., 1994; Fellner & Sulzer-Azaroff, 1984, 1985; Komaki et al., 1978; Ray et al., 1997; Reber & Wallin, 1984; Reber et al., 1990; Saari & Näsänen, 1989; Sulzer-Azaroff et al., 1990). With the exception of Fellner and Sulzer-Azaroff (1985), all reported decreases in injury and illness rates following intervention. However, as discussed earlier, such results must be interpreted with caution. For example, Reber and Wallin described how record-keeping practices changed during the study, thus raising questions about the measure's reliability. Nevertheless, even though inherent difficulties exist with the use of injury and illness measures, researchers should still consider using them as ancillary measures while appropriately acknowledging their limitations.

Costs and Benefits

The costs of a safety program must not outweigh its benefits in order for it to be accepted initially and later maintained. Sulzer-Azaroff et al. (1990) reported the total cost savings of their intervention to be $55,000.00 and Cooper et al. (1994) reported that their program paid for itself. Reber and Wallin (1984) and Reber et al. (1990) reported ‘‘6 figure savings’’ and ‘‘substantial monetary savings,’’ respectively; however, the actual numbers were not provided. In the Fellner and Sulzer-Azaroff (1984, 1985) studies, most of the costs came from the development of the program, and because the running costs were low ($14.00 and $28.00 per week, respectively), substantial savings were assumed. Laitinen et al. (1998) reported the cost of their three-year program to be about $1.4 million. However, the program included many physical improvements that accounted for half of this expenditure. No data were provided on assumed savings. Zohar (1980) and Zohar and Fussfeld (1981) reported the costs of the tokens to be $15.00 and $10.00 per employee, respectively, which, as they pointed out, is less expensive than typical poster campaigns. Therefore, given that the average occupational injury or illness can run into the tens of thousands of dollars, substantial cost savings can be assumed.

There is an inherent obstacle in obtaining cost savings numbers. The measure compares the start-up and daily operation costs of the program from estimates of what the injuries and illnesses that *did not occur* would have cost. Given the problems previously discussed with injury and illness rate measures, any cost savings must be cautiously interpreted. Nevertheless, as stated, the costs of a safety program must not outweigh its benefits. Thus, cost/benefit analyses are an important ancillary measure and should be reported more frequently.

Productivity

Hopkins et al. (1986) obtained measures of productivity and time spent working before and after the intervention. When the program was first implemented, there was an initial decrease in productivity for 3 of the 4 workers. However, performance then rose above baseline levels. No changes were found in the time spent working. Similarly, Komaki et al. (1978) reported that there were no fluctuations in productivity due to their safety interventions. Sulzer-Azaroff and DeSantamaria (1980) did not obtain formal measures of productivity, but they anecdotally reported that productivity may have increased following the implementation of the intervention. The authors speculated that these productivity improvements may have occurred because of

(a) increased time on the job because injuries were reduced and (b) the ''safer'' arrangement of materials and tools. As discussed earlier, in applied settings decision-makers are usually not the researchers and it is certainly reasonable for them to be concerned about improving safety at the expense of productivity. Moreover, however unfortunate, safety professionals are often in conflict with production professionals regarding safety programs and expenditures. Thus, measures of productivity such as those used by Hopkins et al. are important to overcome such concerns, or at the very least, to determine the effects of a safety program on productivity. For these reasons, future researchers should obtain such measures when possible.

MAINTENANCE

Maintenance of the Program

Only 6 studies of the 18 studies mentioned whether the program was maintained in the organization after the study. Of those six, programs in four were maintained; programs in the other two were abandoned, however, similar programs were adopted in other units in the organization. The Sulzer-Azaroff and DeSantamaria (1980) program was kept in place with modifications (e.g., feedback was provided once a week instead of twice a week). The Komaki et al. (1978) program was maintained with modifications (e.g., graphs were posted once a week instead of four days a week) and expanded to include the other work shift after management saw the effects of the reversal phase. The Fellner and Sulzer-Azaroff (1984) feedback program was continued and later used as the site for the participatory versus assigned goal-setting study (e.g., Fellner & Sulzer-Azaroff, 1985). Similarly, the Sulzer-Azaroff et al. (1990) program was continued with other departments added. In two studies, Saari and Näsänen (1989) and Laitinen et al. (1998), the safety program was withdrawn from the units that participated in the study, perhaps to examine post-study/reversal effects, however similar programs were adopted in other units of the organization.

The reasons why these programs were continued and, presumably, not the others were not specified in most studies. Sulzer-Azaroff and her colleagues certainly deserve special recognition, however, because all of the Sulzer-Azaroff programs were maintained. It is likely that this maintenance is due to pre-program systems analyses. A prime example of this approach is provided in Sulzer-Azaroff et al. (1990). A comprehensive system analysis was conducted to insure that the program was compatible with other formal and informal systems within the organization. Not only were all of the program components based on this analysis, but specific pinpoints were identified and communicated to each key organizational official (supervisor, manager and director). Both researchers and practitioners are strongly encouraged to follow suit. While this approach is likely to improve maintenance, researchers should, nonetheless, examine the specific variables that are important in keeping a successful program in place. Organizational officials are not always data-driven: The decision to keep or disband a program may be made for reasons independent of the data, at least independent of the data that are collected as part of the safety program. Thus, determining the reasons why decision-makers choose to keep or remove a successful program would be beneficial in order to design programs that are maintained after the researchers leave the setting.

Maintenance of the Behavior Change

As indicated above, safety programs were maintained in only four studies. In two of the four, researchers reported whether performance was also maintained. Sulzer-Azaroff and DeSantamaria (1980) measured safety 3 days, 2 weeks, 6 weeks and 4 months after the experiment had been formally terminated. Safety remained excellent in all six participating departments. Komaki et al. (1978) reported that their results sustained and the injury rate continued to decline but neither the time period nor data were provided.

Six researchers conducted performance assessments after the program had been withdrawn. Five of the six assessments occurred six or fewer months after withdrawal (Cohen & Jensen, 1984; Laitinen et al., 1998; Zohar, 1980; Zohar et al., 1980; Zohar & Fussfeld, 1981). In all but one, program effects maintained. In the exception, Zohar and Fussfeld, the results maintained in all four departments after three months and in three of the four after six months. Saari and Näsänen (1989) were the only researchers to formally assess maintenance for longer periods of time. Performance remained stable in two units, A and B, for 22 and 13 months respectively. In addition, although no formal measures were obtained, Zohar anecdotally reported that most employees were still wearing earplugs one year after the program's termination.

In Sulzer-Azaroff and DeSantamaria (1980) and Komaki et al. (1978), the programs were maintained and the results maintained. This is, of course, good news, although more data of this kind are required to convincingly state that behavioral safety programs lead to sustained improvements. What is interesting, and more difficult to understand, however, is why performance maintained in sites where the safety programs were withdrawn. That is, by what mechanisms were the behaviors supported after program termination? In Zohar (1980), Zohar et al. (1980), and Zohar and Fussfeld (1981), rates of earplug use remained high at follow-up in spite of the fact that due to high turnover (60%, 65%, and 40%, respectively) large numbers of employees had never been exposed to the safety program. The authors provided three possible reasons. First, at the organizational level, new policies were implemented. Like punctuality and proscribed production levels, earplug use was made mandatory, and supervisors were responsible for maintaining it. Secondly, at the departmental level, new cultural norms regarding earplug use may have been established. Finally, at the employee level, the contrived program contingencies may have induced employees to wear ear plugs (in spite of the immediately punishing consequences), thus bringing employees in contact with the positive natural contingencies (reduction in noise and hearing loss) that maintained the behavior. This analysis, of course, is only relevant for workers who were initially exposed to the program. However, these workers may have prompted and praised ear plug use by newly hired workers. Saari and Näsänen (1989) speculated that the feedback (functioning as a reinforcer) on housekeeping outcomes (e.g., trash can empty) resulted in the outcomes themselves becoming conditioned reinforcers. Thus, once the graphs were no longer displayed, performance maintained because employees received feedback and reinforcement directly from the housekeeping outcomes. Laitinen et al. (1998) and Cohen and Jensen (1984) suggested that the performance levels may have maintained because of the participatory process that involved people from all levels of the organization in the development and execution of the programs. In addition, Cohen and Jensen hypothesized that maintenance may have been due to safer habits and new group norms sustained by peer modeling and management support.

It is important to demonstrate that behavioral interventions improve safety, but it is also important to demonstrate their effectiveness over long periods of time. In addition, the organizational, group, and individual mechanisms that maintain safety after programs are formally withdrawn should be determined. Only then will it be possible to *systematically* design safe ‘‘cultures’’ and develop new group ‘‘norms.’’

INTEGRITY AND RELIABILITY

Integrity of the Independent Variable

Intervention integrity measures were provided in only a few studies. Three examined the integrity of training by measuring employees' knowledge of the training material (Chhokar & Wallin, 1984a; Reber & Wallin, 1984; Reber et al., 1990). Performance tests indicated that training was successful. Subjects in the Ray et al. (1997) study demonstrated their knowledge during the training session, although no data were provided. Sulzer-Azaroff and DeSantamaria (1980) conducted sessions with supervisors to ensure that they understood the feedback form, however, no measures were taken. Sulzer-Azaroff et al. (1990) informally assessed integrity of the independent variables.

Reported problems typically involved the behaviors of supervisors and observers. For example, Komaki et al. (1978) and Cooper et al. (1994) indicated that supervisors did not deliver praise as planned. Similarly, Komaki et al. (1982) noted that supervisors attended fewer and fewer feedback meetings over time. In addition, they found that the employees who did not prefer the consequent condition over the antecedent condition did not understand the graphs. Cooper et al. also reported that observers did not gather data in some weeks. In Zohar et al. (1980), the observer for the control group initially inflated the data and a double-monitoring system had to be put into place. These problems suggest that researchers should take steps to ensure the integrity of independent variables. Special attention should be focused on those who have direct responsibility for implementation: training and monitoring are essential. Moreover, a behavioral safety system must be incorporated into the existing management structure for it to maintain once the researchers leave the organization. Without implementing internal controls to ensure integrity, a behavioral safety program may not maintain over time, regardless of its initial success.

Reliability of the Dependent Variable

Behavioral safety researchers did, for the most part, assess the reliability of the dependent variables, typically in terms of interobserver agreement. Four studies did not report reliability measures (Cooper et al., 1994; Laitinen et al., 1998; Ray et al., 1997; Zohar, 1980), and Sulzer-Azaroff et al. (1990) assessed reliability only informally. Zohar et al. (1980) reported initial reliability problems, however, interobserver agreement was high in the other studies, ranging from 83%-100%. In addition to measuring the reliability of their behavioral targets, Hopkins et al. (1986), commendably, assessed the reliability of the styrene air level measures and productivity measures.

SOCIAL VALIDITY

Social validity (the evaluation of the acceptability of a program by its consumers) is important not only to increase ‘‘buy-in’’ for programs but also to decrease resistance to it and, possibly, increase worker morale. The three main types of social validity include the acceptance of (a) the goals of the program, (b) the procedures employed, and (c) the outcomes of the program (Schwartz & Baer, 1991).

Goals

By its nature, the goals of behavioral safety programs are socially valid. That is, given the rates of occupational injuries and fatalities and the monetary costs involved, few would argue that the improvement of safety is not a worthy goal.

Procedures

A number of studies examined the second type of social validity–the acceptance of the procedures employed. During the course of the study, Chhokar and Wallin (1984a), Cooper et al. (1994), Reber and

Wallin (1984), and Reber et al. (1990) used a questionnaire to determine the acceptance of the program and found that it was acceptable to employees. Laitinen et al. (1998) conducted before and after assessments of the perceived physical and psychosocial (e.g., cooperation, solidarity, and support) working conditions. They found both improved significantly. Others examined the social validity of the procedures *after* the completion of the study. For example, Komaki et al. (1982) assessed employee preference for the antecedent versus the consequent condition, and found that 72% of the employees preferred the consequent condition. Interestingly, as mentioned earlier, they also found that those who preferred the antecedent condition also reported that they did not understand the feedback graphs. In Fellner and Sulzer-Azaroff (1985), employees, responding to a questionnaire, said that they were indifferent to the feedback and goal-setting condition. That notwithstanding, employees indicated that they preferred to set their own goals rather than to have supervisors assign them, even though participative-goal setting was not more effective than assigned goal-setting. Participants in Saari and Näsänen's (1989) study rated the interventions positively. Although Sulzer-Azaroff et al. (1990), Komaki et al. (1978), and Ray et al. (1997) did not formally assess employee acceptance, they anecdotally reported positive reviews of their programs. In fact, both Komaki et al. (1978) and Sulzer-Azaroff et al. (1990) stated that workers would cheer when the new data point was added to the graph. Similarly, Ray et al. reported favorable responses to feedback, and indicated that employees would suggest additional ways to improve safety during feedback sessions.

Outcomes

Sulzer-Azaroff et al. (1990) were the only researchers who reported the acceptance of outcomes, and they did so only anecdotally, noting that the safety director said, ''The program was fantastic. I never dreamed people would be so successful'' (p. 118). Laitinen et al. (1998) hinted at outcome acceptance when they discussed the initial difficulties in obtaining funding for the program: They reported that as the effects in the first department became known, funding was no longer a problem, a clear indication of management's acceptance of the outcomes. Outcome acceptance can also be inferred in those studies where the programs were maintained or expanded. Thus, although it appears that behavioral safety programs are viewed favorably in terms of their procedures, more pervasive use of formal and objective social validity measures of the procedures and the outcomes would be of interest. They would certainly be of value when marketing behavioral safety programs to new organizations.

CONCLUSION

Behavioral safety researchers have demonstrated the effectiveness of a variety of behavioral interventions in a wide variety of manufacturing settings with many different jobs. Although we have restricted our review to applications in manufacturing settings, our conclusions do not differ from those who have examined behavioral applications in other settings. For example, McAfee and Winn (1989) reviewed the results of 24 behavioral safety programs that were implemented between 1971 and 1987. Settings that were represented in their review included human service settings, coal mines, packaging forwarding facilities, city maintenance divisions, city refuse divisions, urban transit, textile weaving mills, and public safety departments. Upon finding this diversity, McAfee and Winn stated, ''In the 24 studies, 20 different job classifications are represented. Certainly, researchers can't be criticized for limiting their studies to only a narrow range of jobs or industries'' (p. 9). They concluded that:

> The major finding was that every study, without exception, found that incentives or feedback enhanced safety and/or reduced accidents in the work place, at least in the short term. Few literature reviews find such consistent results. Although this may be surprising to some, others might argue that this finding is simply further proof of the law of effect which contends that rewarded behavior tends to be repeated. (p. 15)

Perusal of recent studies conducted in settings other than manufacturing (e.g., Austin, Alvero, & Olson, 1998; Austin, Kessler, Riccobono, & Bailey, 1996; Laitinen & Ruohomaki, 1996; Sulzer-Azaroff, in press) and case studies in recent behavioral safety texts (Geller, 1996; McSween, 1995) also adds credence to the generality of our conclusions. Our selected review permitted us to highlight the success of behavioral interventions in an industry where risk of injury is high. It also permitted us to provide a more detailed analysis than a more comprehensive review would have permitted.

While the success rate of behavioral safety interventions is high, questions remain: Questions, that if answered, would lead to continued development of our work in safety. We do not yet conclusively know which independent variables are most important, nor how they relate to a behavioral analysis of safe performance. We encourage safety professionals to conduct functional analyses prior to intervention. In addition, we have not addressed ways to modify those behaviors that appear resistant to change when feedback and/or goal-setting fail. Finally, we need to identify the factors that lead to long-term program maintenance and performance change. This latter suggestion may well be our most important task. Given the humanitarian and economic importance of occupational safety, we encourage additional research that will help determine the most effective and efficient methods that will lead to long-term safety in organizations. We hope that this review will be a spring board to that end.

NOTES

1. An Occupational Injury is, "any injury such as a cut, fracture, sprain, amputation, etc., which results from a work accident or from an exposure involving a single incident in the work environment" (Occupational Safety and Health Administration [OSHA]).
2. An Occupational Illness is, "any abnormal condition or disorder, other than one resulting from an occupational injury, caused by exposure to environmental factors associated with employment. It includes acute and chronic illnesses or diseases which may be caused by inhalation, absorption, ingestion, or direct contact" (OSHA).
3. According to the Bureau of Labor Statistics (1997, Dec.), the incident rates represent, "the number of injuries and illnesses per 100 full-time workers and were calculated as: (N/EH) x 200,000, where N = number of injuries and illnesses; EH = total hours worked by all employees during the calendar year; 200,000 = base for 100 equivalent full-time workers (working 40 hours per week, 50 weeks per year.)" (Table 1.)
4. Two separate studies are discussed in the Cohen and Jensen (1984) manuscript. The current paper only reviews study one in Cohen and Jensen.
5. Two separate studies are discussed in the Zohar (1980) manuscript. The current paper only reviews study two in Zohar.

REFERENCES

Austin, J., Alvero, A. M., & Olson, R. (1998). Prompting patron safety belt use at a restaurant. *Journal of Applied Behavior Analysis, 31*, 655-657.

Austin, J., Kessler, M. L., Riccobono, J. E., & Bailey, J. S. (1996). Using feedback and reinforcement to improve the performance and safety of a roofing crew. *Journal of Organizational Behavior Management, 16*, 49-75.

Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985/1986). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, *7*, 65-89.

Chhokar, J. S., & Wallin, J. A. (1984a). A field study on the effect of feedback frequency on performance. *Journal of Applied Psychology*, *69*, (3), 524-530.

Chhokar, J. S., & Wallin, J. A. (1984b). Improving safety through applied behavior analysis. *Journal of Safety Research*, *15*, 141-151.

Cohen, H. H. & Jensen, R. C. (1984). Measuring the effectiveness of an industrial lift truck safety training program. *Journal of Safety Research*, *15*, 125-135.

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*, 219-240.

Evans, G. W., Palsane, M. N. & Carrere, S. (1987). Type A behavior and occupational stress: A cross-cultural study of blue-collar workers. *Journal of Personality & Social Psychology*, *52(5)*, 1002-1007.

Fellner, D. J., & Sulzer-Azaroff, B. (1984). Increasing industrial safety practices and conditions through posted feedback. *Journal of Safety Research*, *15*, 7-21.

Fellner, D. J., & Sulzer-Azaroff, B. (1985). Occupational Safety: Assessing the impact of adding assigned or participative goal-setting. *Journal of Organizational Behavior Management*, *7(1/2)*, 3-24.

Fitch, H. G., Hermann, J., & Hopkins, B. L. (1976). Safe and unsafe behavior and its modification. *Journal of Occupational Medicine*, *18(9)*, 618-622.

Fox, D. K., Hopkins, B. L., & Anger, W. K. (1987). The long-term effects of a token economy on safety performance in open-pit mining. *Journal of Applied Behavior Analysis*, *20*, 215-224.

Geller, S. E. (1996). *Working safe: How to help people to actively care for health and safety.* Radnor, PA: Chilton.

Haynes, R. S., Pine, R. C., Fitch, H. G. (1982). Reducing accidents rates with organizational behavior modification. *Academy of Management Journal*, *25*(2), 407-416.

Hopkins, B. L, Conard, R. J., Dangel, R. F., Fitch, H. G., Smith, M. J., & Anger, W. K. (1986). Behavioral Technology for Reducing Occupational Exposures to Styrene. *Journal of Applied Behavior Analysis*, *19*(1), 3-11.

Jones, J. W. & Wuebker, L. J. (1993). Safety locus of control and employees' accidents. *Journal of Business and Psychology*, *7*(4), 449-457.

Komaki, J. L. (1982). The case for the single case: Making judicious decisions about alternatives. In L. W. Frederiksen (Ed.), *Handbook of organizational behavior management* (pp. 145-176). New York: John Wiley.

Komaki, J., Barwick, K. D., & Scott, L. R. (1978). A Behavioral Approach to Occupational Safety: Pinpointing and Reinforcing Safe Performance in a Food Manufacturing Plant. *Journal of Applied Psychology*, *63*(4), 434-445.

Komaki, J., Collins, R. L., & Penn, P. (1982). The Role of Performance Antecedents and Consequences in Work Motivation. *Journal of Applied Psychology*, *67*(3), 334-340.

Komaki, J., Heinzmann, A. T., & Lawson, L. (1980). Effect of training and feedback:

component analysis of a behavioral safety program. *Journal of Applied Psychology*, *65*, 261-270.

Laitenen, H., & Ruohomaki, I. (1996). The effects of feedback and goal setting on safety performance at two construction sites. *Safety Science*, *24*, 61-73.

Laitinen, H., Saari, J., Kivistö, M., & Rasa, P. (1998). Improving physical and psychosocial working conditions through a participatory ergonomic process: A before-after study at an engineering workshop. *International Journal of Industrial Ergonomics*, *21*, 35-45.

Loafman, B. (1996). Rescue from the safety plateau. *Performance Management Magazine*, *14*(3), 3-10.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin*, *90*, 125-152.

Mattila, M. & HyÖdynmaa, M. (1988). Promoting job safety in building: An experiment on the behavior analysis approach. *Journal of Occupational Accidents*, *9*, 255-267.

McAfee, R. B. & Winn, A. R. (1989). The use of incentives/feedback to enhance work place safety: A critique of the literature. *Journal of Safety Research*, *20*(1), 7-19.

McSween, T. E. (1995). *The values-based safety process: Improving your safety culture with a behavioral approach*. New York: Van Nostrand Reinhold.

Muchinsky, P. M. (1997). *Psychology applied to work* (5th edition). New York: Brooks/Cole.

National Safety Council. (1997). *Accident Facts, 1996 Edition*. Itasca, IL: Author.

Occupational Safety and Health Administration. *Instructions for OSHA no. 200*. Washington D. C.: Author.

Ray, P. S., Bishop, P. A., Wang, M. Q. (1997). Efficacy of the components of a behavioral safety program. *International Journal of Industrial Ergonomics*, *19*, 19-29.

Reber, R. A., & Wallin, J. A. (1984). The effects of training, goal-setting, and knowledge of results on safe behavior: a component analysis. *Academy of Management Journal*, *27*(3), 544-560.

Reber, R. A., Wallin, J. A., & Chhokar, J. S. (1990). Improving safety performance with goal-setting and feedback. *Human Performance*, *3*(1), 51-61.

Saari, J. & Näsänen, M. (1989). The effect of positive feedback on industrial housekeeping and accidents: a long term study at a shipyard. *International Journal of Industrial Ergonomics*, *4*, 201-211.

Schwartz, I. S. & Baer, D. M. (1991). Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis*, *24*, 189-204.

Skane, D. H. (1985). *Elementary Statistics*. Reading, MA: Addison-Wesley.

Sulzer-Azaroff, B. (1978). Behavioral ecology and accident prevention *Journal of Organizational Behavior Management*, *2*, 11-44.

Sulzer-Azaroff, B. (1990). Strategies for maintaining change over time. *Behaviour Change*, *7*(1), 3-15.

Sulzer-Azaroff, B. (1998). *Who killed my daddy? A behavioral safety fable*. Cambridge, MA: Cambridge Center for Behavioral Studies.

Sulzer-Azaroff, B. (in press). Behavioral safety for the twenty-first century. In H.

Laitinen and J. Saari (Eds.), *People and work*. Finnish Department of Occupational Health: Division of Occupational Safety.

Sulzer-Azaroff, B. & DeSantamaria, M. C. (1980). Industrial Safety Hazard Reduction through Performance Feedback. *Journal of Applied Behavior Analysis*, *13*(2), 287-295.

Sulzer-Azaroff, B. & Fellner, D. J. (1984). Searching for performance targets in the behavior analysis of occupational safety: An assessment strategy. *Journal of Organizational Behavior Management*, *6*(2), 53-65.

Sulzer-Azaroff, B., Harris, T. C., & Blake-McCann, K. (1994). Beyond training: organizational performance management techniques. *Occupational Medicine 9*(2), 321-339.

Sulzer-Azaroff, B., Loafman, B., Merante, R. J., & Hlavacek, A. C. (1990). Improving Occupational Safety in a Large Industrial Plant: A Systematic Replication. *Journal of Organizational Behavior Management*, *11*(1), 99-120.

United States Department of Labor, Bureau of Labor Statistics. (1995, December). *Workplace injuries and illnesses in 1994.* Technical Report #USDL-95-508.

United States Department of Labor, Bureau of Labor Statistics. (1997, August). *National census of fatal occupational injuries, 1996.* Technical Report #USDL-97-266.

United States Department of Labor, Bureau of Labor Statistics. (1997, December). *Workplace injuries and illnesses in 1996.* Technical Report #USDL-97-453.

Zohar, D. (1980). Promoting the use of personal protective equipment by behavior modification techniques. *Journal of Safety Research*, *12*(2), 78-85.

Zohar, D., Cohen, A., & Azar, N. (1980). Promoting increased use of ear protectors in noise through information feedback. *Human Factors*, *22*(1), 69-79.

Zohar, D., & Fussfeld, N. (1981). Modifying earplug wearing behavior by behavior modification techniques: an empirical evaluation. *Journal of Organizational Behavior Management*, *3*(2), 41-52.