Employee Performance

Behavioral Observations

Effects on safe performance

By Joseph R. Sasson, John Austin and Alicia M. Alvero

MUCH HAS BEEN WRITTEN about behavioral safety and the essential components of a behavioral safety system (Geller, 2001; Krause, 1997; McSween, 2003; Sulzer-Azaroff & Austin, 2000). While many different steps are discussed—and while many different terms have been used—authors in the field agree that behavioral safety consists of identifying behaviors which can lead to accidents and injuries; developing interventions to reduce the likelihood of these behaviors; conducting observations of work behavior; and providing feedback on safety performance (Komaki, Heinzmann & Lawson, 1980; Sulzer-Azaroff & Austin, 2000). Some systems even incorporate the use of reinforcement strategies to promote safe behavior (Austin, Kessler, Riccobono et al., 1996; Fox, Hopkins & Anger, 1987).

Although a behavioral safety process encompasses many elements, conducting observations is a key component. Most times, there is no permanent product of safe or unsafe performance; this creates the need to directly observe worker behavior in order to record how safely a person (or group of people) is performing. In theory, these observations could be conducted by a consultant (or other third party), but they are typically conducted by employees and/or supervisors in the organization (McSween, 2003; Krause, 1997).

Joseph R. Sasson, Ph.D., *is a researcher in the fields of behavioral safety, performance management, knowledge management and organizational systems and processes. Affiliated with MedAxiom and Florida State University (FSU), he holds a B.S. in Psychology from FSU, and an M.A. in Industrial/Organizational Psychology and a Ph.D. in Applied Behavior Analysis from Western Michigan University (WMU).*

John Austin, Ph.D., *is an associate professor of psychology at WMU, where he teaches courses in performance management and behavioral safety. In addition, he is a senior consultant with Aubrey Daniels International, where he consults on behavioral safety, leadership and performance improvement systems. Austin holds a B.A. from the University of Notre Dame, and an M.S. and Ph.D. from FSU.*

Alicia M. Alvero, Ph.D., *is an assistant professor of psychology at Queens College, the City University of New York, where she teaches courses in organizational behavior management and behavioral safety. She holds a B.A. from Florida International University, and an M.A. and Ph.D. from WMU. Alvero is an editorial board member of the* Journal of Organizational Behavior Management, *and has consulted with various organizations in the areas of training and behavioral safety.*

One can cite many theoretical (as they are difficult to quantify reliably) benefits to having employees conduct safety observations (as opposed to a contractor hired solely for this purpose). These include increased employee participation; helping the process become part of the organization's culture; assisting with process maintenance; and generating additional discussion about safety in the workplace. In most cases, it is also more cost effective. Thanks to these benefits, many behavioral safety experts have concluded that using employee observers is the preferred method of data collection (McSween, 2003; Krause, 1997).

An underlying premise of research endeavors is to help develop methods and applications of practice. Research should create usable, effective strategies that practitioners can apply to achieve change. While much of the behavioral safety research has been conducted using feedback and incentives to influence behavior change, little research has examined the effects of conducting observations as a subcomponent of the behavioral safety process (for one exception see Alvero & Austin, 2003, 2004). However, logic suggests that if workers are often the primary observers in behavioral safety processes, then it would be beneficial to know the effects of conducting observations within a behavioral safety system.

Therefore, two primary purposes of this study were to: 1) assess the effects of conducting safety observations on the safety performance of the observer in an applied setting; and 2) evaluate the relationship between observation accuracy and observer behavior change.

The Study

A study conducted at a large hospital (323 licensed beds) sought to assess factors beyond the widely accepted theoretical benefits of having employees conduct observations. The research team wanted to examine whether safety performance improved as a result of having employees conduct behavioral safety observations. The study involved 11 participants who worked in the hospital's patient accounting and scheduling departments.

Although 11 participants may appear to be a small sample, it is a reasonable sample size for a study using a within-subject (or single-case) research design. In this type of research design, inferences are usually made about the effects of an intervention(s) by comparing different conditions presented to the same participant(s) over time (Kazdin, 1982). A within-subject design requires fewer participants to detect effects of an intervention (Thompson & Campbell, 2004). Although some speculate about the generalizability of research findings when a small number of participants are employed, "there is no evidence that findings from single-case research are any less generalizable than findings from between-group research" (Kazdin, 1982, p. 288).

The participants were volunteers and all were aware of their participation in an experimental study. All participants were females, with an age range from the early 20s to late 50s. Although the authors were unable to find research that examined the differences between male and female sensitivity to safety performance change, the results of the present study should be interpreted with slight caution because all participants were female.

The participants had been working in their current (or a similar) position (i.e., one that required sitting, typing and talking on the phone) for an average of 6.1 years (*SD:* 4.4 years; range: 1 to 15 years). Eight participants had never received any form of ergonomics training, 2 had received information on how to arrange items on their desk to prevent strain injuries, and 1 was taught hand stretches to reduce tendonitis.

A primary function of these employees' jobs is to use keyboards to enter data. They work at computer-oriented workstations for their entire shift, which places them at risk for various musculoskeletal disorders (MSDs) (NIOSH, 1997; OSHA, 1999a, 1999b). Hospital administrators wanted to focus on this area because of these known risks and because some employees in these units had been treated for MSDs. In general, the hospital had a positive safety culture, meaning that administrators supported safe work practices and would purchase safety equipment when asked to do so. However, since resources were limited, training was lacking as were proactive work analyses to prevent MSD-related injuries.

A research protocol was devised to assess the effects of: 1) providing employees with ergonomics information and training; 2) having employees conduct observations of safety performance; 3) providing daily feedback on safety performance; and 4) accuracy of observations on behavior change. With respect to information and training, participants received a handout depicting safe and unsafe body positions, as well as objective definitions for each behavior to be observed. In addition, an experimenter demonstrated each behavior in person. Information was derived from the International Business Machines (IBM) website (IBM, 2001).

With respect to observations, employees conducted them once a day alongside experimenters. The observations involved coworkers who did not conduct observations as part of the study. Participant observers also were trained on each dependent variable and observed each behavior alongside the experimental staff. Shortly after conducting an observation, participant observers were also observed for safety performance by the experimental staff.

During the feedback condition of the study, all participants received feedback 5 to 20 minutes before a second daily observation session was conducted. This information was presented in numerical form and was based on each participant's individual safety performance (as an average of all dependent variables) during the first observation session of the same day. A baseline phase that followed the same observation protocol as the following phases—but without any intervention being applied (except for observer presence while data were collected)—was used as a comparison for the phases which contained an experimental treatment.

The study lasted approximately 4 months, with 4 days of follow-up data collection occurring 8 months after the beginning of the study (4 months after the first sequence of data collection had ceased). Data collected throughout the study by experimental observers was used for data analysis and feedback, whereas data collected by participant observers were only used for calculations of observer accuracy.

Dependent variables measured were:

1) **Head and neck position.** The neck should be aligned with the back, and the eyes should be level with or slightly above the screen and document. The head should be upright.

2) **Back and shoulder position.** Back should be upright, parallel to and up against the back of the chair. Shoulders should be in line with the back and the hips.

3) **Wrist position.** Wrists should be in line with the elbows, not bent/extended upward or downward.

4) **Foot position.** Both feet should be flat on the floor and the ball of foot and heel should touch the floor or footrest if used.

As a result of the interventions applied, overall safety performance (the combined average of all four dependent variables) improved for all participants. Table 1 shows the mean safety performance and *SD* for each dependent variable per participant across all experimental conditions as well as the overall safety performance for each participant (e.g., Participant 2 averaged 89%, 92%, 99%, 98% and 100% on safe neck position during baseline, information, observation, feedback and follow-up conditions, respectively).

A *d* statistic, which determines the effect size of an intervention, was calculated for the participant observer and nonobserver groups in each condition. This measure provides information on the magnitude of an intervention effect in a standardized metric (Cohen, 1969). The larger an effect size, the greater the difference between the baseline and intervention means.

Table 2 lists the overall effect sizes for each experimental condition across groups. Safety performance for participant observers during the information,

Abstract: *A study was conducted in an applied setting to improve the safety performance of 11 computer terminal operators. The study evaluated behavior change as a function of conducting safety observations. It also provided a preliminary examination of the effects of observer accuracy on behavior change. The four aspects assessed by the research team were: 1) the effects of conducting safety observations; 2) the effects of observation accuracy; 3) the effects of prior injuries on safety performance improvement; and 4) the effects of the time of day that feedback is delivered, or time in reference to the delivery of feedback before or after performance. This article examines each effect, provides data-based analyses of them and discusses their possible implications.*

Table 1

Safety Performance Summary

Participant	Behavior	Baseline M	Baseline SD	Information M	Information SD	Observation M	Observation SD	Feedback M	Feedback SD	Follow-up M	Follow-up SD
1	Neck	85	9.9	90	12.5			99	3.6	99	1.2
	Back/shoulder	24	26.6	18	15.5			52	28.7	98	2.9
	Wrist	97	12	100	0			100	0	100	0
	Feet	53	34.3	88	18.5			97	6.6	100	0
	Overall	57	15.1	68	8.4			84	9.4	99	1
2	Neck	89	11.2	92	9.4	99	2.3	98	3	100	0
	Back/shoulder	39	41.5	1	2.8	6	10.5	35	29.3	100	0
	Wrist	97	10.4	93	13.9	99	1.9	100	0	100	0
	Feet	6	15.8	5	14	4	9.9	76	30.3	85	26.4
	Overall	51	12.1	39	6.4	42	7.3	73	14.1	95	7.5
3	Neck	94	6.9	99	2			100	0.9	100	0
	Back/shoulder	38	35.5	22	28.4			62	27	100	0
	Wrist	97	6.2	98	5.2			100	0	100	0
	Feet	15	21	17	23.2			67	35.6	65	26
	Overall	59	11.4	56	11.2			81	13.4	91	6.9
4	Neck	92	8.4	98	3.2			97	6	100	0
	Back/shoulder	15	27.7	5	9.4			17	23.8	60	21.6
	Wrist	87	29.2	96	10.8			96	14.2	100	0
	Feet	45	48.2	67	39.1			88	16.3	99	1.5
	Overall	57	9.1	60	12.6			67	8.1	86	7.6
5	Neck	99	2.1	97	2.8	95	8.9	99	2	100	1
	Back/shoulder	20	30.5	4	6	15	30.5	34	31.3	99	2.5
	Wrist	92	11.7	98	4	100	0	100	0	100	0
	Feet	74	23.2	92	9.5	92	9.3	97	6.4	59	46.2
	Overall	67	6	68	5.1	69	11.6	78	10.2	87	13.4
6	Neck	88	11.6	97	4.3	94	4.7	99	3.5	99	1.2
	Back/shoulder	16	26.8	2	4	10	15.9	54	41.1	96	7.5
	Wrist	98	4.6	99	2.6	100	0	100	0	100	0
	Feet	27	20.5	89	12.1	84	17.5	97	7	99	1.2
	Overall	57	7.4	68	5.2	68	9.8	85	12.6	98	1.7
7	Neck	95	9.1	97	4.5			100	0.7		
	Back/shoulder	28	33.9	42	29.5			71	29.9		
	Wrist	83	31.2	97	11.2			100	0		
	Feet	69	26.5	92	13.4			99	1.1		
	Overall	66	15.6	79	10.3			91	8.2		
8	Neck	96	4.9	97	4	100	0.6	99	2.7	100	0
	Back/shoulder	10	23.2	1	1.7	33	38.6	50	46.1	34	26
	Wrist	84	27.9	90	30	100	0	98	4.2	100	0
	Feet	6	13.4	34	41.9	95	7.8	94	8.4	88	13.9
	Overall	45	10.9	51	10.4	80	10.8	85	10.3	76	10
9	Neck	97	4.7	94	7.6	99	1.7	100	0	99	2.5
	Back/shoulder	20	21.8	13	12.4	48	32.8	78	19.7	87	26
	Wrist	92	23.9	99	1.7	100	0	100	0	100	0
	Feet	5	10.4	26	29	82	14.3	94	6.1	83	33.2
	Overall	48	6.9	51	10.2	79	12.1	92	6.1	91	17.7
10	Neck	80	15.7	95	5.1	95	5.3	99	2.1	100	0
	Back/shoulder	8	15.8	1	2.5	1	1.9	20	19.1	83	19.4
	Wrist	94	22.9	98	6.6	94	11.1	100	0	100	0
	Feet	5	11.4	17	17.5	66	28.9	70	22.2	64	25.4
	Overall	35	6.9	43	5.2	57	9.6	66	10	83	11.4
11	Neck	96	7.7	98	4			99	1.2		
	Back/shoulder	41	39.9	37	22.3			46	25.9		
	Wrist	95	11.6	98	4.2			100	0		
	Feet	29	41.5	87	15			98	2.9		
	Overall	58	15.2	75	8.3			84	7.6		

Note. *A summary of the mean* (M) *and standard deviation* (SD) *for each dependent variable, for each participant, across all experimental conditions.*

Table 2

Overall Effect Sizes

	Experimental condition			
Participant type	**Information**	**Observation**	**Feedback**	**Follow-up**
Observers	0.85	1.98	3.19	4.06
Nonobservers	0.76	N/A (0.76)	2.06	3.85

Note. *The overall effect sizes for the participant observer and nonobserver groups in each condition (in relation to baseline performance).*

observation and feedback phases resulted in effect sizes of 0.85, 1.98 and 3.19, respectively, in relation to baseline performance. In other words, the difference between baseline and feedback mean safety performance was larger than the differences between observation and baseline mean performance, and information and baseline performance.

The results also indicated that the safety performance of participant observers resulted in larger effect sizes than participants who did not conduct observations. Table 2 shows how similar the performance of the two groups was when exposed to the same ergonomics information and training, as well as the large difference that occurred when one group conducted one behavioral safety observation per day. The data from this study illustrate the benefit of having employees conduct behavioral safety observations and suggest that the process of conducting observations can serve as a performance improvement strategy on its own.

Figure 1

Relationship between Observer Accuracy & Percent Safe Scores, Participant 10

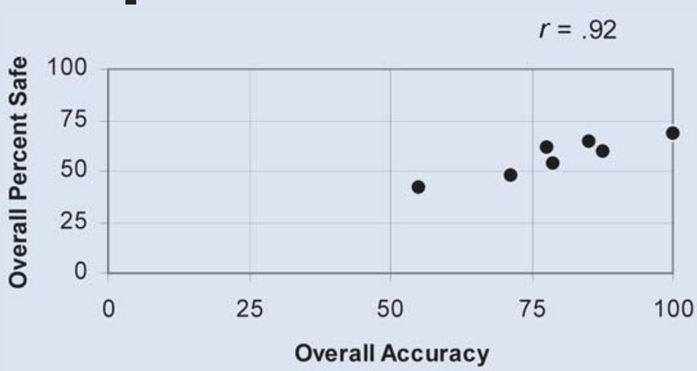
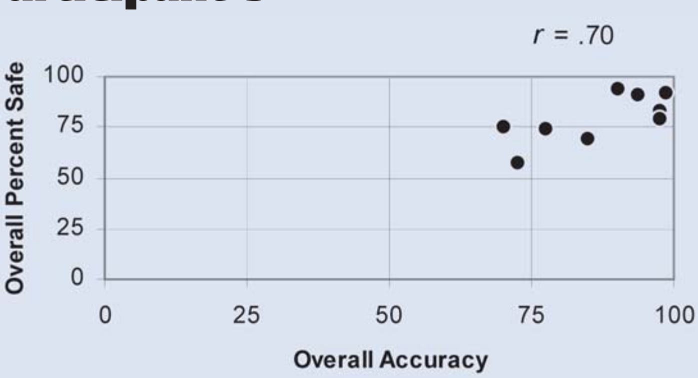


Figure 2

Relationship between Observer Accuracy & Percent Safe Scores, Participant 9



These results also revealed topics for behavioral safety researchers and practitioners to consider. This article shares a selection of those topics and their possible implications for practitioners. Although the current study addresses the effects of behavioral safety observations with a group of computer terminal operators, it is plausible that similar observer effects would carry over to other settings (e.g., manufacturing or heavy industrial settings) because the primary components of behavioral safety processes have little variation. Further research across various settings is necessary to support this hypothesis, however.

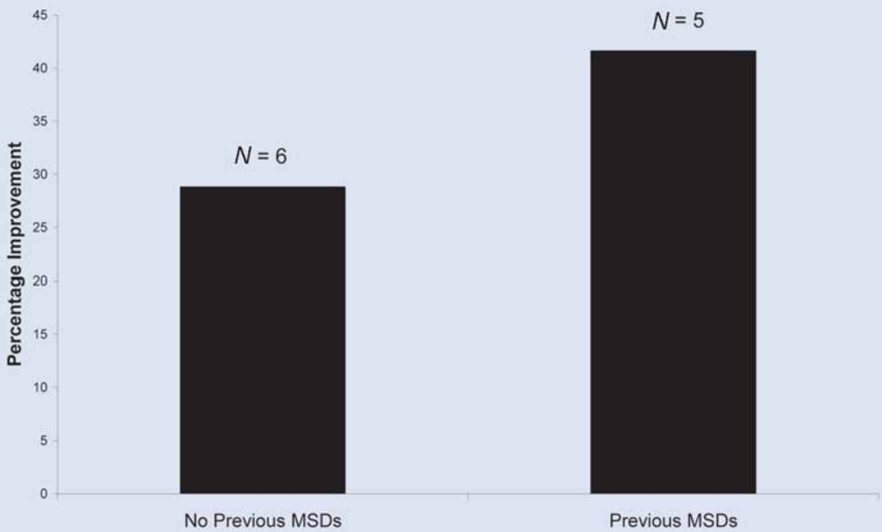
Finding 1: Conducting Observations for Safe Behavior Can Increase Safety Performance

The results show that conducting observations for safe behavior can increase the observer's safety performance. In this case, employees who conducted observations performed more safely than employees who did not conduct observations. All employees were exposed to the same information and training, so the only difference was that one group of participants conducted observations while the other did not.

After the training phase of the study was completed, participants were randomly divided into two groups: observers and nonobservers. Participant

Figure 3

Comparison between Participants Who Had & Had Not Been Diagnosed/Treated for Prior MSDs



Note. *Comparison of the percentage improvement over baseline between participants who had and had not been diagnosed and treated for prior MSDs.*

observers accompanied the experimental staff on one observation per day when the staff conducted an observation of a nonobserver participant. As a result of this, those who conducted the observations performed more than 2.5 times better (on average) than those who did not conduct observations (Table 2). These results suggest that having employees conduct behavioral safety observations may help achieve the maximum effectiveness of a behavioral safety system.

Finding 2: The Accuracy of Observations Might Play a Role in Behavior Change

The second finding was that observer accuracy was highly correlated with behavior change for some participants. In other words, when participants observed behavior more accurately, the observers themselves experienced greater performance gains. To assess the accuracy of employee observations, every employee observer was accompanied by a member of the research team on each occasion. Any deviations from the experimental observer's recordings were counted as instances of inaccurate recording by the employee observer. These inaccuracies were used to calculate the accuracy percentage of participant observations (number of accurate recordings were divided by the number of accurate plus inaccurate recordings and multiplied by 100%).

To provide a measure of relatedness, correlation coefficients were calculated between measures of each employee observer's observation accuracy and his/her level of safety at work. In some cases, the relationship was so strong that correlation coefficients of $r = .70$ and $r = .92$ were obtained (a perfect correlation between two occurrences is calculated as 1.0) (Figures 1 and 2, p. 29). Although the correlation coefficients represent a measure of relatedness and not a cause-and-effect relationship, the magnitude of the correlation coefficients warrants further research in this area. If a cause-and-effect relationship is determined, then future research could concentrate on increasing observer accuracy and finding the point of diminishing returns with regard to observer accuracy training.

Finding 3: The Effects of a Prior MSD on Behavior Change

The third finding of interest was the effects of prior MSD on behavior change. At the beginning of the study, all participants were asked whether they had been treated for any type of MSD. Although participants may have reported that they had an MSD or were experiencing some type of pain, only participants who were diagnosed and treated for an MSD were included in this sample ($N = 5$). The data show that those with a history of MSD experienced greater benefits (more positive behavior change) from the behavioral safety process than those who had not previously experienced an MSD.

One possible explanation for this finding is that a person who has experienced an MSD may have more to gain by changing behavior. Perhaps having experienced discomfort while engaging in the targeted behaviors made these participants more sensitive to the intervention designed to improve those behaviors and, thus, increased the likelihood for behavior change.

Overall measures of behavior change were calculated for each participant by subtracting each individual's overall percent safe during baseline from her overall percent safe during the final phase in which data were collected. The average improvement (calculated in percentage points) for those who had no history of an MSD was 28.8 percentage points, while the average improvement for those with a record of an MSD was 41.6 percentage points (Figure 3).

This difference provides another avenue for future research for several reasons. For example, researchers and practitioners might wish to determine whether different strategies work best with those who have (or have not) had MSDs or a particular MSD, or to determine the effectiveness of different strategies in relation to specific MSD/improvement strategy combinations. This line of investigation could also be followed to determine whether employees who have experienced any type of injury are more likely to respond favorably to behavioral intervention.

Finding 4: The Effects of the Time of Feedback Delivery

Another noteworthy finding was the increase of performance directly after numerical feedback was provided. Participants were observed twice a day—once in the morning and once in the afternoon. During the morning sessions, half of the participants conducted observations of their peers, while the other half were observed by their peers. All participants were also observed in the morning and afternoon by experimental staff.

After the morning session, experimenters calculated a percentage safe for each participant for each target behavior. Participants received written numerical feedback on their morning performance 5 to 20 minutes before the afternoon observation session began. Participants never received feedback on their afternoon performance. Furthermore, participants were informed of the procedure being used to calculate their feedback and understood that each afternoon's feedback was based on their performance that morning.

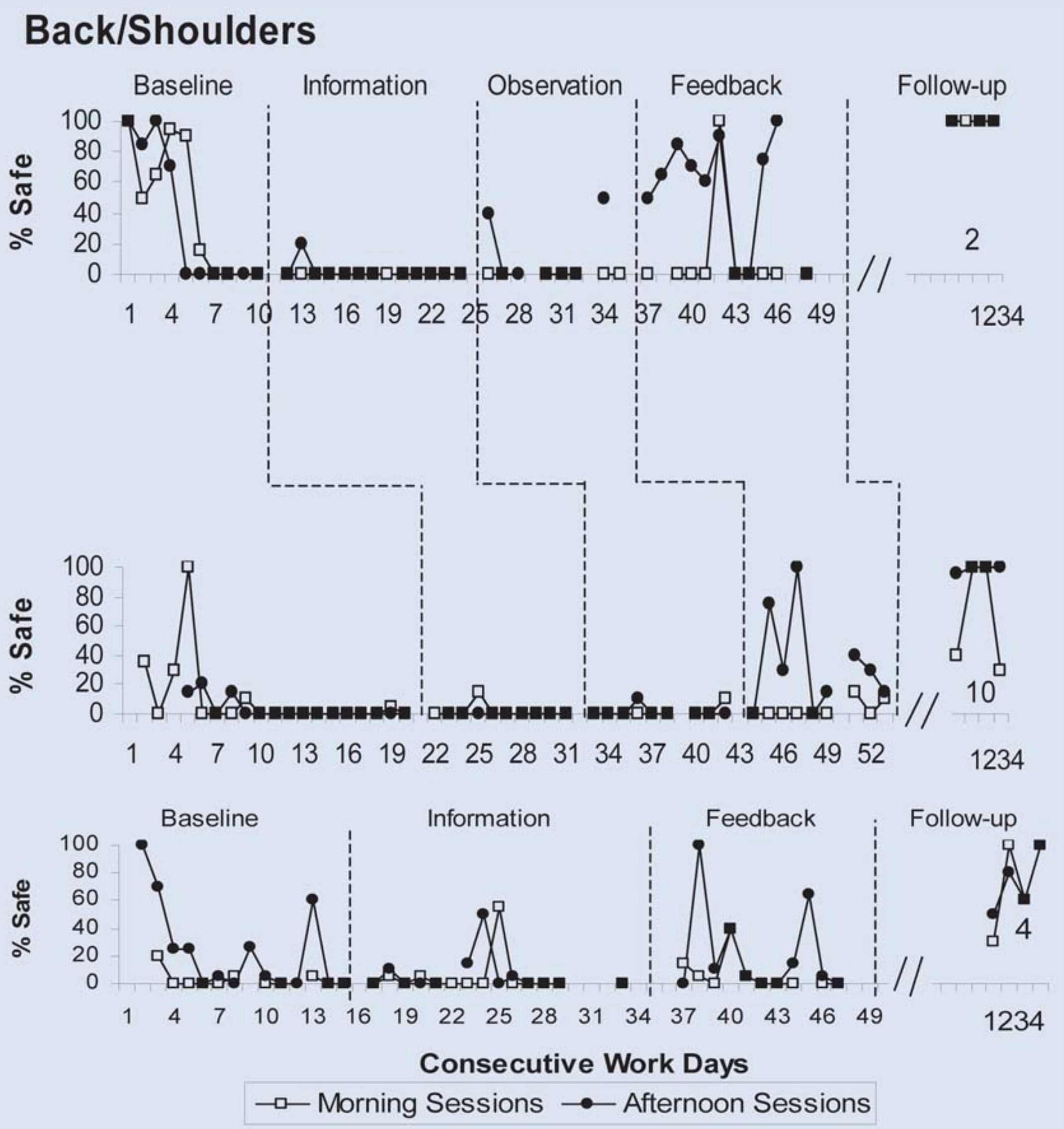
Some participants exhibited a pattern in which afternoon performance was significantly higher than morning performance (Figure 4). Having been told that feedback received in the afternoon was based on morning performance, all participants were aware that to receive higher scores they would have to perform well in the morning, and that afternoon sessions did not affect any feedback ratings. Instead of improving morning performance, some participants improved performance after the delivery of feedback in the afternoon (in which the feedback often reflected low levels of safe behavior).

This trend occurred somewhat cyclically on a daily basis—meaning that each morning performance would be low, then, after receiving feedback, performance would be higher, day after day. These

Figure 4

Figure 4

Effects of Providing Written Feedback



Note. The effects of providing written feedback based on morning performance 5 to 20 minutes prior to conducting an observation of afternoon performance (demonstrated in the feedback phase of each graph).

circumstances suggest that it would be beneficial to provide employees with feedback under an arrangement which differs from that used in the current study. Other arrangements could include providing feedback directly after behavior was observed (a common component of many behavioral safety programs) or providing a person with summary feedback or carefully chosen feedback from an earlier instance of behavior, immediately before the person is to complete a risky task (e.g., one with a high incident rate or associated with severe consequences if an incident occurs).

Results of this study suggest that SH&E professionals should not discount the behavioral observation process as an instrument for improving overall safety performance.

Considerations & Observed Effects

1) Conducting observations for safe behavior can influence safety performance.

2) Increased accuracy of observations might result in greater amounts of behavior change.

3) Special considerations might be adopted when dealing with people who have had prior injuries.

4) Controlling the timeliness and frequency of feedback delivery could be an important variable in promoting safety.

Practical Applications for SH&E Professionals

Results of this study suggest that SH&E professionals should not discount the behavioral observation process as an instrument for improving safety performance. Participants who served as safety observers showed higher levels of safety performance improvements than those who were not observers, and the accuracy of some participant observations was correlated with behavior change.

These findings suggest that SH&E professionals should place a degree of emphasis on: 1) training as many employees as possible to conduct safety observations; and 2) the actual observer training procedures. Frequent reliability checks on observer accuracy should be performed both during training and after the implementation of an observation system. In other words, 2 separate observers should perform various observations on the same persons and a reliability percentage should be calculated (number of agreements between observers/number of agreements plus disagreements and multiplied by 100%). Percentages of 80% or higher indicate reliable observations (Komaki, 1998).

These results also suggest that performance feedback should be delivered frequently, which supports previous literature (Alvero, Bucklin & Austin, 2001), and suggests that increasing the frequency of feedback will increase the degree of performance change. Therefore, SH&E professionals should aim to deliver frequent daily feedback in order to maximize the effects of a behavioral safety process.

Conclusion

Some of these findings have been shown in multiple studies whereas others have been largely ignored by the research literature. These results should be considered exploratory but should shape future research efforts. More importantly, once a sufficient amount of data has been collected, they should shape the way future safety programs are conducted. Although possible reasons for performance improvement have been cited for each effect observed, further research is needed to determine the true causes of the performance increases.

The underlying message is that what behavioral safety consultants have been saying for years could be more than just educated guesses, intuition and years of experience. The theoretical benefits that have been discussed by consultants and safety authors are likely joined by actual performance benefits, and those benefits may be even more powerful when paired with increased observer accuracy. ■

References

Alvero, A.M. & Austin, J. (2003). The observer effect. In T. McSween (Ed.), *The values-based safety process* (2nd ed.). New York: John Wiley & Sons Inc.

Alvero, A.M. & Austin, J. (2004). The effects of conducting behavioral observations on the behavior of the observer. *Journal of Applied Behavior Analysis, 37,* 457-468.

Alvero, A.M., Bucklin, B.R. & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998). *Journal of Organizational Behavior Management, 21*(1), 3-29.

Austin, J., Kessler, M.L., Riccobono, J.E., et al. (1996). Using feedback and reinforcement to improve the performance and safety of a roofing crew. *Journal of Organizational Behavior Management, 16*(2), 49-75.

Cohen, J. (1969). *Statistical power analysis for the social sciences.* New York: Academic Press.

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, E.S. (2001). *Working safe: How to help people actively care for health and safety* (2nd ed.). Boca Raton, FL: Lewis Publishers Inc.

International Business Machines (IBM). (2001). Healthy computing. Purchase, NY: Author. Retrieved March 8, 2007, from http://www.pc.ibm.com/ww/healthycomputing.

Kazdin, A.E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

Komaki, J. (1998). *Leadership from an operant perspective.* London: Routledge.

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.

Krause, T.R. (1997). *The behavior-based safety process: Managing involvement for an injury-free culture* (2nd ed.). New York: Van Nostrand Reinhold.

McSween, T.E. (2003). *The values based safety process: Improving your safety culture with behavior-based safety.* Hoboken, NJ: Wiley-Interscience.

NIOSH. (1997). *A critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity and low back.* Washington, DC: Author. Retrieved March 8, 2007, from http://www.cdc.gov/niosh/docs/97-141.

OSHA. (1999a). Ergonomics. Washington, DC: Author.

OSHA. (1999b). Preventing work-related musculoskeletal disorders. Washington, DC: Author.

Sulzer-Azaroff, B. & Austin, J. (2000, July). Behavior-based safety and injury reduction: A survey of the evidence. *Professional Safety, 45,* 19-24.

Thompson, V.A. & Campbell, J.I.D. (2004). A power struggle: Between- vs. within-subjects designs in deductive reasoning research. *Psychologia: An International Journal of Psychology in the Orient, 47*(4), 277-296.