8-1999

Behavior-Based Safety and Working Alone: The Effects of Self-Monitoring on the Safe Performance of Bus Operators

Ryan Olson

BEHAVIOR-BASED SAFETY AND WORKING ALONE: THE EFFECTS OF SELF-MONITORING ON THE SAFE PERFORMANCE OF BUS OPERATORS

by

Ryan Olson

A Thesis
Submitted to the
Faculty of the Graduate College
in partial fulfillment
of the requirements for the
Degree of Master of Arts
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
August 1999

ACKNOWLEDGEMENTS

I would like to thank Dr. Carl Cheney of Utah State University for introducing me to the experimental analysis of behavior. His enthusiasm for science, marvelous facts, and academic rigor spawned a liberating journey for me.

Secondly, I would like to thank Dr. John Austin of Western Michigan University for his example of productivity and his exceptional efforts in my behalf. I owe my abundant opportunities as a graduate student to him. I would also like to acknowledge the influence of my thesis committee members Dr. Alyce Dickinson and Dr. Wayne Fuqua. Dr. Dickinson introduced me, in outstanding fashion, to behavior analysis applied to work, and I owe an improved understanding of scientific experimentation to interactions with Dr. Fuqua as an instructor.

P. Adam VanAssche contributed greatly to the implementation of key aspects of this project. I hope for opportunities to reciprocate his professional courtesy and friendship. I thank research lab members Alicia Alvero, Scott Traynor, and Barbara Bucklin for their listening ears and input. Lastly, I express my gratitude to Lisa for her partnership and her enhancement of every aspect of my life.

Ryan Olson

BEHAVIOR-BASED SAFETY AND WORKING ALONE: THE EFFECTS OF SELF-MONITORING ON THE SAFE PERFORMANCE OF BUS OPERATORS

Ryan Olson, M.A.

Western Michigan University, 1999

Experimental evaluations of Behavior-based Safety (BBS) processes applied with lone workers are scarce. Further research is needed to determine the power of self-monitoring based interventions for improving safe behavior, and to explore the best practices for improving safety when employees work alone. In the current study, four male bus operators (ages 40-50) self-monitored their safe performance and initialed feedback graphs based on their self-monitoring data at the end of each day. Experimental data collectors observed each participant by riding busses as passengers. A multiple baseline design across performances was used to assess the effects of the intervention on four target performances. The intervention resulted in a 12.5% overall increase in safe performance for the group, with individual increases in safe performance that ranged from 3% to 41% for specific target performances. The results are discussed in terms of the value of BBS processes for employees who work alone and the research needed to determine the components of self-monitoring processes that are most critical for generating improvements in safe performance.

TABLE OF CONTENTS

Table of Contents---Continued

Table of Contents---Continued

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Behavior-based Safety and Working Alone

Behavior-based safety (BBS) has become a popular and successful process for improving safety performance in organizations (Geller, 1988, 1996; Krause, 1997; McSween, 1995). Two components of BBS processes are peer and/or supervisor observations of safe performance and the provision of feedback for observed employees based on those observations. However, practitioners and researchers have begun to apply BBS to jobs where people work alone. When employees work alone they can self-monitor their safe performance and either record their own percent safe scores or receive feedback generated by others from the self-monitoring data. Data from BBS applications where employees self-monitor are limited (Krause, 1997; McCann & Sulzer-Azaroff, 1996), but self-monitoring has been widely studied in other contexts (Kopp, 1988).

The current study involved the implementation of a self-monitoring procedure to improve the safe performance of bus operators. A multiple baseline design across performances was used to assess the effects of the self-monitoring process over a period of six weeks with four bus operators. The results are discussed in terms of the benefits of BBS processes for employees who work alone and the research needed to determine the critical components of self-monitoring processes for improving safe performance.

REVIEW OF LITERATURE

Overview of BBS Research

Over the past 20 years, behavioral research in the field of safety has grown steadily. The first important conceptual articles discussing the benefits of applying behavior analysis technology to improve occupational safety were published in the late 1970's (Smith, Cohen, H., Cohen, A., & Cleavland, 1978; Sulzer-Azaroff, 1978). The first experimental applications of behavioral technology applied to occupational safety occurred during the same time period (Komaki, Barwick, & Scott, 1978; Smith, Anger, & Uslan, 1978). The central foundation of all BBS research since these early applications has been the measurement of safe and at-risk behaviors and conditions, and the use of behavioral technology to increase the frequency of those safe behaviors and conditions. The body of research has demonstrated the effectiveness of many different intervention packages designed to achieve these effects.

Studies employing the use of experimental designs have examined the effectiveness of training (Cohen & Jensen, 1984; Komaki, Heinzmann, & Lawson, 1980; Reber & Wallin, 1984; Reddell, Congleton, Huchingson, & Montgomery, 1992), goal setting and prompts (Austin, Alvero, & Olson, 1998; Berry, Geller, Calef, R. S., & Calef, R. A, 1992; Engerman, Austin, & Bailey, 1997; Fellner & Sulzer-Azaroff, 1986; Ludwig & Geller, 1991, 1997; Phillips, Sutherland, & Makin, 1994;

Reber & Wallin, 1984; Reber, Wallin, & Chhokar, 1990; Saarela, 1989), verbal and graphic feedback (Alavosius & Sulzer-Azaroff, 1986, 1990; Austin, Kessler, Riccobono, & Bailey, 1996; Babcock, Sulzer-Azaroff, & Sanderson, 1992; Chhokar & Wallin, 1984; DeVries, Burnette, & Redmon, 1991; Fellner & Sulzer-Azaroff, 1984; Komaki, Heinzmann, & Lawson, 1980; Nasanen & Saari, 1987; Phillips, Sutherland, & Makin, 1994; Sulzer-Azaroff & Consuelo De Santamaria, 1980), contingent incentives and reinforcement (Fox, Hopkins, & Anger, 1987; Komaki, Barwick, & Scott, 1978; McAfee & Winn, 1989; Petersen, 1984), and self-monitoring procedures (McCann & Sulzer-Azaroff, 1996) at increasing safe behaviors and conditions.

The use of self-monitoring procedures to improve safe behavior is a relatively new development. The field of BBS is growing and reports of successful commercial applications of BBS with lone workers have begun to surface (Krause, 1997). The research base examining the best practice for improving the safe performance of lone workers is small. However, self-monitoring has been widely used in other fields as a behavior change technique.

Self-monitoring

Applications of self-monitoring procedures to improve organizational safety are scarce, but such procedures have been successfully applied to improve other types of organizational behavior. Wilk and Redmon (1990) used goal setting, self-monitoring, and feedback procedures to improve the performance of employees at a university admissions department. The admissions employees self-recorded the number of applications they processed each day and reported those data to their supervisors. One employee processed an average of 22 applications per day during baseline. This person's rate increased to an average of 180 processed applications per day when the intervention package was implemented. Supervisors in the Wilk and Redmon (1990) study assessed reliability of self-monitoring data and agreement between employee data and supervisor data ranged from 93.2% to 100%. Because the Wilk and Redmon intervention was a package of variables, it is uncertain to what extent self-monitoring was critical for generating the effects achieved.

Self-monitoring, as part of intervention packages, has also been used to improve customer servicc (Austin, Wellisley, & Olson, 1998; Troy, 1983), to improve academic performance (Dean, Malott, & Fulton, 1983; DiGangi, Maag, & Rutherford, 1991; Kneedler & Hallahan, 1981; Lan, 1996; Stecker, Whinnery, & Fuchs, 1996), to improve the performance of teachers (Browder, Liberty, Heller, & D'Huyvetters, 1986), to improve the performance of athletes (Kessler, 1985; Srikameswaren, 1992; Whelan, Mahoney, & Meyers, 1991), to increase interactions between staff and patients at an institution (Burgio, Whitman, & Reid, 1983), and to help individuals stop smoking and reduce their caloric intake (Moinat & Snortum, 1976).

It is clear that self-monitoring procedures have contributed to performance improvement across many settings. However, the question of which components of these processes are most critical for generating behavior change is still being explored. Some research suggests that self-monitoring procedures produce performance improvement even when the self-recorded data are not accurate or reliable (Austin, Wellisley, & Olson, 1998; Hayes & Nelson, 1983; McCann & Sulzer-Azaroff, 1996). These results suggest that the critical component of self-monitoring processes might simply be the frequency of exposures to antecedent stimuli that clarify the correct performance. However, when self-monitoring data are more reliable, its effects as a performance improvement tool seem to be enhanced (Baskett, 1985; Kanfer, 1970; McCann & Sulzer-Azaroff, 1996). Whether these enhanced effects occur because of more accurate self-estimations of performance or more accurate self-generated feedback is not known. It would be practical to discover whether BBS processes implemented with lone workers should include controls to assure that employees record accurate or reliable data on their performance. If there are not significant performance gains for assuring reliability or accuracy, then it would not be wise to invest time and money in that direction. With self-monitoring research in BBS being scarce, the field may require more demonstration studies using self-monitoring procedures to improve safe performance before questions about the importance of reliability can be addressed. This issue is illustrated by reviewing two applications of BBS self-monitoring procedures to improve safe performance.

BBS Applications of Self-Monitoring Procedures

Preventing Cumulative Trauma Disorders

McCann and Sulzer-Azaroff (1996) used a behavioral approach to prevent cumulative trauma disorders with employees who spent much of their day typing in an office setting. Part of the intervention package required typists to self-monitor their performance along particular behavioral dimensions. Participants were divided into two groups where one group worked on the target behavior of hand and wrist position and the other worked on the target behavior of posture. Within each group the intervention package was implemented in a staggered fashion as a multiple baseline design across participants. Each participant was exposed to conditions in the following sequence: (a) baseline, (b) training and self-monitoring, and (c) feedback, goal setting and reinforcement. During training participants were taught discriminations between safe and at-risk performance and were required to pass a discrimination test with a score above 80 percent correct. The self-monitoring procedures required participants to estimate the percent of time they performed their target behavior safely. During the final phase of intervention participants met prior to each session. At the meeting they were given both graphic and verbal feedback based on levels of safety observed by experimenters on the previous days. Experimenters guided participants as they set goals to ensure that goals were not set higher than the highest data point from the previous session. And finally, praise was provided for progress and attainment of goals.

The study produced consistent improvements in safe performance across all participants with moderate to high improvements during the training and self-monitoring phase, and very high improvements during the feedback, goal setting, and reinforcement phase. Posture ultimately improved to near perfect levels for all participants in the posture group. Hand and wrist position improved to levels clearly above baseline for all participants in the hand/wrist position group. Average percent improvement figures were not reported. Results were discussed only in terms of the visual appearance of the plotted data.

Experimenters collected data in pairs on the same behavioral dimensions as the typists and achieved acceptable inter-observer agreement for both posture and hand and wrist position (>80%). Participants were not initially given information about the accuracy of their self-estimations of safe performance. Without accuracy information participants achieved acceptable levels of inter-observer agreement between self-monitored data for posture and experimenter data for posture. However, self-monitoring data for hand and wrist position did not agree with experimenter data at this stage. Researchers postulated that the "gross motor" nature of the movements involved with posture made the behavior easier to self-monitor than the "fine motor" hand and wrist position movements, which resulted in the different reliability levels between posture and hand/wrist position. The goal setting, feedback, and reinforcement phase increased the agreement between self-monitoring data and experimental data for hand and wrist position. The reinforcement component (verbal praise) was made contingent upon performance improvement and accurate self-estimations of performance. The researchers reported that high agreement between typists and experimenters was associated with enhanced performance improvement of safe hand and wrist position.

Improving the Safe Performance of Bus Operators

Another application of self-monitoring procedures to improve organizational safety was reported by Krause (1997). This application was a Behavioral Science Technology, Inc. (BST) consultation effort with a public transportation system and represents the only published commercial application of a BBS process with lone workers. Thirty drivers and several supervisors participated in the project. Interviews with drivers were used to develop a checklist that contained 34 specific performances. Drivers estimated their safe performance on these 34 items once or twice daily and plotted their self-monitoring data on graphs. Every two weeks a supervisor rode with each driver and collected data using the same checklist.

When the self-monitoring procedure was initially implemented, drivers reported high percent safe scores that did not agree with supervisors' scores of driver performance. In fact, the two groups were discrepant by over 40 percentage points. Supervisors discussed these discrepancies with drivers and plotted the self-monitoring data and supervisor data together on feedback graphs. Over a period of 20 weeks, supervisor data trended upward and driver data began to trend downward slightly to almost match supervisor data. Agreement between employees and supervisors appeared to take place over time, but no formal assessment of inter-observer agreement was made. Krause (1997) reported a substantial 66 percent decline in injuries and accidents in the organization over the 20 week time period, but the degree to which the drivers' behavior actually changed could not be determined from the data. The decline in injuries was probably due to changes in behavior, but may have been coincidental.

In order to experimentally evaluate the degree to which self-monitoring procedures can improve safe performance of lone workers, a demonstration study similar in design to the McCann and Sulzer-Azaroff (1996) study is needed. BST (as cited in Krause, 1997) did not employ an experimental design and did not collect reliability on supervisor or driver data. The current study is an attempt to synthesize aspects of Krause (1997) and McCann and Sulzer-Azaroff (1996) and experimentally demonstrate the effectiveness of self-monitoring procedures for improving the safe performance of bus operators.

Possible Behavioral Functions of Self-monitoring

There are several behavioral mechanisms that could be responsible for the effectiveness of self-monitoring processes. There are also behavioral concepts that may help explain the effects of self-monitoring processes. Some of the potential behavioral functions of self-monitoring include: (a) an antecedent function (i.e., task clarification), (b) a consequence function (conditioned reinforcement or punishment), (c) a rule generating function (i.e., contingency specifying/function-altering stimuli), and (d) a conditioned establishing operation function.

When a participant is asked to record aspects of his or her behavior, looking at the form and filling it out may clarify performance expectations. Viewing self-monitoring in this way would mean that its effectiveness would be determined by the degree to which filling out the self-monitoring form correctly explained the behaviors that were expected. If self-monitoring functions primarily in this way it would make sense to ask participants to self-monitor at the beginning of the work day. It would also make sense to stress reinforcement for participation, in order to correlate positive consequences with the antecedent process. However, even if participation is reinforced, participants could habituate to self-monitoring antecedents. Therefore, it might be wise to change the behaviors being self-monitored occasionally to avoid deterioration due to habituation.

Aspects of self-monitoring processes may also function as consequences. The consequence function may depend upon the value of making marks on the self-monitoring form (scoring yourself high or low on the monitored performance). If making marks was reinforcing, the sight of the form could function as a discriminative stimulus for those immediately available conditioned reinforcers. If making marks on the form was punishing, the sight of the form may serve as a warning stimulus that punishment was forthcoming. Viewing self-monitoring in this way would suggest that filling out a self-monitoring form would be reinforcing if a participant had been performing well prior to the activity, or as a warning stimulus if the participant had not been performing well prior to the activity. The value of the marks made on self-monitoring forms may also be accentuated by the consequences that supervisors pair with such procedures. This might explain why compliance with self-monitoring processes is not always perfect. Because performance varies, scoring aspects of one's own performance would sometimes be reinforcing and sometimes be aversive.

Due to the fact that management systems utilize numerous performance management strategies, filling out self-monitoring forms may cause participants to generate rules related to other safety outcomes in the organization. If an organization uses a great deal of aversive consequences to discourage unsafe practices and punish employees, filling out a self-monitoring form might promote the generation of rules such as, "If I work really hard and improve this performance, I can avoid punishment from my supervisor (because these performances on this form are what he/she cares about right now)." Schlinger (1993) proposed that a rule such as this one might have behavioral effects because it specifies contingencies and alters the function of stimuli in the immediate environment. The rule described above could change behavior because it specified new contingencies in effect (my supervisor will punish me if I don't improve these behaviors on the form), and alter the function of other stimuli (sight of a stop sign evokes behavior that results in a complete stop).

One more concept that may help explain the effects of self-monitoring procedures is the conditioned establishing operation (CEO). An establishing operation is a procedure that has at least two effects; it (1) alters the value of consequences, and (2) momentarily increases the frequency of behavior that has been correlated with the consequences whose value has been altered (Michael, 1993). A CEO is a procedure with the above effects that functions because of an individual's learning history. For a discussion concerning specific types of CEO's see Michael

(1993). A self-monitoring procedure requiring a participant to record aspects of his or her safe behavior might alter the value of numerous safety-related outcomes. For example, a bus driver may perform rolling stops at stop signs because the brakes squeal less than when he/she performs a complete stop. Participating in a self-monitoring procedure that targeted complete stop might alter the value of this squealing sound, making it less aversive. It might be the case that the squealing sound could become a positive consequence, signaling the successful performance of the behavior being self-monitored, thereby evoking behavior (firm foot pressure on the brakes) that produced that consequence.

The great majority of self-monitoring procedures require verbal skills, therefore, it is likely the case that performance improvement generated by those procedures is caused by a complex set of contingencies and behavioral mechanisms. Considering these mechanisms and explanatory concepts may guide future research and help discover the most effective practices.

The Relevance of Reliability

Although books generated by leaders in the BBS field stress the importance of reliable/accurate behavioral measures during the observation process (Daniels, 1989; Geller, 1988, 1996; Krause, 1997; McSween, 1995), it is not clear that assessment of the reliability of behavioral data regularly takes place at commercial BBS implementation sites.

When proponents of BBS stress the importance of reliable measures in safety improvement efforts, they are most likely concerned with the accuracy of observations. Accuracy is the degree to which data from observations represent the actual state of affairs in nature. To measure accuracy one would need a comparison between data collected by an observer and a perfect measure of the same event as recorded by a machine or expert (Johnston & Pennypacker, 1993). Calculating inter-observer-agreement (IOA) is one way to estimate accuracy. It is simply the degree to which two independent data collectors agree when they measure the same natural event (# agreements divided by # agreements + disagreements, and then multiplying by 100). Assessing IOA is the professionally accepted practice for estimating the accuracy of observations in the field of Applied Behavior Analysis.

Assessment of IOA is rarely reported in BBS implementations in the industrial/organizational community. It is likely the case that even fewer organizations actually adopt some form of reliability assessment system after consultants leave. Demands on employee time in organizations limit IOA procedures, where two observers must collect data at the same time. Without the assessment of IOA it is impossible to know whether reliable behavioral data collection is a crucial variable for creating performance improvement in safety. Questions concerning the importance of reliable data in BBS are especially poignant when participants work alone. Lone workers who self-monitor may over or under estimate their performance to avoid aversive consequences or obtain rewards. Therefore, improvements in safe-monitoring data may not reflect real changes in safe behavior. In addition, reliability is difficult and costly to assess with lone workers. Assessing reliability may require videotaping performance, creative supervisor/employee data comparisons, or increased visits from supervisors or pairs of supervisors. Each strategy for assessing reliability with lone workers is costly. These problems increase the chances that behavioral data in BBS implementations with lone workers will be unreliable or inaccurate.

It is intuitive to argue for reliability of behavioral data in BBS applications. Feedback based on the true or accurate rate of behavior should be more effective at generating performance improvement than feedback based on less accurate data. The attention to reliability in books about BBS suggest that leaders in the field believe that exemplary safety improvement cannot be achieved without reliable measures of behavior (Daniels, 1989; Geller, 1988, 1996; Krause, 1997; McSween, 1995). However, it may be the case that reliable measures of behavior are not as important as the mere presence of observers and/or exposure to checklists that describe desired safe performances. The additional improvement that reliable data might generate may not be valuable relative to the labor costs involved in keeping data accurate.

The lack of IOA measures in most BBS commercial implementations, the increased likelihood of over/under reporting when employees self-monitor, and the possible behavior changing power of accurate data in BBS make IOA assessment an important issue when applying BBS processes with lone workers. Therefore, the reliability of behavioral data in the current study is an important aspect of the experiment.

METHOD

Participants and Setting

A public transportation system that served two midwestern cities with a combined estimated population of 160,000 was the participating organization in the study (Kalamazoo County Visitors Center, personal communication, March, 1999). The organization operated and maintained 17 different bus routes. An operations supervisor managed the performance of seven dispatch supervisors, who in turn supervised 65 bus and other vehicle operators. The operations supervisor and the director of the transit system were interested in using a BBS process with all 65 drivers, but wanted to pilot such a process with one shift of drivers to examine the feasibility and benefits of such an effort. A university campus route was chosen as the location for the pilot study. The university bus route served a campus of approximately 26,000 students. Between two and eight busses operated on the route at different times each weekday.

Four full time drivers participated (male, ages 40-50) and were selected by the operations supervisor. These drivers were some of the most experienced drivers in the organization and included the local union president. Their average bus driving experience was 20.5 years (range: 19-23). The participants worked a 10-hour shift that ran from 7am until 4:30pm, which was the longest shift available in the transit system. Transit management was concerned about safety on this shift because of its duration and because of the busy pedestrian and traffic conditions typical to the university campus. They were also interested in obtaining feedback about BBS from this group of very experienced drivers.

Each participant worked four days a week and was responsible for a specific run number on the campus route. The run number specified the schedule and locations on the route that the driver was responsible for. The bus route was about 30 minutes in length and served all major parts of the campus including all on-campus housing. Busses ran both clockwise and counterclockwise from 7am-12midnight on weekdays only.

Extant Safety Management Strategies

Prior to the study, the transit system used five methods to motivate safe driving. These methods were (1) financial incentives, (2) an escalating discipline program, (3) hiring private investigators to monitor driving, (4) yearly safety awards at a banquet, and (5) bi-monthly performance evaluations by dispatch supervisors.

1. Cash Incentives. The incentive was a $25 bonus for all drivers who worked an entire quarter without having a preventable collision (collisions were labeled preventable and non-preventable by the operations supervisor after an investigation of a collision).

2. Discipline. The discipline program provided 7 escalating aversive consequences for drivers involved in moving violations and/or preventable collisions. The first consequence for a violation was a verbal warning and the final consequence

was termination. Drivers received improved status in the discipline program after a period of collision/moving violation free performance.

3. Private Investigation. Private investigators were hired if the transit system received enough complaints of reckless driving about a particular driver from passengers to warrant suspicion of frequent moving violations. Drivers were disciplined for violations observed by private investigators.

4. Safety Awards. At an annual banquet for the entire organization a yearly safety award was given to all drivers who have gone the whole year without a preventable collision. The award usually consisted of some kind of prize and plaque.

5. Supervisor observations. The final method for motivating safe performance involved supervisor observations of driver performance. Once every two months supervisors were required to ride with drivers or follow drivers in another vehicle, and evaluate their performance with a 32-item checklist. Items on the checklist received either an acceptable rating, a "needs work" rating, or an unacceptable rating. Drivers received verbal feedback on their performance at the discretion of the supervisor and the written evaluation was placed in the driver's employment records.

As a management system these methods achieved a preventable collision rate of 2.08 preventable collisions per month for 1997. The average yearly collision rate (preventable and non-preventable collisions) for the organization had reached a plateau of about 40 collisions per year. The operations supervisor reported that total collisions per year had averaged about 40 for the past five years.

Dependent Variables

Dependent variables were selected through a combination of several review processes. Potential dependent variables were identified by reviewing the targets used in the BST application (Krause, 1997), reviewing the transit system's current performance evaluation checklist, reviewing one year of collision reports from the organization's records and identifying behaviors that could have prevented those collisions, and consulting performance evaluation checklists from other transit systems. Multiple interviews with the operations supervisor and pilot data collection narrowed down the potential dependent variables that could be successfully measured during the study. The degree to which acceptable reliability could be achieved when observations were made from inside of a bus as it traveled was the final consideration for the selection of dependent variables. Six performances were observed throughout the study. They were divided into three categories: (1) loading/unloading passengers, (2) bus in motion, and (3) stopping. The following paragraphs provide the definition of each dependent variable and the rationale for choosing each of them.

Three dependent variables belonged to the "loading/unloading passengers" category. The review of collisions for the transit system discovered that 20 percent of preventable collisions had occurred at loading zones. Another 12 percent of preventable collisions occurred at parking lots, or driveways. The campus bus route drove through six major campus parking lots every 30 minutes and passed numerous exits and entrances to parking lots on campus. In addition, many loading zones on the campus were located near these parking lots and driveways. The dependent variables chosen to prevent collisions as passengers loaded/unloaded from busses were defined as follows:

1. Bus stopping position. Bus doors must remain shut until the bus is completely stopped, and the bus should be positioned so no cars can pass on the right. This definition was taken directly from the transit system's policy. Observers scored this performance by watching the front doors of the bus as it slowed. If the bus was still moving when the doors separated, the performance was scored at-risk. Observers were stationed on the right hand side of the bus and estimated whether or not a car could pass on the right.

2. Two seconds motionless. The bus should remain motionless for at least two seconds after the last loading/unloading passenger either steps behind the yellow line, steps off the bus to the right, or steps clear of the front left corner of the bus. Transit system policy was for passengers to be completely seated before the bus moved. However, this policy did not take into account the safety of passengers outside of the bus. The two seconds motionless criteria addressed this problem. To observe this performance observers started counting to themselves when the last passenger stepped into one of the three specified locations. Observers were instructed to count "one-thousand one, one-thousand two," to themselves, and use a wrist watch during the bus ride to calibrate the pace of their counting. Any movement of the bus before the observer reached "two" was scored as at-risk. If the observer was able to count two seconds before the bus moved, the performance was scored correct.

3. Mirror check. The driver should visually check both side mirrors after loading/unloading passengers as the bus pulls out of a loading zone. The behavior of checking mirrors was identified as a behavior that could have been critical in the prevention of 56 percent of the collisions reviewed for the year 1997. The most common collisions in 1997 involved busses striking an object with its mirror or bumper as the bus pulled out of a driveway, loading zone, or parking area. On a campus route with many pedestrians at curbside, this behavior was a clear choice for inclusion. It was also transit system policy to check the left side mirror before merging with traffic, and check the right side mirror after unloading a passenger. The final version of the dependent variable simply combined these two transit system policies. Observers were instructed to mark this performance as correct if both mirrors were checked before or as the driver started moving. Checking mirrors after the back of the bus cleared the original load/unload location was scored at-risk. From the driver's right hand side of the bus in the second row of forward facing seats, the driver's eyes were visible in the center mirror and head movement could be viewed. If a driver looked in the general direction of either mirror it was assumed he checked that mirror. During pilot data collection sessions, acceptable IOA was not achieved with the requirement of checking mirrors before moving, as the behavior happened almost simultaneously with slight movement of the bus. Changing the requirement to checking before or as the bus pulled out from loading zones solved this problem.

Two dependent variables fell into the "bus in motion" category. These dependent variables were chosen to prevent collisions as the bus traveled and were defined as follows:

4. Cornering (all 90-degree turns). The driver should brake before, not during the corner, and maintain at least 2 feet clearance between the side of the bus, cars, poles, signs, and people. Also, the driver should not hold the change box, change box railing, window frame, or food during a corner. Only objects clearly visible from outside the bus windows were included in the 2 feet clearance criteria. If a bus came very close to a car but actual distance could not be seen from the observers' position on the bus, observers were told to disregard that event. Cornering was scored at-risk if heavy or jerky braking occurred during a corner (unless the bus was avoiding another car that had run a stop sign or passed illegally); if there was less than 2 feet clearance between side of bus and cars, poles, signs, and/or people; if the bus bumped a curb or scraped the pavement; if the driver held the change box, change box railing, or window frame during a corner; or if the driver held food during a corner. Of the collisions reviewed for 1997, 20 percent occurred at corners and/or intersections.

5. Following distance (two-second distance). The bus must maintain a distance behind the leading car of at least two seconds. Following distance was included to promote scanning ahead as the bus traveled on a straightaway. Common collisions on straightaways involved scraping other vehicles while passing, striking tree branches hanging in front of the bus, and striking objects on the curb. Following distance was observed two seconds after every corner and again eight seconds after

that first observation. When a bus straightened out after a corner, observers counted two seconds and then looked up to observe and count following distance using a landmark to judge when to start and stop counting. After the first observation was scored, observers used a wristwatch to count off eight more seconds, and then looked up and observed following distance again. This method was used to increase the frequency of following distance observations.

Only one dependent variable was in the "stopping" category. Coming to a complete stop is a legal requirement and was considered an important safe performance for the campus route. Drivers making complete stops have a better opportunity to scan traffic and pedestrian conditions at busy intersections. There were over 20 stop signals during each 30 minute loop on the route, regardless of the direction the bus was traveling. When supervisors evaluated driver performance, a complete stop was the criteria for acceptable performance. Complete stop was defined as follows:

6. Complete termination of motion at stop signs and red traffic lights. Rolling stops and jumping a traffic signal were scored as at-risk. If the light was green, an observation was not be made. During pilot observations, the method used that achieved reliability included picking out an object like a pole, and watching it as the bus slowed. If that object completely stood still, a complete stop was scored correct.

If a driver's performance on a dependent variable met all of the criteria of a definition, the dependent variable was scored as correct. If the driver's performance failed any part of the safe definition, it was scored at-risk. A percent safe score for each dependent measure was calculated by counting the number of correct scores and dividing that number by the total number of observations for that dependent measure, and then multiplying by 100. An overall percent safe score for each observation session was also calculated by counting the total number of correct scores for all dependent measures and dividing that number by the overall total number of observations during the session, and then multiplying by 100.

Observation Procedures

Observers and Confidentiality for Participants

Three observers collected data on the four drivers throughout the study. Observations were made from the driver's right hand side of the bus from the second row of forward facing seats. This location was about 10 feet from the driver's chair. Each individual driver was assigned a specific run time on the University route by the transit system. To ensure that individual drivers could not be identified from the data shown to the transit system, these run times were used to create a color code for each driver being observed; (a) yellow, (b) blue, (c) purple, and (d) green. The three observers met weekly to discuss observation assignments, solve problems, and plan a schedule for the upcoming week. Observation assignments were made so that each driver color would be observed each day. Each observer used a bus schedule marked with four colors to identify times and locations on the route where a specific driver (according to their color code and run time) could be observed. Observers learned to recognize drivers but did not know their names during any part of the study.

Frequency of Observations and Session Length

All four run times were observed once each day. The color code for the driver observed was circled on the data sheet, as well as the date, time and location when the observer boarded and exited the bus, and the name of the observer assigned to that run time that day. Each observation session lasted at least 30 minutes and constituted one complete trip around the entire route. A session sometimes lasted longer than 30 minutes as observers were required to observe at least 10 instances of loading/unloading of passengers during a session. If a session lasted for one complete trip around the route, there were 10 or more load/unload of passengers observed, about 30 corners observed, between 6 and 10 following distances observed, and over 20 stops observed. Once or twice each week all four run times were observed by two observers, and IOA was calculated by dividing the number of agreements by the number of agreements + disagreements, and then multiplying by 100.

Reliability

During reliability sessions, the author was the primary observer. To ensure that both observers were attending to the same events at the same time, the primary observer had the responsibility to announce upcoming opportunities to observe corners and stops. The primary observer also announced the times to observe following distance. Observations were made when the primary observer said "following, now." To protect the independence of observations, the observer sitting on the right hand seat next to the window used a three-ring binder with the left cover held upright to block the visibility of the data sheet. The observer sitting on the left hand seat covered his/her data sheet with his/her right arm and hands (all observers were right handed).

Methods to Minimize Driver Reactivity to Experimental Observers

Participating drivers were not aware of the purpose of experimental observers until a debriefing meeting at the conclusion of the study. This was hoped to have diminished the impact of driver reactivity to the presence of experimental observers during baseline and intervention conditions. It was odd for passengers to ride the entire route without ever arriving at a destination, and drivers occasionally asked questions. If observers were asked questions about what they were doing by drivers, they were instructed to say, "I'm collecting a survey on passengers for a class." To minimize the possibility of such interaction with drivers, observers were instructed to wear headphones when collecting data by themselves.

Independent Variables

Training

Intervention began when all four drivers attended a training session together at the transit system hub. The training lasted for one and a half hours and consisted of three basic components, (1) introduction to BBS and rationale for piloting such a process at the transit system, (2) introduction to and rationale for the self-monitoring process, and (3) description of the details of running the project (i.e., completing self-monitoring forms, handing in self-monitoring forms, the nature of feedback, distribution of color codes, weekly visits by dispatch supervisors, etc.). The training was conducted by a doctoral student consultant (male, age 26) not involved in data collection and by the operations supervisor. The student consultant was introduced as an external safety consultant without mentioning his ties to the university. The student consultant had previous experience implementing BBS processes at two large paper mills and at a paper products manufacturing plant. This experience was shared with the drivers.

To reward drivers for participating fully in the process, they were informed during the initial training session that lunch would be provided for them at the conclusion of the pilot project. Drivers were told that one or two additional brief meetings with the consultant would be scheduled over the next few weeks. Drivers were also informed that the transit system would not be given any information that could identify individuals from the data and that data that were collected from the process could never be used to punish drivers for any reason (transit officials signed a site approval form stating these conditions).

Immediately after the training session with drivers, the student consultant and the operations supervisor met with dispatch supervisors at their weekly meeting. They were informed about the process and their duties for the duration of the pilot project, which included prompting drivers to use self-monitoring forms twice each day and observing driver performance once each week with new checklists (observing occurrences of the performances being self-monitored by drivers). Supervisors were not informed of the presence of experimental observers.

Self Monitoring

Three different self-monitoring forms were used during the three phases of intervention (see experimental design section). Drivers used these forms to record their safe performance twice each day during their 10-hour shift. They were asked to estimate the percent of time they performed each of the target performances safely. Blank squares were provided on the form for drivers to write their estimations in. This format is recommended by researchers who study self-report measures in order to avoid shaping respondents' answers (Schwarz, 1999). Self-monitoring data sheets were handed in at the transit hub each day at a locked drop box in the drivers' lounge at the transit system hub. The participants chose the location of the drop box. Drivers were told that they would be prompted twice a day by their dispatch supervisors via CB radio when it was time to self-monitor.

The first self-monitoring form included only the "complete stop" performance. Drivers used this form for eight workdays. At the beginning of the second phase of intervention, the operations supervisor and the consultant (who conducted the initial training) met the drivers at the transit system hub for a brief meeting and introduced a new self-monitoring form. The new form required drivers to continue self-monitoring complete stop performance and begin self-monitoring the performance of remaining two seconds motionless after loading/unloading passengers. This second checklist was used for five workdays. To begin the third and final phase of intervention, another brief meeting with drivers at the transit hub was arranged. The third and final self-monitoring form included complete stop, and all three of the passenger load/unload performances (two seconds motionless, mirror check, and bus stopping position). After five more working days using this final form, the route stopped running for the semester and the study was concluded. The "bus in motion" performances of cornering and following distance were never introduced to drivers because there was little room for improvement. The drivers averaged over 90 percent safe for these two behaviors throughout the study.

Feedback

The author generated a daily graph of individual and group performance based on self-monitoring data and posted it on the wall in the drivers' lounge where data sheets were handed in. Each driver was asked to initial the group graph each day in a box labeled with their run number to demonstrate that they viewed the graph. Color codes were used on all graphs so individuals could not be identified from the data.

Participants did not self-plot their graph like the drivers in Krause (1997) because of the risk of revealing their individual graph to people watching them in the drivers' lounge when they signed their initials.

Supervisor Prompts and Observations

Dispatch supervisors were instructed to prompt participating drivers via CB radio twice each day to use the self-monitoring forms. Supervisors were asked to record the date and time of their prompt on a chart posted in the dispatch office. These prompts were planned to occur at 11am and 3pm. In addition to delivering prompts, a supervisor rode once with each driver during the study to observe performance on the dependent variables that the driver was currently self-monitoring. Experimental observers arranged to measure performance concurrently with supervisor observations. Experimental observers boarded the bus prior to the supervisor visit and left the bus one or two stops after the supervisor left the bus. This procedure was added to the design of the study as a type of probe, where performance changes generated by the presence of a supervisor could be measured and compared to data collected on the same day without supervisor presence. To create this comparison, each driver was observed for an additional session on the same day either before or after the supervisor probe was completed.

Independent Variable Integrity

Three measures of independent variable integrity were calculated. The first of these was percent compliance with the self-monitoring procedure for each participant. This was calculated by counting the actual number of self-monitoring forms completed by each driver, dividing that number by the expected number of completed self-monitoring forms for each driver (two per day), and then multiplying by 100.

The second measure of independent variable integrity was percent compliance with feedback procedures. This was calculated by counting the number of days each driver signed the feedback graph, dividing that figure by the number of days the driver was expected to sign the feedback form, and then multiplying by 100. The third measure was the percent of supervisor compliance with delivering prompts. This was calculated by counting the number of prompts recorded on the supervisor form, dividing that figure by the number of prompts that were expected to be given, and then multiplying by 100.

Experimental Design

A multiple baseline design across performances was used to assess the effects of the intervention. Intervention began after a baseline of 10 sessions for each driver for each of the six dependent measures was obtained. During phase one of the intervention drivers self-monitored complete stop performance while baseline conditions continued for all five of the remaining dependent variables. This first intervention condition lasted for eight workdays. Phase two introduced one new performance to the drivers as they self-monitored complete stop and the performance of remaining motionless for two seconds after loading/unloading passengers. This second condition lasted for five workdays while baseline conditions continued for the remaining four dependent variables. The third and final phase of intervention introduced two new performances to drivers (four performances total) and they self-monitored complete stop, two seconds motionless, checking mirrors, and bus stopping position. The final phase also lasted for five workdays. As mentioned previously in the self-monitoring section of this paper, the “bus in motion” performances were never introduced to drivers because little opportunity for improvement existed. A supervisor observation probe was arranged for each driver each week. However, absences and competing demands on supervisor time only allowed the successful completion of one supervisor probe session for each driver over the course of the study.

RESULTS

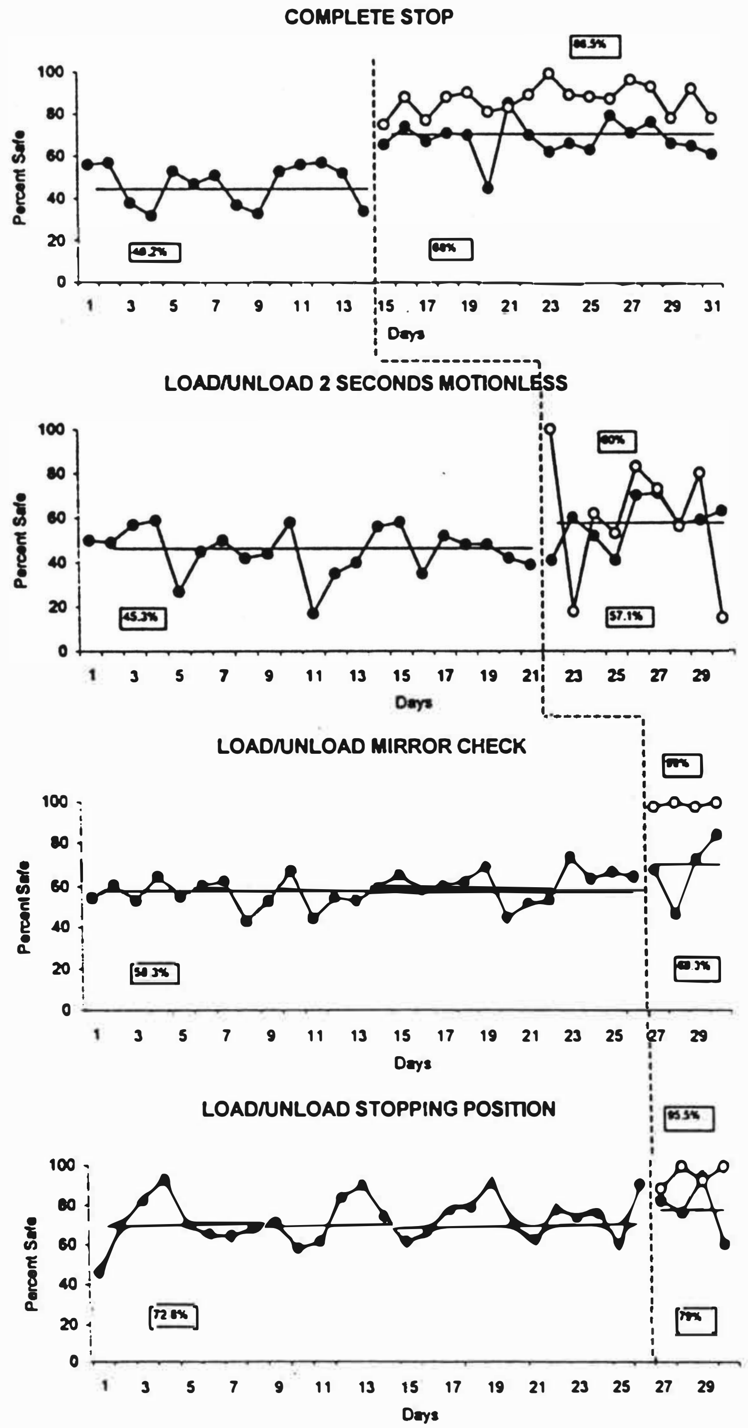
Group Performance

The group of four participants improved their safe driving by an average of 12.5% over baseline conditions. The dependent variable realizing the largest improvement for the group was complete stop which improved by an average of 21.8% (range: 14%-41%). Two seconds motionless after loading/unloading passengers improved by an average of 11.8% (range: 3%-19%), mirror check improved by an average of 10% (range: 3%-15%), and bus stopping position improved by an average of 6.2% (range: 2%-12%). The results obtained for each driver during supervisor probes is contained in the following section on individual performance. Figure 1 represents the grouped data (i.e, averaged across the four drivers for each session) for each of the four dependent variables in a multiple baseline design across performances.

Individual Performance

The results of individual participants are presented in order of the participant with the greatest average overall improvement to the participant with the least average overall improvement.

The yellow participant improved by an average of 14% over baseline levels. He realized his greatest average improvement for the behavior of two seconds

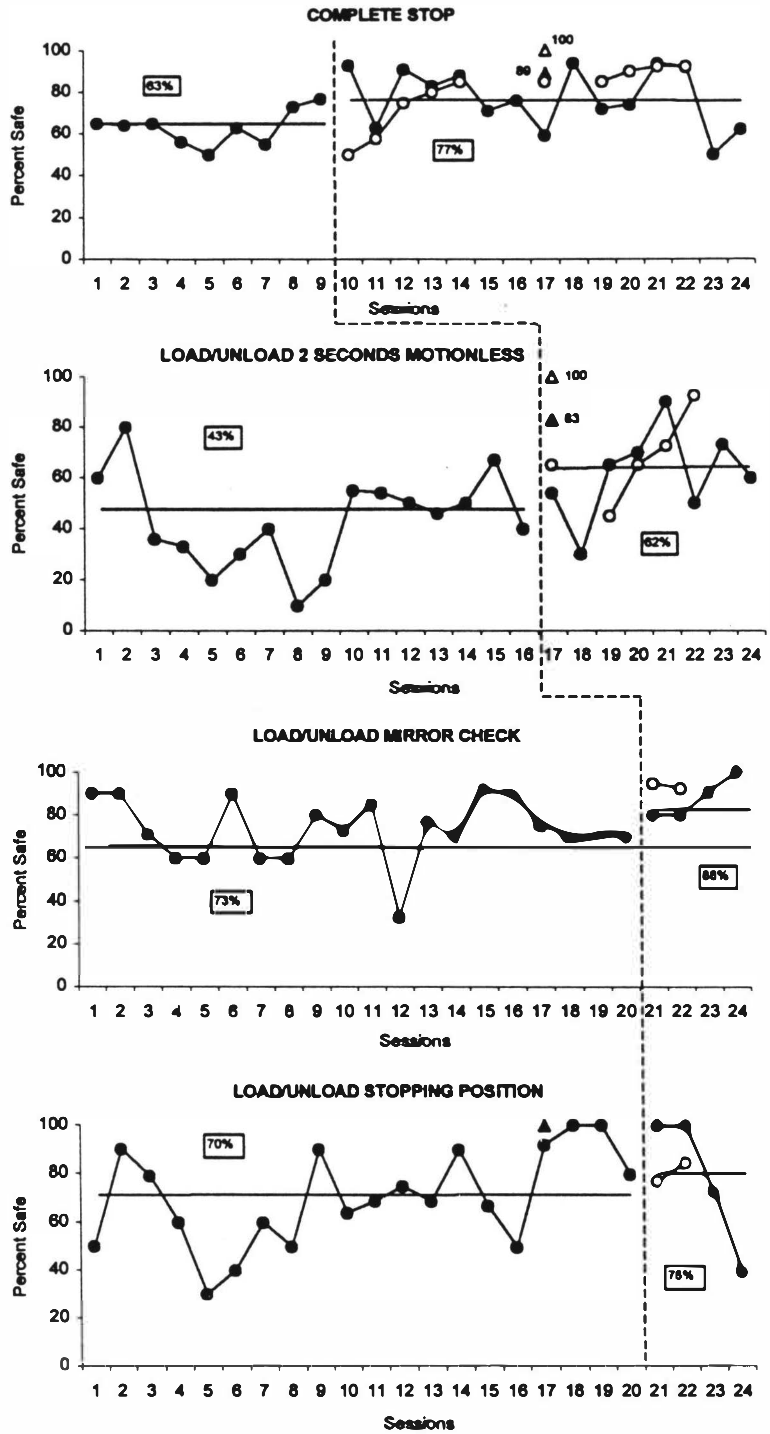


Legend. Closed circle data points are experimenter data averaged for each session and open circle data points are self-monitoring data averaged for each session.

Figure 1. Group Results in Multiple Baseline Design Format.

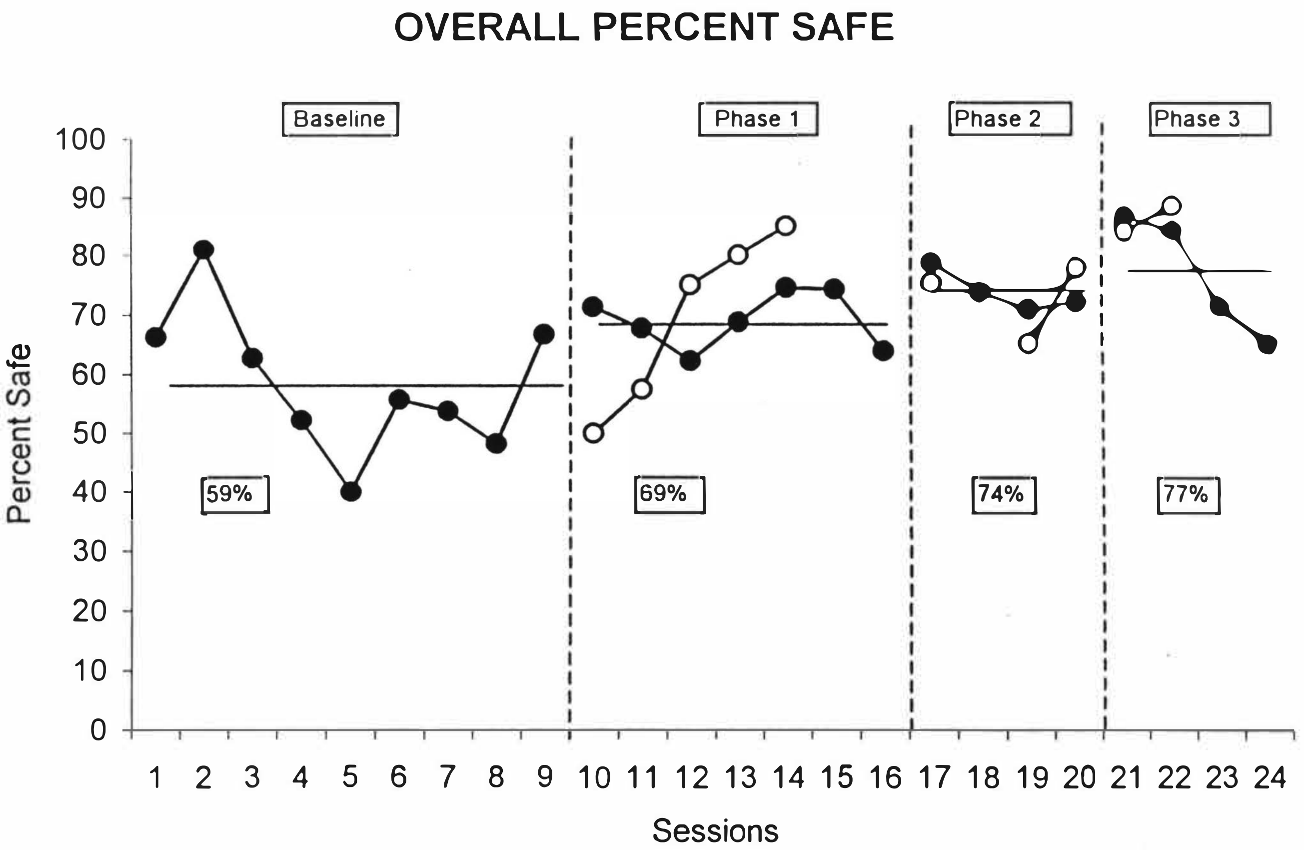
motionless with a 19% improvement (baseline, 43% safe; intervention, 62% safe). Bus stopping position improved 8% (baseline, 70% safe; intervention, 78% safe), mirror check improved 15% (baseline, 73% safe; intervention, 88% safe), and stopping improved 14% (baseline, 63% safe; intervention, 77% safe). The yellow participant achieved the greatest overall improvement of all drivers with an average improvement of 14%. A supervisor probe on the first day of phase two of the intervention created systematic effects on the performance of the yellow participant. Complete stop and two seconds motionless, which were being self-monitored, improved to over 20% above the levels measured on the same day without supervisor presence. Mirror check and bus stopping position, which were still under baseline conditions, did not change in the presence of the supervisor. For a graphic display of these data see Figures 2 and 3 on pages 35 and 36.

The green participant improved by an average of 11% over baseline conditions. He realized his greatest average improvement for the behavior of complete stop with a 41% improvement (baseline, 51% safe; intervention, 92% safe). This improvement stands out as the most clear and dramatic effect of the intervention procedures. Bus stopping position improved 3% (baseline, 49% safe; intervention, 52% safe), two seconds motionless improved 3% (baseline, 28% safe; intervention, 31% safe), and mirror check also improved 3% (baseline, 38% safe; intervention, 41% safe). For a graphic display of these data see Figures 4 and 5 on pages 37 and 38.



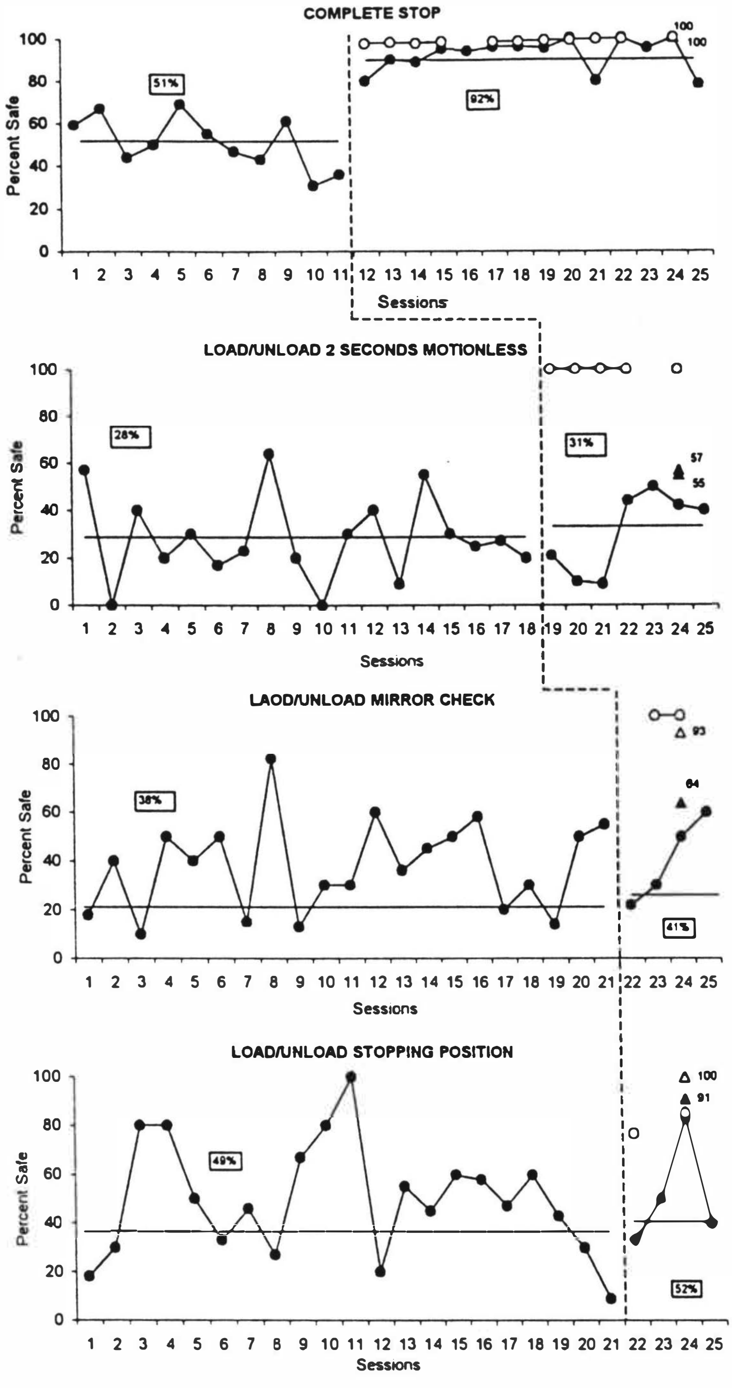
Legend. Closed circle data points are experimenter data, open circle data points are self-monitoring data, closed triangles are experimenter data during supervisor probes, and open triangles are supervisor data.

Figure 2. Yellow Participant Results in Multiple Baseline Design Format.



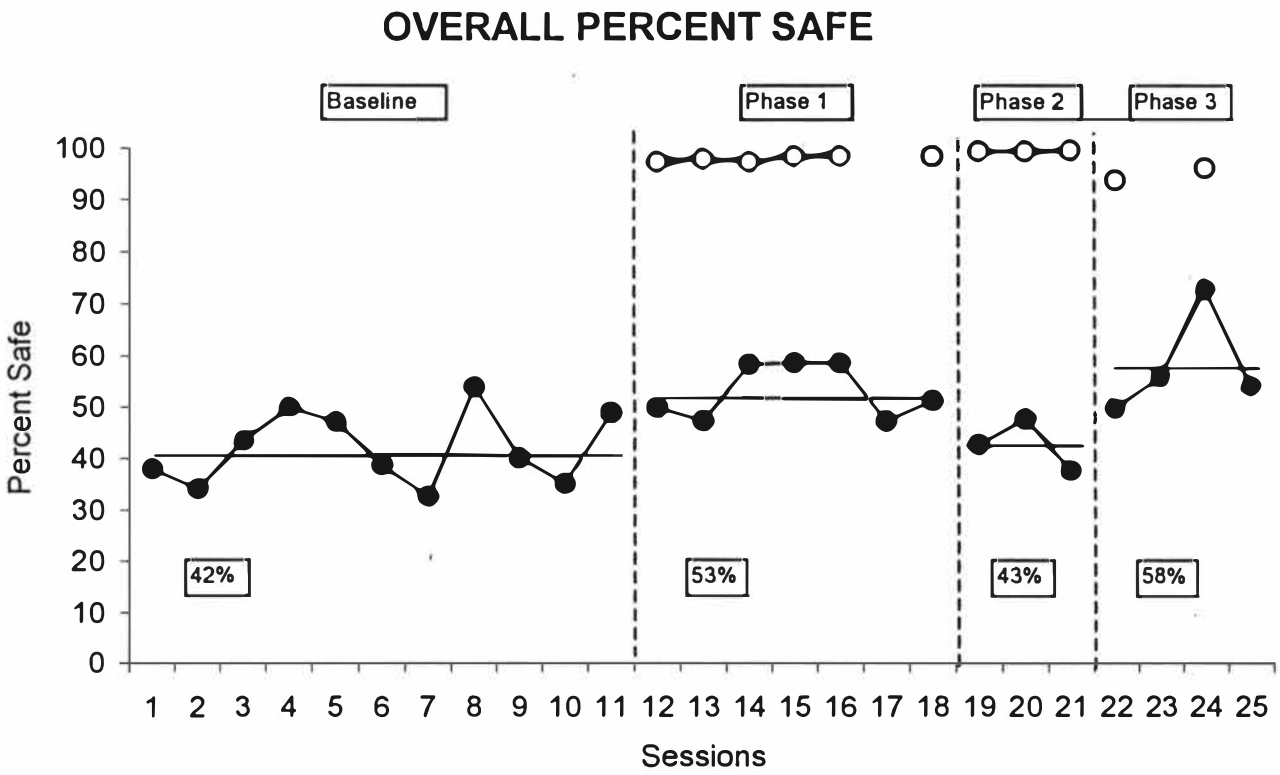
Legend. Closed circle data points are experimenter data, open circle data points are self-monitoring data, and vertical broken lines represent intervention phase changes.

Figure 3. Yellow Participant Overall Percent Safe Scores Across Phases.



Legend. Closed circle data points are experimenter data, open circle data points are self-monitoring data, closed triangles are experimenter data during supervisor probes, and open triangles are supervisor data.

Figure 4. Green Participant Results in Multiple Baseline Design Format.



Legend. Closed circle data points are experimenter data, open circle data points are self-monitoring data, and vertical broken lines represent intervention phase changes.

Figure 5. Green Participant Overall Percent Safe Scores Across Phases.

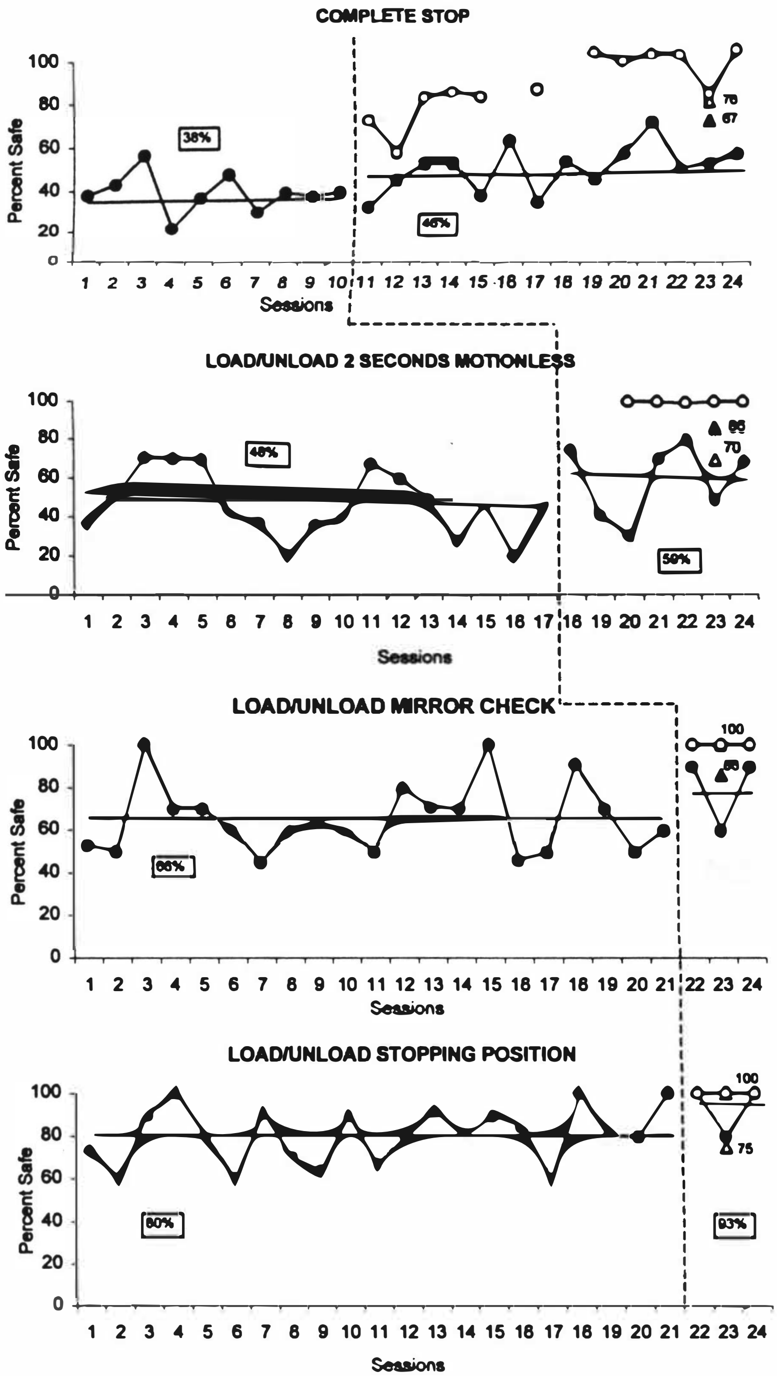
The purple participant improved by an average of 9% over baseline conditions. He realized his greatest average improvement for the behavior mirror check with a 15% improvement (baseline, 65% safe; intervention, 80% safe). Bus stopping position improved 12% (baseline, 81% safe; intervention, 93% safe), two seconds motionless improved 12% (baseline, 47% safe; intervention, 59% safe), and complete stop improved 9% (baseline, 38% safe; intervention, 47% safe). For a graphic display of theses data see Figures 6 and 7 on pages 40 and 41.

The blue participant improved by an average of 8% over baseline conditions. The blue participant realized his greatest average improvement for the behavior of complete stop with a 19% improvement (baseline, 38% safe; intervention, 57% safe). Bus stopping position improved 2% (baseline, 94% safe; intervention, 96% safe), two seconds motionless improved 5% (baseline, 66% safe; intervention, 71% safe), and mirror check improved 15% (baseline, 58% safe; intervention, 73% safe). For a graphic display of these data see Figures 8 and 9 on pages 43 and 44.

Results of Self-monitoring Estimations

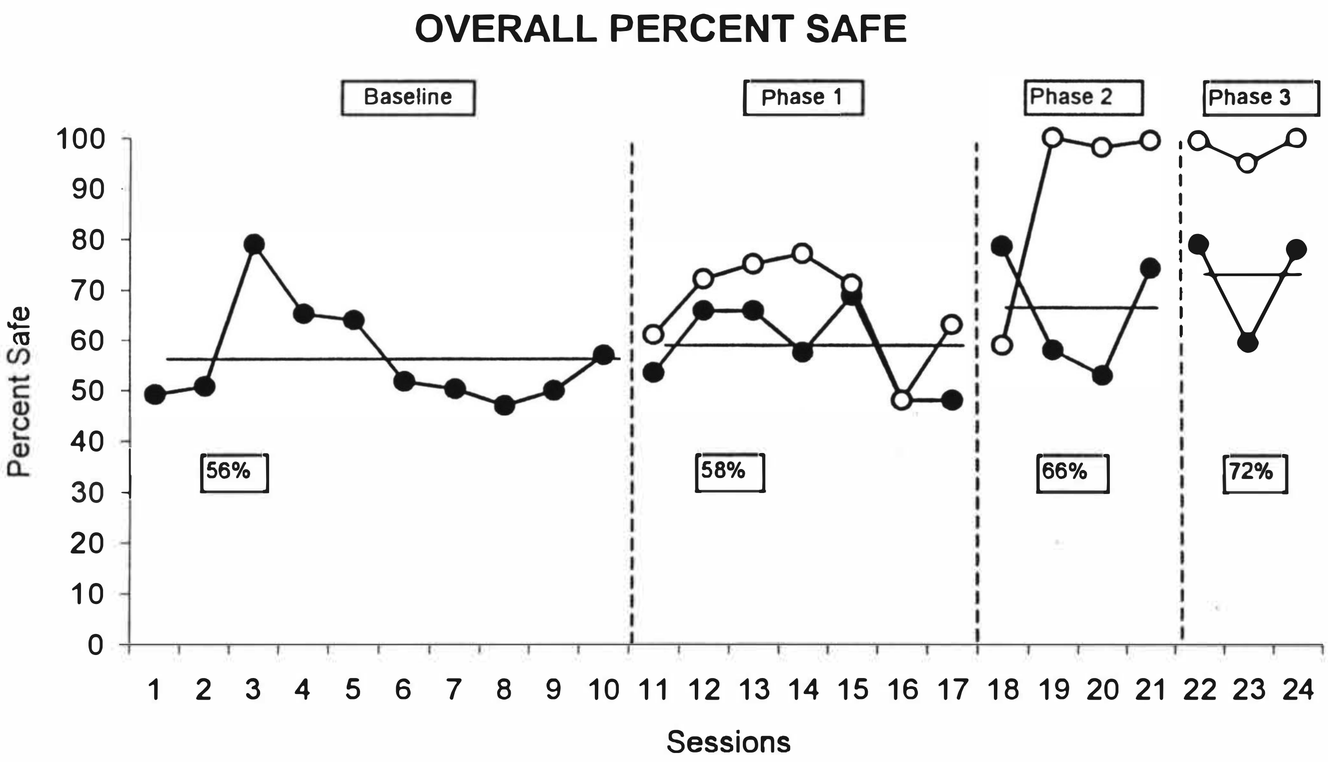
It should be noted that drivers estimated their performance for an entire day with two self-observations, and experimenters only sampled their behavior for between 30 minutes to 60 minutes each day. Therefore, the comparison between experimenter and self-monitoring data is not an exact comparison.

The average of the yellow participant's percent safe estimations across all intervention phases was 72% (complete stop, 79%; two seconds motionless, 67%;



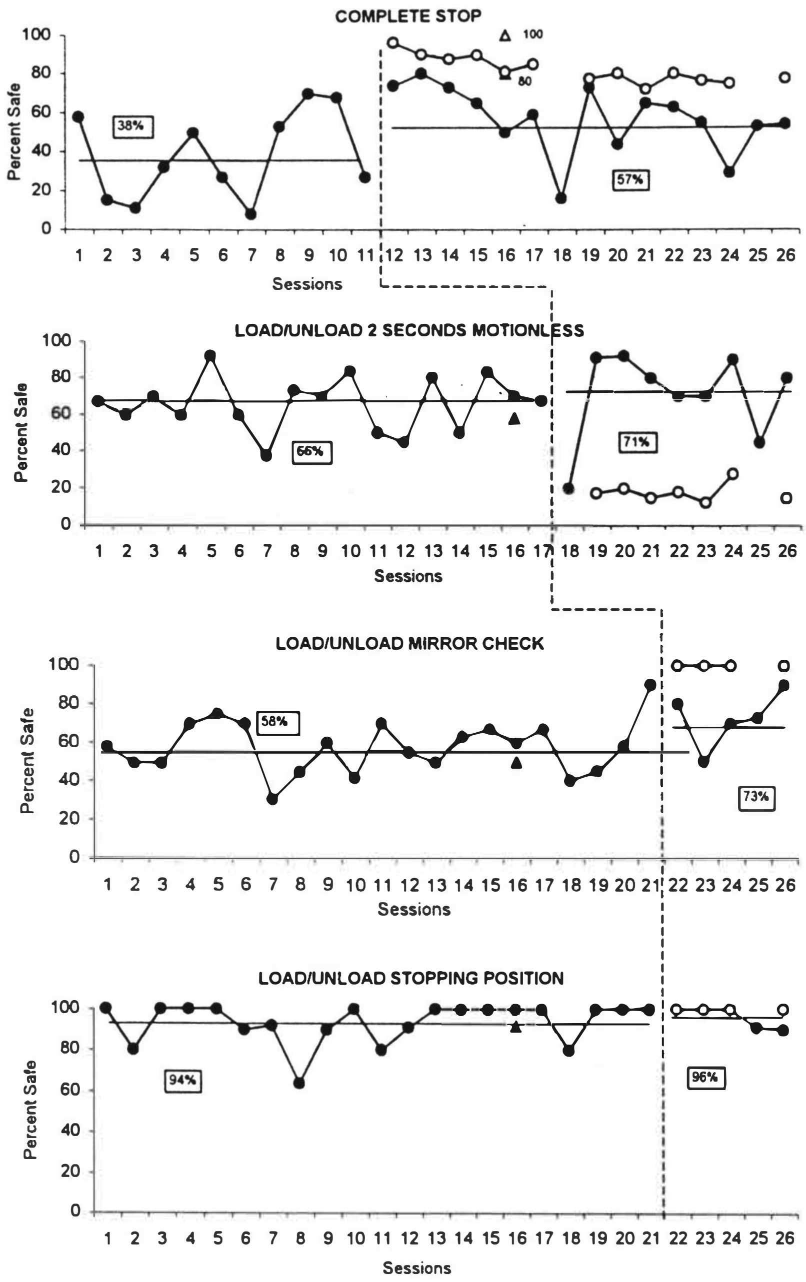
Legend. Closed circle data points are experimenter data, open circle data points are self-monitoring data, closed triangles are experimenter data during supervisor probes, and open triangles are supervisor data.

Figure 6. Purple Participant Results in Multiple Baseline Design Format.



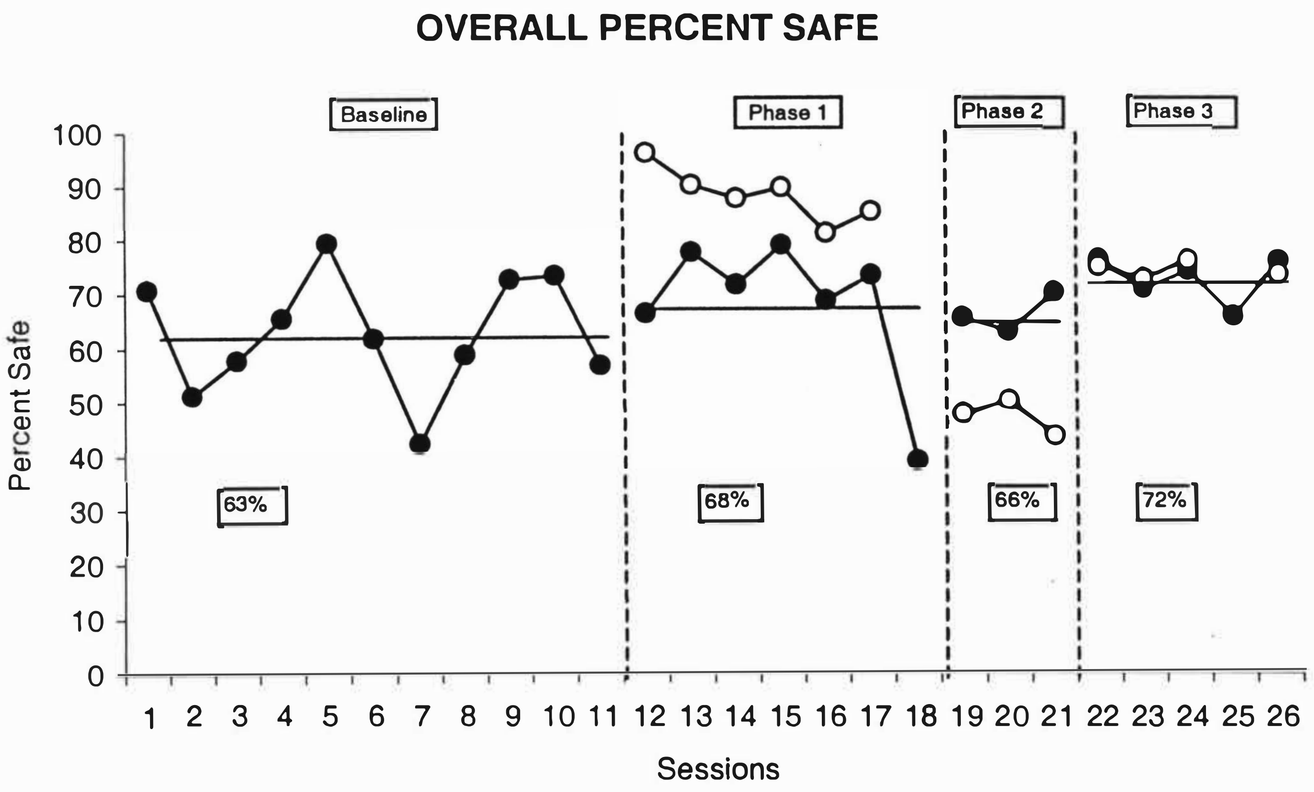
Legend. Closed circle data points are experimenter data, open circle data points are self-monitoring data, and vertical broken lines represent intervention phase changes.

Figure 7. Purple Participant Overall Percent Safe Scores Across Phases.



Legend. Closed circle data points are experimenter data, open circle data points are self-monitoring data, closed triangles are experimenter data during supervisor probes, and open triangles are supervisor data.

Figure 8. Blue Participant Results in Multiple Baseline Design Format.



Legend. Closed circle data points are experimenter data, open circle data points are self-monitoring data, and vertical broken lines represent intervention phase changes.

Figure 9. Blue Participant Overall Percent Safe Scores Across Phases.

mirror check, 94%; bus stopping position, 81%). His actual overall percent safe score, as calculated from experimental observations, was 73% (complete stop, 77%; two seconds motionless, 62%; mirror check, 88%; bus stopping position, 78%). The largest discrepancy between his self-monitoring data and experimenter data occurred for mirror check, with a difference of 6%. The smallest discrepancy occurred for complete stop, with a difference of 2%. Refer to Figures 2 and 3 to view these data comparisons.

The average of the green participant's percent safe estimations across all intervention phases was 98% (complete stop, 99%; two seconds motionless, 100%; mirror check, 100%; bus stopping position, 81%). His actual overall percent safe score as calculated from experimental observations was 53% (complete stop, 92%; two seconds motionless, 31%; mirror check, 41%; bus stopping position, 52%). The largest discrepancy between his self-monitoring data and experimenter data occurred for two seconds motionless, where the difference was 69 percent. The smallest discrepancy occurred for complete stop, where the difference was 7%. Refer to Figures 4 and 5 to view these data comparisons.

The average of the purple participant's percent safe estimations across all intervention phases was 78% (complete stop, 85%; two seconds motionless, 99.9%; mirror check, 100%; bus stopping position, 100%). His actual overall percent safe score as calculated from experimental observations was 65% (complete stop, 47%; two seconds motionless, 59%; mirror check, 80%; bus stopping position, 93%). The largest discrepancy between his self-monitoring data and experimenter data occurred for two seconds motionless, where the difference was 40.9 percent. The smallest discrepancy occurred for bus stopping position, with a difference of 7%. Refer to Figures 6 and 7 to view these data comparisons.

The average of the blue participant's percent safe estimations across all intervention phases was 74% (complete stop, 82%; two seconds motionless, 18%; mirror check, 100%; bus stopping position, 100%). His actual overall percent safe score as calculated from experimental observations was 71% (complete stop, 57%; two seconds motionless, 71%; mirror check, 73%; bus stopping position, 96%). The largest discrepancy between his self-monitoring data and experimenter data occurred for two seconds motionless, where the difference was 53 percent. The smallest discrepancy was for bus stopping position, where the difference was 4%. Refer to Figures 8 and 9 to view these data comparisons.

Independent Variable Integrity

The degree to which the intervention was delivered as planned was measured in three different ways, (1) percentage of self-monitoring forms completed, (2) percent compliance in signing the feedback graph, and (3) percentage of supervisor prompts delivered to drivers. These results are presented first for the entire group, and then for each individual participant

Group Independent Variable Integrity

Group compliance with the rule to fill out two estimations of safe performance each day was 76.5%. This means that drivers completed 76.5% of the estimations they were assigned to make. During phases one, two and three of the intervention, compliance was 91.5%, 72.5%, and 60.5% respectively.

Group compliance with the rule to sign the feedback graph at the end of each shift was 58.8%. This means that drivers signed the graph on 58.8% of the days of intervention when they were working. During phases one, two, and three of the intervention, compliance was 43.3%, 52%, and 85.5% respectively.

Supervisors were asked to initial a form each time they provided a prompt for a participating driver. This measure of independent variable integrity reflects the number of prompts that were recorded on that form. Drivers received 68.3% of the prompts that were planned. Individual participants received at least one prompt on 88.3% of the days during the project, and received two daily prompts on 48.3% of the days during the project. During phases one, two, and three of the intervention, compliance was 66%, 81.5%, and 57.5% respectively.

Individual Participant Independent Variable Integrity

The yellow participant was 82% compliant with the rule to estimate safe performance twice each day (phase one, 100%; phase two, 100%; phase three, 33%). He was 50% compliant with the rule to sign the feedback graph at the end of each shift (phase one, 50%; phase two, 33%; phase three, 67%). The yellow participant received 73% of the planned supervisor prompts. He received at least one prompt on 91% of days during the intervention, and received two prompts on 55% of days during the intervention. A number representing overall independent variable integrity was calculated by averaging the compliance with self-monitoring forms, compliance with signing the feedback graph, and percent of supervisor prompts received. For the yellow driver, overall independent variable integrity was 68.3%. For a summary of the yellow participant's independent variable integrity see Table 1 below.

Table 1

Yellow Participant Independent Variable Integrity

Independent Variable	Phase One	Phase Two	Phase Three	Overall
Self-monitoring	100.0	100.0	33.0	82.0
Feedback	50.0	33.0	67.0	50.0
Supervisor Prompts	80.0	66.7	66.7	73.0
Overall IV Integrity	76.7	66.6	55.6	68.3

The green participant was 83% percent compliant with the rule to complete two estimations of safe performance each day (phase one, 100%; phase two, 69%; phase three, 67%). He was 85% compliant with the rule to sign the feedback graph at the end of each shift (phase one, 83%; phase two, 75%; phase three, 100%). The green participant received 65% of the planned supervisor prompts. He received at least one prompt on 85% of days during the intervention, and received two prompts on 46% of days he worked during the intervention. Overall independent variable integrity for the green driver was 77.7%, which was the highest score for the group of participants. For a summary of independent variable integrity for the green participant see Table 2 below.

Table 2

Green Participant Independent Variable Integrity

Green Driver	Phase One	Phase Two	Phase Three	Overall
Self-monitoring	100.0	69.0	67.0	83.0
Feedback	83.0	75.0	100.0	85.0
Supervisor Prompts	58.3	87.5	50.0	65.0
Overall IV Integrity	80.4	77.2	72.3	77.7

The purple participant was 73% compliant with the rule to complete two estimations of safe performance each day (phase one, 83%; phase two, 63%; phase three, 67%). He was 62% compliant with the rule to sign the feedback graph at the end of each shift (phase one, 40%; phase two, 75%; phase three, 75%). The purple participant received 62% of the planned supervisor prompts. He received at least one prompt on 85% of days during the intervention, and received two prompts on 38% of days during the intervention. Overall independent variable integrity for the purple driver was 65.7 percent. For a summary of independent variable integrity for the purple participant, see Table 3 on the following page.

Table 3

Purple Participant Independent Variable Integrity

Purple Driver	Phase One	Phase Two	Phase Three	Overall
Self-monitoring	83.0	63.0	67.0	73.0
Feedback	40.0	75.0	75.0	62.0
Supervisor Prompts	58.3	75.0	50.0	62.0
Overall IV Integrity	60.4	71.0	64.0	65.7

The blue participant was 68% compliant with the rule to complete two estimations of safe performance each day (phases one, 83%; phase two, 38%; phase three, 75%). He was 38% compliant with the rule to sign the feedback graph at the end of each shift (phase one, 0%; phase two, 25%; phase three, 100%). The blue participant received 73% the planned supervisor prompts. He received at least one prompt on 92% of days during the intervention, and received two prompts on 54% of days during the intervention. Overall independent variable integrity for the blue participant was 59.7%, which was the lowest score for the group of participants. For a summary of independent variable integrity for the blue participant see Table 4 on the following page.

Table 4

Blue Participant Independent Variable Integrity

Independent Variable	Phase One	Phase Two	Phase Three	Overall
Self-monitoring	83.0	38.0	75.0	68.0
Feedback	0.0	25.0	100.0	38.0
Supervisor Prompts	80.0	75.0	62.5	73.0
Overall IV Integrity	54.3	46.0	79.2	59.7

Reliability

A total of 99 experimental observations of driver performance took place over the course of the study. Two independent observers collected data simultaneously for 30 sessions (30.3% of total sessions). The average agreement percentage was 89.8% (range, 70-100). Inter-observer agreement (IOA) scores were calculated for each dependent variable for every IOA session. Agreement scores under 80 percent were limited to 11 out of 120 total IOA calculations. Table 5 on page 51 shows ranges of IOA scores for each dependent variable over the course of the study.

Debriefing and Survey Results

At the conclusion of the study the participants met with the operations supervisor and student consultant for lunch and debriefing. A survey was administered to the drivers to solicit their opinions about the process. After

Table 5

Inter-observer Agreement Percentages for Each Dependent Variable

Dependent Variables	Average IOA	Range IOA	Sessions <80%
Bus stopping position	93.2%	70.0-100	2
2 seconds motionless	90.0%	73.0-100	0
Check mirrors	84.1%	70.0-100	8
Complete Stop	91.8%	77.2-100	1

the survey was administered, participants were informed about the experimental observers from Western Michigan University. A summary of the self-monitoring results and the average percent improvement for each individual (as observed by experimental observers) were delivered to the group with only color codes to identify individuals. After discussion about the process, the operations supervisor left the room and the participants received instructions about signing consent forms for the use of data. All four participants chose to sign the consent form.

For the survey question "How accurate do you think your self-monitoring estimations were (1 = too low; 3 = completely accurate; 5 = too high)," the group average score was 3.5 (yellow, 2; blue, 4; purple, 5; green, 3). When participants were asked how much their behavior actually changed (1 = no change, 5 = a great deal of change) the group average score was 4.25 for complete stop (yellow, 5; blue, 4; purple, 4; green, 4), 3.5 for two seconds motionless (yellow, 4; blue, 3; purple, 3;

green, 4), 3.25 for mirror check (yellow, 4; blue, 3; purple, 3; green, 3), and 3 for bus stopping position (yellow, 3; blue, 3; purple, 3; green, 3). These self report measures corresponded to actual improvements where complete stop realized the largest improvement of 20.8 percent (Likert score, 4.25) and bus stopping position was the least improved performance for the group at 6.3 percent (Likert score, 3).

Participants were asked "If you changed your driving even a little bit, why did you change?" The following options were provided and participants could circle as many as they felt applied: (a) Possibility of punishment for not complying, (b) using the self-monitoring forms prompted me to think about my driving more often, (c) seeing my estimations posted in the drivers lounge, (d) It was important to me to do my best at my job (it was rewarding to improve), (e) I care about passenger safety, (f) encouragement from co-workers, (g) pressure/harassment from co-workers, (h) The behaviors we self-monitored were important because they were related to locations where collisions are likely to happen, (i) pressure from supervisors, and (j) other. All four drivers reported using the self-monitoring forms as a reason they changed their performance (option b) and that they cared about passenger safety (option e). The yellow participant circled items b, d, and e. The blue participant circled items b and e. The purple participant circled items b, c, d, e, and h. The green participant circled items b and e.

Next, participants were asked "If you didn't change your driving much, why was this the case?" The options were (a) thought the behaviors we self-monitored were not very important; (b) too much hassle to worry about; (c) not any financial incentive to change; (d) making complete stops, pausing for passengers, etc. prevents me from getting ahead of schedule; (e) did not understand the graphs; (f) accidents/collisions just don't happen enough to warrant any extra effort to prevent them; (g) definitions of behaviors were hard to understand; and (h) other. The yellow participant circled items a, d, and g. Blue circled option b. Purple wrote, "it helped my driving" in option h. Green circled option f.

Participants were asked to rank order different aspects of the process from (1), the most useful/favorite aspect to (8), the least useful/least favorite aspect of the process. The items mentioning supervisor involvement received the lowest rankings. Being able to share opinions about the project and talking with co-workers about safety and aspects of the route received the highest rankings. The group scores were averaged and fell in order as follows: (1) Being able to share opinions about the project, (2) talking with co-workers about safety and aspects of the route, (3) Meetings to discuss the project, (4) using self-monitoring forms, (5) graphs of safe performance, (6) process not attached to discipline in any way, (7) supervisors observed the same behaviors we did, and (8) more frequent contact from supervisors.

Participants were also asked to make comments about their choices on the ranking procedure. The yellow participant wrote that the most useful aspect of the process was "It caused me to consider the effects on others (students) of my errant behavior (rolling stops)." About his least favorite aspect of the process he wrote, "Supervisor riding the bus caused suspicion about other motives in their observations, could be used for later unrelated discipline." The blue participant enjoyed meetings as escape from routine and the ability to give input about safety. With regard to contact from supervisors he said, "less is better." The purple participant liked the graphs and did not comment about least favorite aspects of the process. Concerning the most useful aspect of the process the green participant wrote "Complete stops are important. A lot can happen in a short amount of time at an intersection. Really have to stop completely to see the whole picture," and "Safety comes first in my mind."

All four participants recommended extending the use of a self-monitoring process to other parts of the organization for both new and experienced drivers. They also responded favorably to having the union participate in the process of choosing behaviors for the checklist. When asked to "Rate the value of this project for you (1 = complete waste of time, 3 = somewhat valuable, 5 = very beneficial)" the group average score was 4 (yellow, 5; blue, 3; purple, 4; green, 4). With regard to supervisor performance, participants reported that supervisors delivered the prompts about once each day on average. When asked, "How would you rate your supervisor's observation of your performance during the project (1=not serious, mostly casual; 3, casual/neutral; 5, serious and professional)," the group average score was 2.75 (yellow, 2; blue, 3; purple, 3; green, 3).

DISCUSSION

Discussion of Group Performance

The object of the study was to demonstrate the effectiveness of a self-monitoring process at increasing critical safe behaviors. The intervention was designed to resemble procedures used in the BST consultation effort with bus operators (Krause, 1997). Experimental procedures were used to evaluate improvements and ensure the reliability of behavioral data. The self-monitoring procedures created a small to moderate overall improvement for the group over baseline levels (12.5%), with a wide range of individual levels of improvement on specific dependent variables (range: 2%-41%). See Figure 1 to review the grouped data.

Applied Implications

One of the main implications regarding group data is the promising nature of this process for improving the safe performance of lone workers. The intervention phases were relatively short, with the entire intervention lasting only three weeks. Without any opportunity to generate participant "buy in" or allow participants to familiarize themselves with the new process, a 12.5% improvement in overall safe performance was achieved. It is important to consider how moderate improvements, like those produced in the current study, would impact an organization over time. An example of such impact can be discussed in terms of the performance improvement of the purple participant, whose overall average improvement was 9%.

Performance improvement for the purple participant was not clearly visible in graphic form until after average lines were added to the line graph (see Figures 6 and 7). Small to moderate improvements like these may or may not be practically important. To explore this issue we should consider the hypothetical cumulative effects of this driver's moderate improvements in safe behavior in our setting. For example, there were usually about 10 instances of loading/unloading passengers every 30 minutes on the route. During a ten-hour shift with a constant flow of passengers, each driver could potentially stop to unload or load 200 times each day. During baseline conditions, the purple driver checked both side mirrors 65% of the time when loading/unloading passengers. This represents 130 safe mirror checks out of 200 opportunities each day. During intervention conditions, he checked both side mirrors 80% of the time. This represents 160 safe mirror checks out of 200 opportunities each day. During one month performing at his baseline level for mirror check the purple participant would achieve 2080 safe mirror checks out of 3200 opportunities. Intervention level performance would achieve 2560 safe mirror checks out of 3200. So his 15% average improvement, not clearly visible in a line graph, could result in as many as 480 fewer at risk load/unload instances each month. If the remaining 64 drivers working in the transit system were also participating in the project and improved to similar levels (with similar passenger rates), the transit system could realize 31,200 fewer at-risk behaviors each month. Consultants applying BBS with lone workers would find these cumulative effects promising, especially considering the absence of "buy in" activities that could increase the power of this type of intervention.

Even though the short term improvements in the current study could be considered promising, it would be important for future research to examine performance improvements over longer periods of time. It is possible that the participants in the current study would have improved to much greater levels with more time and higher levels of participation or "buy in." It is also possible that performance would eventually return to baseline levels as drivers habituated to the process (if it functioned primarily as an antecedent intervention). To truly demonstrate the effectiveness of this process as an organizational practice, sustained improvements would need to be demonstrated.

Discussion of Group Self-monitoring Results

The average of group self-monitoring estimations during intervention was 86.3% while the average of experimenter observations of safe performance during intervention was 68.1% (18.2% difference). This same general tendency for self-estimations to be inflated was reported in McCann and Sulzer-Azaroff (1996) where participants overestimated their performance during the first phase of intervention and in Austin, Wellisley, and Olson (1998) where participants overestimated their performance by 12%. One exception to this general tendency in the current study occurred for the blue participant, who underestimated his performance on two seconds motionless by 53%. Some participants were much more accurate in their self-estimations than others. For example, the discrepancy between self-estimations and experimenter data for the yellow participant was only 1%. In contrast, the discrepancy between self-estimations and experimenter data for the green participant was 45%. It is interesting to note that, during the survey, participants were able to accurately judge which target performance they improved the most and which target performances remained the same, in spite of their generally inaccurate estimations during the project. This suggests that participants only improved the performance they "wanted" to improve, meaning each individual's unique set of values may have played an important role in individual patterns of performance improvement. This is interesting because values can be defined as a set or constellation of conditioned reinforcers (Malott, Whaley, & Malott, 1997). The implications for BBS practitioners is that employee involvement in the selection of target performances may allow participants to choose targets related to their own unique set of conditioned reinforcers. With targets in place that are related to participant reinforcers, greater or more consistent performance improvement could be expected.

Possible Behavioral Functions of the Self-monitoring Processes

The group improvements in safe performance achieved in this demonstration study are similar in size to improvements achieved with antecedent interventions targeting safety belt use. For example, Austin, Alvero, and Olson (1998) found a

20% increase in safety belt use when patrons of a restaurant were verbally prompted at the door. Complete stop for the group of participants in the current study improved by a very similar 21.8%. In another safety belt study, Engerman, Austin, and Bailey (1997) observed a 12% increase in safety belt use when grocery store patrons were prompted at the side of their vehicles. This average percent safe figure is just 0.5% different than the group average improvement in the current study (12.5%).

Another similarity between the current study and safety belt studies is the variability of the data. For example, the range of group percent safe scores for complete stop during baseline was 30%-58% percent with a mean of 46.2%. During intervention, the range for complete stop was 42%-80% with a mean of 68%. Average improvement was substantial, but considerable overlap in range occurred. In Austin, Alvero, and Olson (1998), the range of safety belt use during baseline conditions was 39%-83% with a mean of 57%. During intervention conditions, safety belt use ranged from 61%-100% with a mean of 77%. As with the current study, average improvement was substantial (20%), but the data were highly variable with considerable overlap in range. Safety belt studies have not been conducted with single subjects, so comparisons to the current study in that respect are not direct. However, safety belt use involves the same underlying contingencies that most safety improvement efforts face, which is that immediate and probable consequences support risky performance, while delayed and improbable consequences fail to support correct/safe performance. Most BBS demonstrations in which the use of consequences are explicit demonstrate, on average, much larger changes in behavior.

This further supports the notion that the self-monitoring procedure functioned primarily as an antecedent in the current study.

Whether self-monitoring procedures with lone workers tend to function primarily as antecedents as does a verbal prompt is an interesting research question. Future research could explore this question by requiring the self-monitoring to take place either just before (antecedent function), or just after (consequence function) a work shift. If self-monitoring tends to function primarily as an antecedent, practitioners should stress correlating such processes with reinforcement to ensure their prolonged effectiveness. Some individual performance suggests that the self-monitoring intervention became less effective over time in the current study. The blue participant did not sign the feedback form (consequence related to self-monitoring) once during the first phase of intervention. It appears that the missing feedback component made his self-monitoring less effective over time. The data for the blue participant during phase one resemble data of an organism experiencing extinction (see Figure 8).

A key component missing from the current study was the absence of employee participation in the design stages of the project and other activities said to generate "buy in." Employee participation is heavily promoted by BBS expert Tom Krause (1997). Krause argues that such "buy in" and participation activities are critical for achieving exemplary improvements in safe behavior. Such employee involvement may function as a conditioned establishing operation, where the value of consequences related to safety improvement are increased, and behavior correlated with those improvements is more frequently evoked. It may be the case that an important aspect of employee "buy in" is the degree to which employees value the performances targeted for improvement. Participant comments on the debriefing survey suggest that employees may make the greatest improvements when they value the target performance. The green participant improved complete stop by 41%. His survey comments regarding the self-monitoring process emphasized this specific target performance. He wrote, "Complete stops are important. A lot can happen in a short amount of time at an intersection. Really have to stop completely to see the whole picture." The yellow driver also realized a substantial improvement for the target complete stops (14%). He mentioned stopping performance explicitly when he wrote about the benefits of the self-monitoring process, "It caused me to consider the effects on others (students) of my errant behavior (rolling stops)." These results and self-report measures suggest that learning experiences prior to the onset of BBS observations may function as important establishing operations for consequences associated with making improvements in safe behavior. Future research should examine more closely this potential relationship between employee buy in activities and the effectiveness of self-monitoring.

Discussion of Individual Performance

It was hoped that very consistent effects would be observed across participants, or at least systematic improvements related to the degree to which participants complied with intervention procedures. Consistent improvement was observed for the yellow participant, who was also the most accurate self-estimator of performance in the group. However, each participant's largest improvement was not necessarily the most accurately self-estimated performance. Among individual participants there were very small to very large improvements for specific target performances. Understanding individual differences in performance requires a consideration of the accuracy of each participant's self-monitoring estimations, the integrity of the independent variables for each participant, the self-report data obtained from each participant, and anecdotal information obtained by experimental observers.

Discussion of Yellow Participant Performance

The yellow driver realized the greatest average improvement and the most consistent improvements of any participant. He was also the most accurate estimator of his safe performance. The average of his self-estimations of safe performance was 73%, which was only 1% higher than experimenter data. The largest discrepancy between his self-monitoring data and experimenter data occurred for mirror check , with a difference of 6%. Upon visual inspection of his data, it is clear that his estimations closely tracked his actual performance (see Figures 2 and 3). These results support the findings of McCann and Sulzer-Azaroff (1996) who found that the greatest improvements in safe performance occurred when participants recorded their most accurate self-estimations of safe performance.

Overall independent variable integrity for the yellow participant was 68.3% (Phase one, 76.7%; Phase two, 66.6%; Phase three, 55.6%). The decline in integrity percentages was largely the result of decreased participation in self-monitoring procedures (33% compliant during phase three). This might partially explain the sharp drop in his performance on bus stopping position during the last two days of intervention.

The yellow participant appeared to be very deliberate and conscientious and seemed to take great pride in his profession. It is possible that certain personality characteristics could predict initial compliance with self-monitoring procedures. This could be useful knowledge with regard to planning for implementations of self-monitoring procedures. The yellow participant also responded very systematically to the presence of a supervisor, where his performance on the variables being self-monitored was about 20% higher than his performance on the same day without supervisor presence. Dependent variables that were not being self-monitored remained at baseline levels. This effect demonstrates relatively low reactivity to experimental observers as compared to reactivity to supervisor presence.

Discussion of Green Participant Performance

The green participant performed in paradoxical fashion with a very large (41%) average improvement for complete stop and small (3%) average improvements on all other dependent variables. His self-estimations of safe performance for complete stop averaged 98%, which were only 6% higher than experimental data.

However, the accuracy of these estimations may reflect a response bias, as he estimated all four dependent variables very high throughout the study (average of all self-estimations, 98%). His self-estimations for the other dependent variables were very inaccurate, with discrepancies as high as 69%. The green participant's responses to survey questions indicated that he felt his estimations of performance were "completely accurate," and that he did not think his mirror check or bus stopping position performances changed at all during the project. Although his estimations were often over 90% for these variables, his actual performance was 41% safe for mirror check and 52% safe for bus stopping position. There are many possible reasons for this discrepancy. It may have been the case that he did not understand the definitions of the correct performances, or perhaps he feared some type of discipline for reporting low percent safe scores. Interactions with him during the debriefing meeting seemed to indicate that he simply did not value these target performances as much as he valued complete stops, and that he did not want to risk future punishment for reporting low scores. However, the results of the supervisor probe for this participant "muddy the waters" of this interpretation. The probe occurred during the final phase of the study where all four dependent variables were being self-monitored. With the supervisor present he scored 100% safe for complete stop, 57% safe for two seconds motionless, 64% safe for mirror check, and 91% safe for bus stopping position. If fear of discipline was a significant motivator for this participant, we would have expected all dependent variables to score at least as high as his self-estimations of performance when a supervisor was present. His estimations on that day were 100% for complete stop, 100% for two seconds motionless, 100% for mirror check, and 89% for bus stopping position. In light of these data, it seems most probable that he really did not completely understand the criteria for the correct performance of these targets. Future research should control for this potential experimental confound.

Integrity of the independent variables for the green participant were the highest of all participants at 77.7%, with a general decline in integrity as new performances were introduced (Phase one, 80.4%; phase two, 77.2%; phase three, 72.3%). This high participation may reflect his readiness to actively work on safety. If he had participated in the selection of target performances (where he could impact the inclusion of performances he valued), it is possible that his performance improvement would have been more consistent. His paradoxical performance may represent the need for employee participation in the development stages of BBS (activities said to generate participant "buy in").

Observers noticed that the green participant's demeanor and performance varied dramatically on occasion. When he seemed upset or short on patience, his performance on bus stopping position and two seconds motionless became very risky (door open much too soon at loading zones and no pause after passengers loaded/unloaded). These target behaviors may have produced preferred aggressive reinforcers whose value was established when the green participant had experienced aversive stimulation. It has been postulated that frustrating aversive events may act as establishing operations for aggressive reinforcers (i.e., a raised heart beat, raised adrenaline levels, sight of others in pain or afraid, etc), evoking behaviors that generate such biological stimulation (Malott, Whaley, & Malott, 1997). It is also possible that competing reinforcement contingencies, aggressive or otherwise, were simply of much greater value than those provided by the intervention (see contingency analysis section following the discussion of individual performance).

Discussion of Purple Participant Performance

It was mentioned previously that performance improvement for the purple participant was not clearly visible in graphic form until after average lines were added to the line graph (see Figures 6 and 7). However, it is interesting to analyze the performances for which he realized his greatest average improvements during the study. The purple participant was highly sociable with passengers and made his greatest improvements on dependent variables explicitly related to passenger and pedestrian safety with a 15% improvement for mirror check, a 12% improvement for two seconds motionless, and a 12% improvement for bus stopping position. This may be further anecdotal evidence that participants made the greatest improvements on performance targets that were personally valued prior to the study.

The average of his self-estimations of performance was 78%, which was 13% higher than experimenter data. He estimated his performance at or near 100% for two seconds motionless, mirror check, and bus stopping position throughout the study. This strong pattern of response bias makes it difficult to interpret any relationship between the accuracy of his estimations and concurrent performance improvement. A supervisor probe occurred for this participant during the final phase of the intervention, and he scored between 15% and 20% higher on each dependent variable than he did on the same day without the presence of a supervisor.

The purple participant did not seem to take the project seriously and was overheard poking fun at his "safety self-monitoring" on several occasions by experimental observers. In spite of this apparent lack of respect for the process, he did make improvements in his safe behavior and recommended the process highly in his survey responses. This paradox should encourage consultants to make data-based decisions when experiencing resistance and difficulties with employees who are participating in BBS processes, rather than making choices based on casual observations of behavior. The data revealed that performance changed and that the driver enjoyed the process. Casual observations of his behavior alone would have discouraged the transit system from adopting the self-monitoring process permanently.

Discussion of Blue Participant Performance

The blue participant achieved the smallest overall improvement for the group with an average 8% increase in safe performance over baseline conditions. He also had the lowest overall independent variable integrity of all participants at 59.7%. The clearest effects for this participant occurred during the first phase of intervention. During baseline conditions he seemed to come to a complete stop only when he was forced to do so by traffic conditions. His typical pattern of performance was to roll slowly through stop signs. This distinctive pattern of performance observed during baseline made behavior changes observed on the first day of intervention very dramatic. The deterioration of this improvement in performance was also distinctive as it gradually returned to baseline levels over the six sessions of phase one (see Figure 8). Contributing to this effect may have been the fact that the blue participant did not sign the feedback graph at all during phase one of the intervention, thereby eliminating the built in consequence component of the intervention. Another clear effect achieved during the first phase of the study for this participant was the results of the supervisor probe. He scored almost 40% higher on complete stop when the supervisor was present than he did when the same performance was measured on the same day without the presence of a supervisor. In addition, the baseline dependent measures all showed slightly lower performance with the supervisor present than they did without the presence of the supervisor, showing that the participant was reactive only to the performance being self-monitored.

At the onset of phase two, the blue participant's performance dropped to about 20% on the target performances for that phase. At the time it was postulated that this might represent counter controlling behavior in response to the intervention procedures. However, this extremely low pattern of performance did not continue beyond the first day of phase two of the intervention.

The accuracy of the blue participant's self-estimations did not seem to systematically vary with his performance improvement. The smallest discrepancy between his self-estimations and experimenter data occurred for bus stopping position with an average difference of 4%. This performance, however, improved by only 2% over baseline levels. His estimations were discrepant from experimenter data by 25% for complete stop, where he realized his greatest improvement (19% over baseline levels). He realized a small 5% improvement for two seconds motionless, where his self-estimations were 53% discrepant from experimenter data.

Contingency Analysis

A central dilemma underlies most safety improvement efforts, and this is that accidents and injuries or, in the present setting, collisions occur infrequently. Aversive outcomes in safety are infrequent, low probability consequences that often fail to maintain avoidance responses. Therefore, more immediate and probable consequences tend to shape behavior instead of the logical rules related to avoiding potentially fatal aversive outcomes. One driver's answer to a survey question highlights this issue. When explaining why some of his behavior did not change very much, the green driver circled the statement "accidents/collisions just don't happen often enough to warrant any extra effort to prevent them." Aversive outcomes like collisions, as horrific as they may be, tend to be too delayed and too uncertain to motivate safe behavior. In addition, the safest way of doing things often requires the person to endure immediate aversive conditions (taking longer to complete a task, wearing uncomfortable personal protective equipment, etc.). The goal of BBS is to overcome the contingencies that favor risk taking (unsafe performance) by providing more immediate and probable consequences for safe performance.

Both Aubrey Daniels (1989) and Tom Krause (1997) have developed coding systems for describing consequences for specific target performances. Both systems examine three key issues: (1) the value of consequences, (2) the temporal relation of the consequence to the target performance, and (3) the probability of the consequence. When analyzing safe behavior, labeling consequences with one of these coding systems can reveal whether the environment or organizational context generally favors risky or safe performance. To explore possible reasons for the effectiveness of the intervention used in the current study, Daniels system (1989) was used to analyze the driving performances measured. Daniels suggests analyzing the problem (at-risk) performance first, and then analyzing the desired (safe) performance next. The analysis consists of listing the antecedents and consequences for each target performance and then scoring each consequence according to its behavior strengthening qualities. Each consequence is scored as being either positive or negative (P/N), immediate or in the future (I/F), and certain or uncertain (C/U). Positive, immediate, and certain consequences (PIC) tend to maintain or increase behavior and are likely to qualify technically as reinforcement or as an analog to reinforcement (Malott, Whaley, & Malott, 1997; Skinner, 1953). Negative, future, and uncertain consequences tend to decrease or eliminate behavior (NFU) and are likely to qualify technically as punishment or an analog to punishment (Malott, Whaley, & Malott, 1997; Skinner, 1953). Tables showing this analysis for each of the four target performances are included below. Table 6 on page 71 shows an analysis of the problem performance "rolling stop" and consequences hypothetically

Table 6

Analysis of the At-Risk Performance of Rolling Stops

Antecedents	Consequences	Rating
Traffic is approaching rapidly	Acquire position in traffic without waiting.	PIC
Hear aversive sound of squealing brakes	Aversive sound less intense and of shorter duration than during a complete stop.	PIC NIC value may have been altered by intervention
Bus approaches a stop sign	Bus strikes a passenger or hits a vehicle.	NIU(very uncertain)
Bus approaches stop sign	Forward motion continues.	PIC NIC value may have been altered by intervention
Pedestrians approaching the intersection to cross in front of bus	Pedestrians stop and wait for bus to pass because it rolls through intersection.	PIC NIC value may have been altered by intervention
Bus approaches stop sign	Traffic ticket, Disciplined for a moving violation.	NIU, NFU
Bus approaches stop sign	Minimal muscular exertion on brake pedal.	PIC
Bus approaches stop sign	Must score percent safe estimation lower or be dishonest.	NF(within hours)C
Bus approaches stop sign	Sees percent safe scores posted in driver's lounge.	NF(next day)C
Bus approaches stop sign	Talks about performance at a meeting.	NFC

*Highlighted areas represent intervention conditions

available for that performance before and after intervention. Table 7 on page 72 shows an analysis of the correct performance "complete stop" and the consequences hypothetically available for that performance before and after intervention.

Table 7

Analysis of the Safe Performance of Complete Stops

Antecedents	Consequences	Rating
Traffic is approaching rapidly	Opportunity to merge is removed.	NIC
Hear aversive sound of squealing brakes	Aversive sound more intense and of longer duration than during a rolling stop.	NIC PIC value may have been altered by intervention
Bus approaches stop	Sight of vehicle or passenger in a risky position in relation to bus.	PIU
Bus approaches stop sign	Forward motion stops.	NIC PIC value may have been altered by intervention
Passenger approaches intersection to cross in front of bus	Passenger crosses in front of bus, causing the driver to wait several seconds.	NIC PIC value may have been altered by intervention
Bus approaches stop sign	Maximal muscular exertion on brake pedal.	NIC PIC value may have been altered by intervention
Bus approaches stop sign	Score percent safe estimation higher and enjoy being honest.	PF(within hours)C
Bus approaches stop sign	Sight of reported percent safe scores graphed and posted in driver's lounge.	PF(next day)C
Bus approaches stop sign	Will talk about this performance with co-workers at a meeting.	PFC

*Highlighted areas represent intervention conditions

Performing a complete stop could be considered safe for many reasons. A few reasons include (a) it is a legal requirement, (b) it creates more time to see traffic and pedestrian conditions and risks, and (c) it allows other drivers to make clear discriminations about right of way and opportunities to proceed forward motion. Several general contingencies discourage this performance.

A driver making a complete stop may be punished by the loss of the opportunity to merge with traffic. In fact, stopping completely may cause the driver to remain motionless for as much as several minutes of time if traffic is heavy and opportunities to merge are limited. There is also good reason to assume that forward motion of a vehicle is reinforcing most of the time. Opportunities to obtain a lunch, take a break and read a paper, or take a restroom break were only available if the bus was ahead of schedule. This condition or policy likely increased the value of forward motion as a reinforcer. Therefore, behavior that caused the bus to stand still could become aversive. In these three examples related to complete stop we can see that safe behavior tended to be discouraged and risky behavior tended to be reinforced. These contingencies may have all contributed to the at-risk stops that were observed during both baseline and intervention phases.

An immediate consequence that seemed to reinforce premature lifting of the foot from the brake involved a screeching sound made by the brakes of many busses. If a driver performed a rolling stop, this stimulus was presented with less intensity for a shorter period of time. In this sense, braking was punished by the presentation of an aversive sound and rolling stops were reinforced by the more immediate cessation the aversive sound. Rolling through a stop sign may also be reinforced frequently by obtaining an immediate position in traffic that is going in the desired direction.

Rolling stops were supported by at least five positive contingencies and discouraged by two uncertain negative contingencies prior to intervention. Intervention procedures created at least three more negative contingencies for rolling stops and may have altered the value of some positive contingencies making them negative or neutral.

Complete stops were discouraged by at least five negative contingencies prior to intervention. One positive consequence was identified for pre-intervention conditions but was not literally contingent on the performance of complete stop. This positive consequence was the sight of pedestrians or other vehicles in risky positions in relation to the bus. A driver could have identified pedestrians or vehicles in risky positions regardless of the motion status of the bus, as sight is contingent upon head position and eye movement. However, the opportunity for "looking behaviors" increases during a complete stop. To summarize, pre-intervention conditions discouraged the safe performance of complete stops. The intervention procedures added negative consequences for the at-risk behavior of rolling stops and positive consequences for the safe performance complete stops. Intervention procedures may have also altered the value of some previously negative contingencies. This same general pattern is evident for all four dependent variables.

Table 8 on page 75 shows an analysis of the problem performance "less than two second pause after loading/unloading passengers" and hypothetical consequences available for that performance before and after intervention. Table 9 on page 76 shows an analysis of the correct performance "two second pause after

Table 8

Analysis of an Inadequate Pause After Loading/Unloading Passengers

Antecedents	Consequences	Rating
Passenger loads/unloads	Immediate forward motion with loading accomplished.	PIC NIC value may have been altered by intervention
Pedestrians waiting near the curb to walk in front of the bus after passenger's finish loading/unloading	Pedestrians remain at the curb and the bus continues without delay.	PIC NIC value may have been altered by intervention
Passenger loads/unloads	Passenger is injured on the bus, or is struck by the bus as it pulls away from the loading zone.	NIU(very uncertain)
Passenger loads/unloads	Passenger complains to the transit system.	NFU
Passenger loads/unloads	Sight of pedestrians or exiting passengers in risky positions in relation to the bus.	PIU(even less probable than during a correct two second pause)
Passenger loads/unloads	Passengers are seated quickly in order to escape the risky position of standing while the bus moves.	PIC NIC value may have been altered by intervention
Passenger loads/unloads	Passengers walk safely down the aisle as the bus moves.	PIC NIC value may have been altered by intervention
Passenger loads/unloads	Must record percent safe estimation lower or be dishonest.	NF(within hours)C
Passenger loads/unloads	Sees percent safe scores posted in driver's lounge	NF(next day)C
Passenger loads/unloads	Talks about performance at a meeting.	NFC

*Highlighted areas represent intervention conditions

Table 9

An Analysis of the Safe Performance of a Two Second Pause After Loading/Unloading Passengers

Antecedents	Consequences	Rating
Passenger loads/unloads	Bus remains motionless with loading accomplished.	NIC PIC value may have been altered by intervention
Pedestrians waiting near the curb to walk in front of the bus after passenger's finish loading/unloading	Pedestrians walk in front of the bus which causes several seconds of delay.	NIC PIC value may have been altered by intervention
Passenger loads/unloads	Passenger falls and is injured on the bus, or is struck by the bus as it pulls away from the loading zone.	NIU(very uncertain)
Passenger loads/unloads	Passenger phones in complements to the transit system.	PFU(very uncertain)
Passenger loads/unloads	Sight of pedestrians or exiting passengers in risky positions in relation to the bus.	PIU
Passenger loads/unloads	Passengers take extra time finding a seat and delay the bus.	NIC
Passenger loads/unloads	Passengers walk safely down the aisle as the bus moves.	PIC
Passenger loads/unloads	Can score percent safe estimation high and enjoy being honest.	PF(within hours)C
Passenger loads/unloads	Sees percent safe scores posted in driver's lounge.	PF(next day)C
Passenger loads/unloads	Talks with co-workers at a meeting.	PFC

*Highlighted areas represent intervention conditions

loading/unloading passengers" and the hypothetical consequences available for that performance before and after intervention.

The behavior of remaining motionless for two seconds after loading/unloading passengers bears some similarity in analysis to complete stops, because both dependent variables require the driver to hold the brake pedal down and keep the bus motionless. As mentioned previously, forward motion probably functioned as a reinforcer most of the time. Considering the reinforcing nature of forward motion, it could be concluded that holding the bus motionless after a passenger boarded or exited would have been aversive. In addition, pedestrians could walk in front of the bus and cause further delays for the driver while the bus was stopped. It was observed many times throughout the study that when a driver starting forward motion of the bus as soon as passengers loaded/unloaded, pedestrians remained on the curb until the bus passed them and cleared the loading zone.

Table 10 on page 78 shows an analysis of the problem performance "looking at fewer than two side mirrors after loading/unloading passengers" and hypothetical consequences available for that performance before and after intervention. Table 11 on page 79 shows an analysis of the correct performance "checking side mirrors after loading/unloading passengers" and the hypothetical consequences available for that performance before and after intervention.

Table 12 on page 80 shows an analysis of the problem performance "poor bus stopping position (door open early or cars able to pass on right) before loading/unloading passengers" and the hypothetical consequences available for that

Table 10

Analysis of the At-Risk Performance of Looking at Fewer Than Two Side Mirrors

Antecedents	Consequences	Rating
Passenger loads/unloads	Minimal neck muscle exertion.	PIC PIC value may have been altered by intervention
Passenger loads/unloads	Passenger falls and is injured on the bus, or is struck by the bus as it pulls away from the loading zone.	NIU(very uncertain)
Passenger loads/unloads	Another vehicle is struck by the bus.	NIU(very uncertain)
Passenger loads/unloads	Sight of vehicles, pedestrians, or exiting passengers in risky positions in relation to the bus.	PIU
Passenger loads/unloads	Sight of interesting things happening around the bus (not visible in mirrors).	PIC
Passenger loads/unloads	Must score percent safe estimation low or be dishonest.	NF(within hours)C
Passenger loads/unloads	Will see reported percent safe scores graphed and posted in driver's lounge.	NF(next day)C
Passenger loads/unloads	Will have to talk about this performance with co-workers at a meeting.	NFC

*Highlighted areas represent intervention conditions

Table 11

Analysis of the Safe Performance of Checking Both Side Mirrors After Loading/Unloading Passengers

Antecedents	Consequences	Rating
Passenger loads/unloads	Maximal neck muscle exertion for the occasion.	NIC PIC value may have been altered by intervention
Passenger loads/unloads	Passenger falls and is injured on the bus, or is struck by the bus as it pulls away from the loading zone.	NIU(very uncertain)
Passenger loads/unloads	Another vehicle is struck by the bus.	NIU(very uncertain)
Passenger loads/unloads	Sight of vehicles, pedestrians, or exiting passengers in risky positions in relation to the bus.	PIU PIU value may have been altered by intervention (even more positive)
Passenger loads/unloads	Sight of mirrors without vehicles, pedestrians, or exiting passengers in risky positions in relation to the bus.	NIC
Passenger loads/unloads	Must score percent safe estimation low or be dishonest.	NF(within hours)C
Passenger loads/unloads	Will see reported percent safe scores graphed and posted in driver's lounge.	NF(next day)C
Passenger loads/unloads	Will have to talk about this performance with co-workers at a meeting.	NFC

*Highlighted areas represent intervention conditions

Table 12

Analysis of the At-risk Performance of Poor Stopping Position

Antecedents	Consequences	Rating
Approaching loading/unloading zone	Passenger loads/unloads quickly because door is open before bus is stopped. Forward motion can begin earlier.	PIC NIC value may have been altered by intervention
Approaching loading/unloading zone	Passenger loads/unloads safely.	PIC
Approaching loading/unloading zone	Passenger loads/unloads while bus is still in motion and is injured, or is struck by another vehicle passing the bus on the right.	NIU(very uncertain)
Approaching loading/unloading zone	Must score percent safe estimation low or be dishonest.	NF(within hours)C
Approaching loading/unloading zone	Will see reported percent safe scores graphed and posted in driver's lounge.	NF(next day)C
Approaching loading/unloading zone	Will have to talk about this performance with co-workers at a meeting.	NFC

*Highlighted areas represent intervention conditions

performance before and after intervention. Table 13 on page 81 shows an analysis of the correct performance "correct stopping position (door opens after complete stop and no cars can pass on right)" and the hypothetical consequences available for that performance before and after intervention.

Table 13

Analysis of the Safe Performance of Correct Stopping Position

Antecedents	Consequences	Rating
Approaching loading/unloading zone	Passenger loads/unloads slowly because door is shut when passenger is ready to board. Forward motion is delayed.	NIC PIC value may have been altered by intervention
Approaching loading/unloading zone	Passenger loads/unloads safely.	PIC(very certain)
Approaching loading/unloading zone	Passenger loads/unloads and is injured, or is struck by another vehicle passing the bus on the right.	NIU(even more uncertain than during poor stopping position
Approaching loading/unloading zone	Must score percent safe estimation low or be dishonest.	NF(within hours)C
Approaching loading/unloading zone	Will see reported percent safe scores graphed and posted in driver's lounge.	NF(next day)C
Approaching loading/unloading zone	Will have to talk about this performance with co-workers at a meeting.	NFC

*Highlighted areas represent intervention conditions

This exercise using Daniels (1989) method of analysis sheds light upon some of the possible contingency changes that were responsible for performance improvements in the current study. Each table illustrated the central safety dilemma, where at-risk performance tended to be supported by positive and immediate consequences and safe performance tended to be discouraged with only delayed and uncertain consequences available. The intervention, in theory, was effective because of the degree to which added intervention contingencies were more powerful than "natural" existing contingencies supporting at-risk performance.

Strengths and Weaknesses of the Current Study

Strengths of the current study include the reliable measures of performance, measures of independent variable integrity, collection of self-report measures at the conclusion of the study, supervisor probes, and the use of an experimental design. The assessment of reliability of observations was absolutely essential for determining the effectiveness of the self-monitoring based intervention. With IOA scores calculated for 30.3% of experimental observations, we can be fairly certain that the intervention actually changed the behavior of participants. Measuring the integrity of the independent variables for each participant was a nice feature because it creates a context for viewing performance changes. In some cases independent variable measures created insight into unusual patterns in the data, such as the blue participant's "extinction like" performance during phase one of the intervention. Measures of the independent variable showed zero compliance with the consequence aspect of the intervention for that phase, which adds strength to the interpretation that the self-monitoring procedure functioned primarily as an antecedent during the study. Participant responses on the survey instrument gave participants a chance to give their opinions about aspects of the process, and gave the experimenter a chance to collect information about covert behavior that may have impacted their performance. Supervisor probes affected performance in a systematic way and may represent one method for assessing participants' understanding of the target performances. For some participants the probes demonstrated that they understood and were capable of performing the target behaviors at high percent safe levels. For the green participant, however, the probe showed that he may not have understood the target behaviors. His performance improved with the supervisor present, but not to the levels expected. In general, the probes demonstrated that supervisor presence was a more powerful intervention than self-monitoring, and that participants improved only the behaviors that supervisors were observing. Finally, the experimental design (in conjunction with reliable measures of performance) demonstrated that behavior changed as a function of the systematic introduction of the intervention. This conclusion could not be made from the data reported from the BST consultation effort with a bus system (Krause, 1997), even though a 66% reduction in accidents and injuries occurred.

Weaknesses of the current study include the relatively short duration of the intervention, the absence of meaningful outcome measures (also due to the short duration), the small number of participants, the lack of employee buy in, and the apparent "low power" of the intervention. The duration of the study was cut short because the particular bus route terminated for summer break. We cannot determine whether performance changes would maintain, improve, or deteriorate with longer duration. Future studies with lone workers should consider designs that provide more time for the stabilization of performance under each experimental condition.

Although collision reduction was the primary interest of the transit system, the absence of collisions among the four participants during the study is not very meaningful. More participants would be needed to truly impact the preventable collision rate of 2.08 per month for the transit system. Also, participants in the current study essentially started "cold" with self-monitoring procedures. This feature does not resemble real world BBS consultation efforts where employees are often heavily involved in the planning stages of the process. It would be an important research innovation to use methods that would allow employees to participate in the selection of dependent measures, and give them time to buy into the process, without compromising the ability of experimental observers to collect reliable measures. In the current study, employee buy in was compromised for the sake of ensuring dependent variables that could be measured reliably. It is possible that using video cameras to tape lone worker performance could become a part of future research studies examining self-monitoring procedures. The low power of the intervention made it difficult to see some of the effects generated by the intervention. An intervention that included more salient establishing operations (eg., employee participation in the planning stages), more powerful antecedent prompts (eg., the presence of a video camera), or more powerful consequences (e.g., incentives contingent upon reaching participation goals) may have generated larger performance improvements. However, the purpose of the current study was to demonstrate the effects of self-monitoring and feedback procedures with lone workers. Additional components would have complicated the ability to see effects generated by the central self-monitoring and feedback procedures.

Conclusions and General Discussion of the Results

The results of the study suggest that self-monitoring procedures alter the safe behavior of bus drivers who work alone. In this sense the study was successful demonstration of the effects reported by Krause (1997) during a consultation effort with a bus transit system. However, because of the small number of participants and short duration of the study, it cannot be concluded that changes in safe behavior led to an important decline in collisions in the current setting. All four participants were "collision free" for five weeks, but the transit system as a whole had three separate months without collisions in 1997.

The independent variables did produce behavior changes, but only demonstrated moderate to low power as a behavior changing package intervention (12.5% overall average improvement). This may have been due to the lack of participant involvement in activities such as dependent variable selection and design of the process. The fact that participants were aware of the short-term nature of the project may have also contributed to this effect. Perhaps they did not take the procedures "seriously" because the process was presented as temporary rather than permanent. A solution to these issues might involve a research partnership with a consultation effort working with greater numbers of lone workers. This might limit the experiment to an ABC design (baseline, process design and dependent variable selection, and intervention), but self-monitoring could be studied in a realistic context, where participants are exposed to important "buy in" activities and remain involved for many months or even years after the initial start date.

The results also suggest that in order to generate the most consistent performance improvement, accurate self-estimations are needed. Aside from the yellow participant, patterns in the other participants' data supported this conclusion by the degree to which low accuracy of self-estimations co-varied with variable performance improvement. However, accuracy did not systematically vary with each participant's largest performance improvements. Further research is needed to specifically address this issue. It may have been the case that accurate self-monitoring and consistent improvements simply happened together for the yellow participant, but were not causally related events.

The study does represent "good news" for BBS practitioners considering the use of self-monitoring processes with lone workers. This good news is two fold: (1) participants rated the process as valuable and recommended that the transit system adopt a form of the self-monitoring procedure permanently, and (2) the intervention caused an improvement in performance, even in the absence of activities to generate "buy in." It should also be noted that all participants rated "sharing my opinions about safety" and "talking with co-workers about safety" among their favorite aspects of the project.

The self-monitoring process seemed to function as an antecedent for the target performances. Evidence for this function occurred for the blue participant, who improved his performance during phase one with 0% compliance with signing the feedback graph. The feedback graph was signed at the end of the day. Signing the graph in that sense functioned as a consequence, but also may have served as an antecedent for performance on the following day of work. Less clear, but very likely, were the consequence functions, rule functions, and conditioned establishing operation functions of the self-monitoring process itself. These principles of behavior may explain, in part, some of the individual specific patterns of performance improvement in the study. The value of consequences related to self-monitoring, the generation of rules, and the effects of conditioned establishing operations would theoretically be reliant upon each participant's unique learning history. When interventions are applied with adult humans having many years of work experience, idiosyncratic responses to the procedures are likely, and are clues that simple operant conditioning is not completely responsible for the effects. Unfortunately the operation of these behavioral principles is only speculative. To conclude that conditioned reinforcers, rule generation, and conditioned establishing operations caused the idiosyncratic performance improvements would be an "ad hoc" and perhaps erroneous explanation. This study did not experimentally control for, or explicitly examine, any of these principles and concepts of behavior.

Closing Comments

The results of the current study suggest that self-monitoring is an effective method for improving the safe performance of bus operators and possibly the performance of other lone workers. The variables that might create a more powerful intervention package are not known, and it is suggested that future research include an intervention component that resembles the consulting practice of employee participation in dependent variable selection and process design. It would also be useful to study lone workers for more extended periods of time. Other future research could explore the behavioral mechanisms responsible for performance improvements generated by self-monitoring procedures. One simple manipulation, suggested earlier in the discussion, could involve requiring participants to self-monitor at different times, either just before work or just after work and look for differential effects on performance. Regarding any future advances in improving the safe performance of lone workers, researchers must address important methodological issues to study these phenomena. Central methodological limitations related to BBS research with lone workers are (a) reactivity to experimental observers, (b) reliability of behavioral data, and (c) scale and expense. Successful extensions of the current study must use methods that minimize participant reactivity to experimental observers, include methods that assess the reliability of behavioral measures, and find ways to study more participants for longer periods of time.

Appendix A

Protocol Clearance From the Human Subjects Institutional Review Board

Human Subjects Institutional Review Board

Kalamazoo, Michigan 49008-3899

WESTERN MICHIGAN UNIVERSITY

Date: 3 March 1999

To: John Austin, Principal Investigator
Ryan Olson, Student Investigator for thesis

From: Sylvia Culp, Chair

Re: HSIRB Project Number 98-12-07

This letter will serve as confirmation that your research project entitled "Evaluation of a Behavior-Based Safety Process with Lone Workers" has been **approved** under the **exempt** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may **only** conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: 3 March 2000

Appendix B

Experimenter Data Collection Form

Checklist Type (circle one): Regular/Reliability
Observer (circle one): Primary/Secondary

LOADING/ UNLOADING PASSENGERS – at least 10 observations, write # of passengers in C or A box																	%
Bus stopping position • Door shut until completely stopped • Positioned so no cars may pass on the right		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	C																
	A																
2 seconds motionless AFTER • Last person steps behind yellow line • Last person steps off bus to the right • Last person steps clear of the front left corner of the bus when exiting to the left		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	C																
	A																
Visually check both SIDE mirrors after loading/ unloading passengers.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	C																
	A																
BUS IN MOTION – Primary observer should announce the number observation prior to each corner.																	%
Cornering (90 degree turns) • Brake BEFORE the turn • At least 2 ft. clearance between side of bus from cars, poles, signs, and people		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
		C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	
AT-RISK: • Heavy or jerky braking during turn • less than 2 feet clearance • Bumping curb or scraping the bottom of the bus • Holding change box or window frame during turn • Holding food during a turn		[illegible]	[illegible]	18	19	20	21	22	23	24	25	26	27	28	29	30	
		C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	
		31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
		C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	
Following distance on straightaways – • Observe this 2 seconds after every corner • Then, count 8 seconds more on watch and observe the distance of the same car again		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
		C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	
STOPPING – Primary observer should announce the number observation prior to each stop.																	%
Complete termination of motion at ~~stop signs, railroad crossings, and~~ red traffic lights. No rolling stops, no creeping at traffic ~~signals~~. If light is green, do not make an observation. Watch the pole of the stop sign as the bus slows to judge cessation of motion.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
		C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	
		16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
		C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	
TOTALS		Correct							At-Risk								Ovr all %

DATE: OBSERVER: BROWN/GOLD

TIME ON	TIME OFF
LOCATION	LOCATION

COLOR: YELLOW/BLUE/PURPLE/GREEN/ WHITE(sub)

Appendix C

Sample Supervisor data Collection Form (Phase Two)

Supervisor Observation Instructions- PHASE TWO

1. Fill in the date, time, and location of the observation at the bottom of the data sheet.
2. Circle the run number of the driver you are observing. **Do not list the drivers name.**
3. Sit toward the front of the bus to make your observation. You should ride with the driver for about ½ hour. **As you travel you will give the driver a score for every stop encountered on the route.** Circle C if the driver correctly comes to a complete stop according to the definition below. If the driver performs a rolling stop or jumps the traffic signal circle A for that instance. When you are finished write the total number of C's and divide that number by the total number of stops observed, and then multiply that by 100. This will give you a percent safe score for complete stops.
4. **For Load/Unload 2 seconds motionless, count "one thousand one, one thousand two" after the last loading/unloading passenger clears the yellow line, exits the bus to the right, or clears the bus on the front left.** If the bus stays motionless for 2 seconds, circle C. If the bus moves at all before you count to two, circle A. You only give one score for each load/unload instance, whether there is one passenger or many. When you are finished write the total number of C's and divide that number by the total number of stops observed, and then multiply that by 100. This will give you a percent safe score for complete stops.
5. Write in the location and end time of the observation
6. Return the data sheet to the locked drop box

STOPPING																%
Complete termination of motion at stop signs and red traffic lights. No rolling stops, no creeping at traffic signals. If light is green, do not make an observation. Watch the pole of the stop sign as the bus slows to judge cessation of motion.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	

LOAD/UNLOAD TWO SECONDS MOTIONLESS																%
The bus should remain completely motionless for 2 seconds AFTER • Last person loading steps behind yellow line • Last person unloading steps off bus to the right • Last person unloading steps clear of the front left corner of the bus when exiting to the left	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	C A	

DATE: SUPERVISOR: BROWN/GOLD

RUN#: (78211), (78221), (78231), (78241)

TIME ON	TIME OFF
LOCATION	LOCATION

Appendix D

Sample Self-monitoring Form (Phase Three)

SELF-MONITORING FORM

Estimate each behavior twice each day. The goal is to actively work on improving the behaviors on the checklist.

Run number:_______________ Date:_______________

BEHAVIOR	ESTIMATION ONE Percent safe	ESTIMATION TWO Percent Safe	AVERAGE Percent Safe
COMPLETE STOP Wheels of the bus completely stationary at all stops signs and red traffic lights.			
PASSENGER SAFETY ~ 2 SECONDS MOTIONLESS AFTER • Last person loading steps behind yellow line • Last person unloading steps off bus to the right • Last person unloading steps clear of the front left corner of the bus when exiting to the left			
PEDESTRIAN/PASSENGER SAFETY ~ MIRROR CHECK Visually check both SIDE mirrors as you pull out of a loading zone every time the bus stops to load or unload passengers.			
PASSENGER SAFETY ~ BUS STOPPING POSITION • Bus doors shut until completely stopped • Bus positioned so no vehicles may pass on the right			
TIME			

BIBLIOGRAPHY

Alavosius, M. P., & Sulzer-Azaroff, B. (1986). The effects of performance feedback on the safety of client lifting and transfer. Journal of Applied Behavior Analysis, 19, 261-267.

Alavosius, M. P., & Sulzer-Azaroff, B. (1990). Acquisition and maintenance of health-care routines as a function of feedback density. Journal of Applied Behavior Analysis, 23, 152-162.

Austin, J., Alvero, 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 (2), 49-75.

Austin, J., Wellisley, J., & Olson, R. (1998). The behavior engineering model at work on a small scale: Improving customer service with task clarification, employee self-monitoring, and public posting. Manuscript in preparation.

Babcock, R. A., Sulzer-Azaroff, B. & Sanderson, M. (1992). Increasing nurses' use of feedback to promote infection control practices in a head injury treatment center. Journal of Applied Behavior Analysis, 25, 621-627.

Baskett, L. M. (1985). Self-monitoring in children: Accuracy and reactivity. Journal of Genetic Psychology, 146, 107-116.

Berry, T. D., Geller, E. S., Calef, R. S., & Calef, R. A. (1992). Moderating effects of social assistance on verbal interventions to promote safety belt use: an analysis of weak plys. Environment and Behavior, 24 (5), 653-669.

Browder, D. M., Liberty, K., Heller, M., & D'Huyvetters, K. K. (1986). Self-management by teachers: Improving instructional decision making. Professional School Psychology, 1 (3), 165-175.

Burgio, L. D., Whitman, T. L., & Reid, D. H. (1983). A participative management approach for improving direct-care staff performance in an institutional setting. Journal of Applied Behavior Analysis, 16, 37-53.

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

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

Daniels, A. (1989). Performance management (3rd ed.). Tucker, GA: Performance Management Publications.

Dean, M. R., Malott, R. W., & Fulton, B. J. (1983). The effects of self-management training on academic performance. Teaching of Psychology, 10 (2), 77-81.

DeVries, J. E., Burnette, M. M., & Redmon, W. K. (1991). AIDS prevention: Improving nurses' compliance with glove wearing through performance feedback. Journal of Applied Behavior Analysis, 24, 705-711.

DiGangi, S. A., Maag, J. W., & Rutherford, R. B. (1991). Self-graphing of on-task behavior: Enhancing the reactive effects of self-monitoring on on-task behavior and academic performance. Learning Disability Quarterly, 14 (3), 221-230.

Engerman, J. A., Austin, J., & Bailey, J. S. (1997). Prompting patron safety belt use at a supermarket. Journal of Applied Behavior Analysis, 30, 577-579.

Fellner, D. J., & Sulzer-Azaroff, B. (1984). Increasing industrial safety practices and conditions through posted feedback. Journal of Safety Research, 15, 141-151.

Fellner, D. J., & Sulzer-Azaroff, B. (1986). Occupational safety: Assessing the impact of adding assigned or participative goal setting. Journal of Organizational Behavior Management, 7, (½), 3-24.

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. (1988). Managing occupational safety. Blacksburg, VA: Make-A-Difference, Inc.

Geller, E. S. (1996). The psychology of safety: How to improve behaviors and attitudes on the job. Blacksburg, VA: Make-A-difference, Inc.

Hayes, S. C., & Nelson, R. O. (1983). Similar reactivity produced by external cues and self-monitoring. Behavior Modification, 7 (2), 183-196.

Johnston, J. M, & Pennypacker, H. S. (1993). Strategies and tactics of behavioral research. New Jersey: Lawrence Earlbaum and Associates, Inc.

Kanfer, F. H. (1970). Self-monitoring: Methodological limitations and clinical applications. Journal of Consulting and Clinical Psychology, 35, 148-52.

Kessler, D. L. (1985). The effectiveness of different sources of feedback on outcome and behavior in improving the performance of university baseball pitchers. Dissertation Abstracts International, 45 (8-B), 2733-2734.

Kneedler, R. D., & Hallahan, D. P. (1981). Self-monitoring of on-task behavior with learning disabled children: Current studies and directions. Exceptional Education Quarterly, 2 (3), 73-82.

Komaki, J. L., Barwick, K. D., & Scott, L. W. (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. L., Heinzmann, A. T., & Lawson, L. (1980). Effect of training and feedback: Component analysis of a behavioral safety program. Journal of Applied Psychology, 65 (3), 261-270.

Kopp, J. (1988). Self-monitoring: A literature review of research and practice. Social Work Research and Practice, 24, 8-20.

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

Lan, W. Y. (1996). The effects of self-monitoring on students' course performance, use of learning strategies, attitude, self-judgement ability, and knowledge representation. Journal of Experimental Education, 64 (2), 101-115.

Ludwig, T. D., & Geller, E. S. (1991). Improving the driving practices of pizza deliverers: Response generalization and moderating effects of learning history. Journal of Applied Behavior Analysis, 24, 31-44.

Ludwig, T. D., & Geller, E. S. (1997). Assigned versus participative goal setting and response generalization: Managing injury control among professional pizza deliverers. Journal of Applied Behavior Analysis, 82 (2), 253-261.

Malott, R. W., Whaley, D. L., & Malott, M. E. (1997). Elementary principles of behavior. Upper Saddle River, NJ: Prentice Hall.

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, 7-19.

McCann, K. B., & Sulzer-Azaroff, B. (1996). Cumulative trauma disorders: Behavioral injury prevention at work. Journal of Applied Science, 32 (3), 277-291.

McSween, T. E. (1995). The values-based safety process: Improving your safety culture with a behavioral approach. New York: John Wiley & Sons.

Michael, J. L. (1993). Concepts and principles of behavior analysis. Western Michigan University: Society for the Advancement of Behavior Analysis.

Moinat, S. & Snortum, J. R. (1976). Self-management of personal habits by female drug addicts: A feasibility study. Criminal Justice & Behavior, 3 (1), 29-40.

Nasanen, M., & Saari, J. (1987). The effects of positive feedback on housekeeping and accidents at a shipyard. Journal of Occupational Accidents, 8, 237-250.

Petersen, D. (1984). An experiment in positive reinforcement. Professional Safety, 29 (5), 30-35.

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.

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.

Reddell, C. R., Congleton, J. J., Hutchingson, R. D., & Montgomery, J. F. (1992). An evaluation of a weightlifting belt and back injury prevention training class for airline baggage handlers. Applied Ergonomics, 23 (5), 319-329.

Saarela, K. L. (1989). A poster campaign for improving safety on shipyard scaffolds. Journal of Safety Research, 20 (4), 177-185.

Schlinger, H. D. (1993). Separating discriminative from function-altering effects of verbal stimuli. The Behavior Analyst, 16 (1), 9-23.

Schwarz, N. (1999). Self reports: How the questions shape the answers. American Psychologist, 54 (2), 93-105.

Skinner, B. F. (1953). Science and human behavior. New York: Macmillan.

Smith, M. J., Anger, W. K., & Uslan, S. S. (1978). Behavioral modification applied to occupational safety. Journal of Safety Research, 10 (2), 87-88.

Smith, M. J., Cohen, H. H., Cohen, A., & Cleavland, R. J. (1978). Characteristics of successful safety programs. Journal of Safety Research, 10 (1), 87-88.

Srikameswaren, S. M. (1992). Standardized coaching sessions and a self-evaluation package for improving performance of figures by young figure skaters. Dissertation Abstracts International, 53 (1-B), 554.

Stecker, P. M., Whinnery, K.W., & Fuchs, L. S., (1996). Self-recording during unsupervised academic activity: Effects on time spent out of class. Exceptionality, 6 (3), 133-147.

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

Sulzer-Azaroff, B., & Consuelo De Santamaria, M. (1980). Industrial safety hazard reduction through performance feedback. Journal of Applied Behavior Analysis, 13, 287-295.

Troy, A. (1983). Improving customer service behaviors at a restaurant. Unpublished masters thesis, Western Michigan University, Kalamazoo, MI.

Whelan, J., Mahoney, M. J., & Meyers, A. W. (1991). Performance enhancement in sport: A cognitive behavioral domain. Behavior Therapy, 22 (3), 307-327.

Wilk, L. A., & Redmon, W. K. (1990). A daily-adjusted goal-setting and feedback procedure for improving productivity in a university admission department. Journal of Organizational Behavior Management, 11, 55-75.