A COMPARISON OF MULTIPLE INDICATORS

Observations, Supervisor Report, and Self-Report as Measures of Workers' Hearing Protection Use

SALLY LECHLITNER LUSK
University of Michigan
DAVID L. RONIS
University of Michigan and Department of Veterans Affairs
LESLIE MARTEL BAER
University of Michigan

The purpose of this study was to compare multiple indicators of behavior and identify the most viable measure of blue-collar workers' use of hearing protection. Three measures of use were employed: observations, supervisor report, and self-report. Supervisor report was highly discrepant from both self-report and observed use; self-report and observations were highly correlated, and discrepancies between the two were slight. These results suggest that, for this type of measurement, self-report is an appropriate measure and may be the best choice when time and monetary resources restrict measurement to one indicator.

AUTHORS' NOTE: This study was supported by Grant No. R01NR02050 from the National Center for Nursing Research and the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The authors would like to thank Madeleine J. Kerr, PhD, MS, RN, for her contributions to preparing the data for this report.

EVALUATION & THE HEALTH PROFESSIONS, Vol. 18 No. 1, March 1995 51-63
© 1995 Sage Publications, Inc.

from the SAGE Social Science Collections. All Rights Reserved.

The Occupational Safety and Health Administration (OSHA) estimated that 14 million workers in the United States are exposed to hazardous noise at the worksite (Occupational Safety and Health Administration [OSHA], 1980), which can cause deleterious physiological and psychological stress reactions, as well as destroy hearing. As the National Institute of Occupational Safety and Health (NIOSH) has stated, "noise-induced hearing loss is a progressive injury" that can be traced to a single causative agent—noise—and is therefore preventable (National Institute of Occupational Safety and Health [NIOSH], 1986). Although hearing protection equipment is available to U.S. workers as required by OSHA, many do not consistently use it, although they have been advised to do so.

Several levels of personnel in a company—workers, supervisors, and managers—are responsible for protecting workers' hearing (Royster & Royster, 1986). Although workers are responsible for their individual use, supervisors are typically given the burden of enforcing hearing protection equipment policies; active support by management also increases the effectiveness of hearing conservation programs (Royster & Royster, 1986). A series of appropriate interventions, in which workers exposed to noise had repeated contact with information on and instruction in the use of hearing protection, would encourage the use of hearing protection equipment. However, very few empirical studies have been conducted to guide the focus and content of these interventions. Finding a valid indicator of this behavior is an important first step to developing the research necessary to guide future interventions.

Observations, often seen as a superior indicator of behavior, require considerable investment of time and money and may be limiting for some research, both financially and in terms of the range of behavior that may be studied. Observations may also affect the performance of the behavior due to the presence of the observers. Proxy reports (reports given by someone other than the individual carrying out the behavior) are often greeted with skepticism due to concerns over a proxy's ability to access the needed information to report reliably (Blair, Menon, & Bickart, 1991; McLaughlin, Mandel, Mehl, & Blot, 1990; Mingay, Shevell, Bradburn, & Ramirez, 1994); in addition, social desirability bias can influence proxy reports, just as it can affect self-reports (McLaughlin et al., 1990).

Self-report is a frequently used indicator of behavior in research studies and often considered a fast and relatively low-cost means of collecting data. However, the reliability and validity of such information is frequently brought into question due to the following concerns: (a) limitations in subjects' ability to recall the events and to locate them correctly and in time (Sudman & Bradburn, 1974); (b) biases from the use of cognitive processes other than recall to estimate the frequency of subjects' own past behaviors (Blair et al., 1991; Bradburn, Rips, & Shevell, 1987; Burton & Blair, 1991; Tversky & Kahneman, 1974); and (c) the effects of social desirability bias on self-reports (Rosenberg, 1969).

Studies that are adequate comparisons of multiple indicators of health-related behaviors generally fall into three categories of verification: biochemical confirmation, observations, or reports by others. We are concerned here only with the latter two. Several studies have compared the relative reliabilities and validities of self-report and observations of health-related behaviors, reporting a wide range of methods and results (Crockett, Potter, Wright, & Bacheller, 1992; Jette, 1987; Robertson, 1992). For example, one researcher found high discrepancies between self-report and observations of drivers' seat belt use and alcohol consumption; however, the self-reporters were not the same subjects as those observed (Robertson, 1992). In contrast, researchers who compared self-report and trained interviewer report of home shelf inventories, did evaluate the same group of subjects with both indicators; self-report was found to have high overall sensitivity and specificity, as well as significant overall agreement with the interviewer-completed survey (Crockett et al., 1992).

Several studies have compared proxy and self-reports in the ongoing debate regarding these two indicators (Blair et al., 1991; Mingay et al., 1994; Moore, 1988; Williams, Thomas, Young, Jozwiak, & Hector, 1991; Wynd, 1992). Although results from such studies have remained inconclusive as to the best methods of comparing the two indicators and the nature of their relationship, some progress has been made in identifying the different methods respondents use in proxy and self-reporting. Blair et al. (1991), for instance, found that proxies tend to estimate the frequency of events, whereas self-reporters use more recall.

The authors of the Health Habits Scale specifically built into their study a method for comparing proxy and self-responses to the scale items; self-reports and proxy reports demonstrated good agreement based on Pearson correlations (Williams et al., 1991). However, results from a range of studies have produced inconclusive evidence as to which type of report is most accurate (Moore, 1988). Although many comparisons between proxy and self-reports have been made, few have met the criteria outlined by Moore (1988) for adequately examining this issue: (a) investigating behaviors that are not *shared* experiences, but have a clearly defined "self"; (b) a clear control for how proxy and self-reports are solicited, thus reducing self-selection bias; and (c) a means of evaluating both responses by an objective measure. Even among studies that meet all three criteria, Moore (1988) concluded the data to be equivocal; for example, some studies found overreporting in proxy responses, whereas others found it in self-report.

In designing a study that meets all three criteria (Moore, 1988), our aim was to test multiple indicators of a health-related behavior, the use of hearing protection, and to compare results from observations, supervisor report, and self-report.

METHOD

Data regarding the use of hearing protection were collected in one midwestern automotive plant in three phases. Blue-collar workers' use of hearing protection was observed in the first phase. In the second phase, supervisors of the observed workers were interviewed regarding their subordinates' use of hearing protection equipment. Finally, all workers in the plant (N = 4,473) were invited to participate in a survey about health and hearing protection, which included the self-report items on use of hearing protection equipment.

SAMPLE

Comparisons of the three indicators were made for blue-collar workers only. Of the observed workers (n = 185), supervisor reports were available from eight supervisors. Of the workers for whom both

observed report and supervisor report of use of hearing protection was available (n = 174), 48 volunteered to fill out the questionnaire, thus supplying their self-report of use.

A total of 645 blue-collar, skilled trades, and white-collar workers completed the questionnaire, and of the 185 blue-collar workers who had been observed, 55 completed a questionnaire. Although a relatively small percentage of the observed population, this number represented a higher proportion of the observed workers (30%) than usually obtained in other voluntary activities at the plant (typically 2% to 3%).

For the total sample of blue-collar workers (n = 318) who gave self-reported data on the questionnaire, ages ranged from 24 years to 64 years with a mean of 41.6 years; most were male (78.9%), non-Hispanic white (86.9%), and married (72.2%), and a large minority (43.2%) had completed some education beyond high school. This sample was similar in these characteristics to the smaller sample of 48 blue-collar workers whose use of hearing protection was assessed by all three indicators; their ages ranged from 25 years to 56 years with a mean of 38.0 years; most were male (68.8%), non-Hispanic white (84.8%), and married (76.6%). A large minority (38.3%) had completed some education beyond high school.

Observations

Using noise dosimeters, two trained observers measured the noise level in the vicinity of each worker in five departments on day and evening shifts. Measurements were taken every half hour for a half day at a time for a minimum of two nonconsecutive half days. While recording the noise level, the observers surreptitiously recorded the use or nonuse of hearing protection. This process resulted in 14 to 58 observations for each worker, with a mean of 31 observations, over a period of 5 weeks. Interrater reliability was established at .94 by having a third trained observer join each of the other two observers for four to five sessions of observations.

Supervisor Report

Supervisors of the observed blue-collar workers were asked for their impressions of their subordinates' use of hearing protection.

Their responses to the following items, for each of the workers they supervised, were obtained: "What percent of the time would you say each of your workers wore their hearing protection while in this workstation area during the (1) last week? (2) last month? (3) last 3 months?" Supervisor reports were unavailable for 11 of the workers for one of three reasons: (a) some workers were no longer available for the study by this phase; (b) one of the supervisors refused to be interviewed; and (c) no supervisor could be identified for some of the workers.

Self-Report

In the final phase, all workers in the plant ($N = 4,473$) were invited to complete a written questionnaire that included a self-report measure, in the form of three questions, of the percentage of time that hearing protection was used: "(1) During the past *week* in your work area, what percent of the time would you say you *actually* used hearing protection? (2) During the past *month* in your work area, what percent of the time would you say you *actually* used hearing protection? (3) During the past *3 months* in your work area, what percent of the time would you say you *actually* used hearing protection?" The observation of hearing protection use and the administration of the questionnaires were conducted by different research teams to prevent workers from connecting the observers to the subsequent administration of the health and hearing protection questionnaire through which self-report was obtained.

The three indicators could not be compared for 7 of the 55 workers who were observed and provided self-reports, due to the missing supervisor data. Therefore, the subsequent analyses include only the 48 blue-collar workers whose use of hearing protection was assessed by all three indicators.

ANALYSIS AND RESULTS

Table 1 shows the frequency of use of hearing protection as indicated by the three approaches to measurement. In the case of supervi-

TABLE 1
Percentage of Use of Hearing Protection by Method of Assessment

	Observed		*Supervisor Report*		*Self-Report*	
	n	*% of Workers*	n	*% of Workers*	n	*% of Workers*
0%[a]	15	31.3	7	14.6	13	27.1
20%	4	8.3	3	6.3	4	8.3
40%	3	6.3	5	10.4	1	2.1
60%	2	4.2	0	0.0	1	2.1
80%	4	8.3	4	8.3	5	10.4
100%	20	41.7	29	60.4	24	50.0
Total	48	100.0	48	100.0	48	100.0
Mean %		54.4		73.5		62.3
SD		44.1		37.3		43.9
Range		0.0-100.0		0.0-100.0		0.0-100.0

a. Rounded to nearest 20%.

sor report and self-report, these frequencies represent means of the reported use for the three time periods (past week, past month, and past 3 months). The differences among the three time periods were essentially nil; for self-report, the correlations among the measures for the three time periods ranged from .96 to .99, and for supervisor report, correlations ranged from .91 to .96.

The extremely high correlations among the reports of use during the past week, past month, and past 3 months suggest that one of two processes may have been used by respondents: (a) workers and supervisors have mental schema—information about experiences of a series of similar events organized together—about their own and their subordinates', respectively, use of hearing protection, and these schema are independent of time (Brewer, 1994; Menon, 1994); or (b) workers and supervisors construct an answer based on the past week (a relatively easy mental task) and assume that percentage of use is the same for longer periods (Blair & Burton, 1987; Menon, 1994). Although actual use may be very consistent, we do not conclude that these reports are reliable just because of their high correlations.

As may be seen in Table 1, all three measures indicated that hearing protection was worn the majority of the time and that different workers varied considerably in their use. Use of hearing protection also displayed a bimodal distribution; all three measures indicated that most

workers used hearing protection almost all of the time or almost never, whereas fewer used hearing protection around half the time.

We computed Pearson, Spearman, and Kendall correlations to assess convergent validity of the three indicators. Nonparametric correlations were computed in addition to Pearson's r because of the extreme nonnormality of the distributions. Observed and self-reported use correlated highly with each other (from .69 to .89, depending on the specific correlation coefficient examined; see Table 2). Observations and self-report had low correlations with supervisor report, ranging from .33 to .47. Thus the convergence between observations and self-report was much higher than between supervisor report and either of the other measures.

We used repeated measures of analysis of variance to test the statistical significance of the apparent differences in mean use of hearing protection among the three indicators. Mean hearing protection use differed significantly across the three indicators, $F(2, 94) = 6.93, p = .002$, with lowest use indicated by observation ($M = 54.5\%$). Self-reported use was 7.9% higher (62.3%), and supervisor report was an additional 11.2% higher (73.5%). Dependent t tests indicate that the difference between observation and self-report was statistically significant, $t(47) = 2.72$, $p = .009$, as was the difference between observations and supervisor report, $t(47) = 3.15$, $p = .003$. However, the difference between self-report and supervisor report did not reach significance, $t(47) = 1.90, p = .063$. The difference between supervisor and self-report, although larger than the difference between observations and self-report, was nonsignificant because the standard error of the difference between supervisor and self-report (.85) was much larger than the standard error of the difference between observations and self-report (.42). The higher standard error was due to the lower correlation between supervisor report and self-report.

Discrepancies between observations and self-report were small, but supervisor report was highly discrepant from both observed and self-reported use (Table 3). As indicated in the lower panel of Table 3, the mean absolute discrepancy (or average size of a discrepancy) was smallest between self-report and observations and largest between supervisor report and observations. Self-report and observations were within 5% of each other for 56.3% of the workers. They were within

TABLE 2
Correlations Among the Measures of Hearing Protection[a]

	Self-Report	*Supervisor Report*
Pearson product moment correlations		
Observed use	.89**	.47**
Supervisor report	.50**	
Kendall's tau-b		
Observed use	.69**	.33*
Supervisor report	.38**	
Spearman's rho		
Observed use	.84**	.42*
Supervisor report	.49**	

a. $N = 48$ for all correlations.
*$p < .01$; **$p < .001$.

10% of each other (with discrepancies in Table 3 shown rounded to 0%) for 70.8% of the workers. In contrast, supervisor report and self-report were within 5% of each other for 47.9% of workers and within 10% of each other for 56.3% of the workers. Supervisor report and observations were within 5% of each other for 31.3% of the workers and within 10% of each other for 39.6% of the workers.

For some workers, there were also large discrepancies among the indicators, especially for supervisor report. One worker had a discrepancy greater than 70% between self-report and observed, whereas eight had discrepancies as large between supervisor and self-reports, and nine had discrepancies as large between supervisor report and observations. Some of the discrepancies between supervisor report and the other measures were 100%, indicating that the supervisor said the worker always used hearing protection and the other measures indicated the worker never used hearing protection, or vice versa.

DISCUSSION

When time and money are limited for a behavioral study, it is not unusual for one indicator to be selected; naturally, if more than one indicator is feasible, the most workable one needs to be identified. In determining which one of several indicators of a behavior is the most

TABLE 3
Discrepancies Among the Measures of Use of Hearing Protection

	Self-Report Minus Observed		*Supervisor Minus Self-Report*		*Supervisor Minus Observed*	
Difference[a]	n	*% of Workers*	n	*% of Workers*	n	*% of Workers*
100%	0	0.0	3	6.3	4	8.3
80%	1	2.1	3	6.3	3	6.3
60%	2	4.2	0	0.0	0	0.0
40%	3	6.3	5	10.4	9	18.8
20%	5	10.4	4	8.3	6	12.5
0%	34	70.8	27	56.3	19	39.5
–20%	2	4.2	3	6.3	5	10.4
–40%	1	2.1	0	0.0	0	0.0
–60%	0	0.0	1	2.1	0	0.0
–80%	0	0.0	1	2.1	0	0.0
–100%	0	0.0	1	2.1	2	4.2
Total	48	100.0	48	100.0	48	100.0
Mean %		7.9		11.2		19.2
SD		20.2		40.9		42.2
Range		–36.1 to 76.7		–100.0 to 100.0		–100.0 to 100.0
Mean absolute discrepancy		11.7%		25.7%		30.7%
SD		2.6		4.8		5.0

a. Rounded to nearest 20%.

viable, several factors must be taken into consideration: validity and reliability, time constraints, and financial limitations. Because an indicator is of little use if it is not valid and reliable, we begin with those requirements in the process of narrowing our field of three indicators.

Supervisor report weakly correlated with self-report and observations, suggesting substantial random error, whereas self-report and observations were strongly correlated, implying little random error. Discrepancies of supervisor report from the other measures were often large, averaging 26% away from self-report and 31% away from observed use. Large discrepancies were found in both directions, again suggesting substantial random error. These results suggest that supervisor report provides only a crude assessment of use of hearing protection, and they are consistent with the higher errors of proxy reports rather than self-reports found by Mingay et al. (1994). They

are also consistent with the argument that proxies generally have less of the needed information available to them (Mingay et al., 1994).

During their interviews, supervisors acknowledged their difficulty in estimating the percentage of time that their workers used hearing protection. Although supervisors are charged with the responsibilities of training workers to use hearing protection equipment and enforcing the policies requiring its use, they apparently feel unable to focus adequate attention on this aspect of worker behavior. The other demands of their positions, such as production schedules, appear to require more of their attention; in addition, because of their felt or acknowledged responsibility for this worker behavior, they may tend to expect and perceive greater compliance by workers than actually occurs. Whether due to biases or the lack of data available to the respondent, results from this study clearly indicated that supervisors' perceptions did not provide accurate data.

Because supervisor report is an inadequate indicator of workers' use of hearing protection, the remaining two indicators, self-report and observations, must be considered. Although average self-reported use of hearing protection was significantly higher statistically than observed use, the actual values reported by these two methods were usually close to each other, and self-reported and observed use of hearing protection correlated very highly with each other. In addition, self-report and observations correlated about equally with supervisor report, providing no evidence favoring one or the other. These results strongly suggest that both self-report and observations are valid and reliable methods of assessing use of hearing protection. Additional information must be taken into account when selecting the better of these two remaining indicators.

The reliability of self-report measures may be specific to the behavior studied or to the social desirability of the behavior, but our results suggest little social desirability bias in this self-reported measure (self-reported use was only 7.9% higher than observed use). The case can be made that, for the use of hearing protection, social desirability would be a stronger factor in supervisor report than in self-report because supervisors may feel that a high report of the use of hearing protection among their subordinates indicates that the supervisors are performing an adequate job of monitoring this behav-

ior (their report of workers' use was 19.2% higher than the workers' observed use). Therefore, for this behavior, social desirability is not a major factor in choosing the best indicator.

Obviously, the personnel costs of obtaining data through multiple observations are much greater than through self-report measures. Even when observations are as accurate and representative of behavior as another measure such as self-report, they are expensive and time consuming. In addition, management may be unwilling to have observers in the work setting. Finally, systematic observations have been thought to be the best means of obtaining true measures of behavior and are held as the "gold standard." However, this perception fails to recognize both that observations are limited in the amount of behavior that can be observed and that the behavior may be influenced by the presence of the observer.

Self-report of the use of hearing protection among blue-collar workers demonstrated high correlations with observation of use (.69 to .89). Although the difference between self-report and observation of use (7.9%) was significant, it was the smallest difference between any two of the indicators; this result implies that little social desirability bias exists for blue-collar workers giving reports of their own use. Thus, in consideration of the accuracy of the self-reported data when compared to observations, the apparently small social desirability bias for this type of data, and time and cost factors, the results from this study support the use of self-report measures. Although these conclusions may not extend to all types of behaviors, self-report is clearly an adequate, comparatively inexpensive means of obtaining large amounts of reliable data regarding workers' use of hearing protection.

REFERENCES

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research, 14*, 280-288.

Blair, J., Menon, G., & Bickart, B. (1991). Measurement effects in self vs. proxy responses to survey questions: An information-processing perspective. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 145-166). New York: Wiley.

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*, 157-161.

Brewer, W. F. (1994). Autobiographical memory and survey research. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 11-20). New York: Springer-Verlag.

Burton, S., & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly, 55*, 50-79.

Crockett, S. J., Potter, J. D., Wright, M. S., & Bacheller, A. (1992). Validation of a self-reported shelf inventory to measure food purchase behavior. *Journal of the American Dietetic Association, 92*, 694-697.

Jette, A. M. (1987). The Functional Status Index: Reliability and validity of a self-report functional disability measure. *Journal of Rheumatology, 14*(Suppl. 15), 15-19.

McLaughlin, J. K., Mandel, J. S., Mehl, E. S., & Blot, W. J. (1990). Comparison of next-of-kin with self-respondents regarding questions on cigarette, coffee, and alcohol consumption. *Epidemiology, 1*, 408-412.

Menon, G. (1994). Judgements of behavioral frequencies: Memory search and retrieval strategies. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 161-172). New York: Springer-Verlag.

Mingay, D. J., Shevell, S. K., Bradburn, N. M., & Ramirez, C. (1994). Self and proxy reports of everyday events. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 235-250). New York: Springer-Verlag.

Moore, J. C. (1988). Self/proxy response status and survey response quality: A review of the literature. *Journal of Official Statistics, 4*, 155-172.

National Institute of Occupational Safety and Health. (1986). *A proposed national strategy for the prevention of work-related, noise-induced hearing loss*. Unpublished manuscript, National Institute of Occupational Safety and Health, Cincinnati, Ohio.

Occupational Safety and Health Administration. (1980). *Noise control: A guide for workers and employers*. Washington, DC: U.S. Department of Labor and the Occupational Safety and Health Administration.

Robertson, L. S. (1992). The validity of self-reported behavioral risk factors: Seatbelt and alcohol use. *Journal of Trauma, 32*, 58-59.

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279-349). New York: Academic Press.

Royster, L. H., & Royster, J. D. (1986). Education and motivation. In E. H. Berger, W. D. Ward, J. C. Morrill, & L. H. Royster (Eds.), *Noise and hearing conservation manual* (pp. 383-416). Akron, OH: American Industrial Hygiene Association.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Williams, R. L., Thomas, S. P., Young, D. O., Jozwiak, J. J., & Hector, M. A. (1991). Development of a Health Habits Scale. *Research in Nursing and Health, 14*, 145-153.

Wynd, C. A. (1992). Personal power imagery and relaxation techniques used in smoking cessation programs. *American Journal of Health Promotion, 6*, 184-196.