Managing Distribution Quality Through an Adapted Incentive Program with Tiered Goals and Feedback

Melodie J. Bateman & Timothy D. Ludwig

ABSTRACT

The purpose of this study was to evaluate a program designed to improve the work quality of employees at a food distribution warehouse. Participants (n = 23) were responsible for selecting food items to be distributed to restaurants. Selector errors were targeted using an adapted incentive program featuring tiered goals and individualized normative feedback. A pre-existing incentive program, wherein selectors forfeited money for each error, was adapted to allow the selectors the opportunity to earn back their disincentive money upon attainment of weekly goals that were increasingly demanding. The initial goal for the first two weeks was set at 2.0 errors per 1000 cases and every two weeks thereafter the new goal was reduced by 0.1 errors per 1000 cases until the final goal of 1.5 errors per 1000 cases. Weekly graphic feedback included public posting of individual errors rates. The mean errors per 1000 cases for the experimental group decreased 10% from 3.16 during baseline to 2.54 during the intervention. The largest decreases in errors was observed in the mid-level performers (with baseline error rates between 2 and 3.5 errors per 1000 cases). A cost-benefit analysis revealed potential savings of $9,799.50.

ARTICLE

Incentives have been shown to be more effective than hourly pay in influencing employee performance because they provide monetary reinforcement contingent on the worker's performance (Abernathy, Duffy,&O'Brien, 1982; Gaetani, Hoxeng,&Austin, 1985; Orpen, 1982). Incentive program contingencies differentiate them from other types of pay-for-performance compensation systems, such as profit sharing, and gainsharing. Incentive programs provide reinforcement closer in time to the targeted behavior because the monetary reward can be distributed in weekly paychecks. The reinforcer is also probable because employees are guaranteed incentive pay contingent on individual performance instead of on group or organizational performance. Although individual incentives may be considered small, they are available each and every time a unit is produced. Therefore, this type of compensation system is a popular technique because it sets up the conditions needed to encourage the desired behaviors that lead to increased performance (Redmon & Agnew, 1991).

Whereas an incentive program provides reinforcement for desirable performance, a disincentive program includes a response cost of money loss for undesirable performance, such as errors (Abernathy et al., 1982). Disincentives can be analogous to punishers or negative reinforcers depending on the targeted response. When disincentives are used in attempts to decrease an undesirable response, they are analogous to punishers. If the targeted behavior is the desirable response, dis-

incentives are analogous to negative reinforcers. In this case, the employee would exhibit the desired response in order to avoid the negative consequence of losing money.

Often incentives and disincentives are paired together. Gaetani et al. (1985), for example, provided billing feedback to auto mechanics, followed by feedback plus incentives (i.e., 5% of the money billed to customers over a goal) or disincentives (i.e., pay reduced by a percentage when below the goal) depending on their performance. The results showed that performance increased when feedback was provided, but increased more during the feedback and monetary incentive/disincentive program.

Likewise, Abernathy et al. (1982) implemented a feedback intervention, followed by a feedback and incentive/disincentive program intervention among proof operators in a bank. Goals were established for incentive pay and monetary penalties were assessed for avoidable errors. This incentive/disincentive program was associated with an increase of over 150% in proof operators' performance and remained at the same level for over 2 years.

These studies demonstrate that incentive/disincentive programs can be effective in increasing desired performance immediately, as well as over the long-term. No studies, to our knowledge, have investigated the repayment of the disincentive component contingent on the decrease of undesirable performance or the increase of desirable behavior. In this situation employees lose money for incorrect behaviors and then would

have the opportunity to earn back disincentive money when corrected

behaviors meet a certain goal.

The present study examined the effects of an adapted incentive program

that included the repayment of disincentive money when the undesired

performance was below a goal. Instead of simply having money

taken away for errors, employees were allowed the opportunity to earn

back their money when their number of errors occurred at low rates. A

goal of low rates was used because although error-free performance

was desired, it was next to impossible to achieve.

VARIATIONS OF FEEDBACK
WITHIN INCENTIVE PROGRAMS

By their very nature, incentive programs are paired with feedback.

Employees are informed of their performance levels when they receive

or don't receive incentive pay contingent on performance. Feedback becomes

a discriminative stimulus for the reinforcement when paired with

reinforcers such as incentive pay (Daniels, 1989). A discriminative

stimulus indicates that a response is likely to be followed by a reinforcer

(Sulzer-Azaroff & Mayer, 1991). Additionally, when reinforcement is

attained for increases in performance, the feedback delivered with the

reinforcer becomes a conditioned reinforcer (Sulzer-Azaroff & Mayer,

1991). Feedback paired with tangible reinforcement has been shown to

be more effective in increasing performance than feedback alone

(Balcazar, Hopkins,&Suarez, 1986; Alvero, Bucklin,&Austin, 2001).

This evidence supports the use of feedback in combination with an incentive program.

Typically, the feedback provided in incentive programs is presented as single data points (e.g., Honeywell, Dickinson, & Poling, 1997) or as individual graphs of data over time (e.g., LaMere, Dickinson, Henry, Henry, & Poling, 1996). The current study presented individual employee data represented on a publicly displayed graph so that employees could see their individual performance, as well as that of their peers. Although peer performance did not affect individual incentive pay, individuals were able to compare themselves to the rest of the group to determine if they were performing above or below average.

This type of "normative" feedback has been shown to be effective when paired with goals. Ludwig and Geller (2000) reported a study that posted individual drivers' names on graphs next to their corresponding data points. The results demonstrated that this type of normative feedback is more effective than averaged group feedback. In another study, Ludwig, Biggs, Wagner, and Geller (2001) paired normative feedback with an individual competition among pizza deliverers and demonstrated that even deliverers who did not win an award for their safe driving still changed their behavior during the program. However, there have been no individual monetary incentive program studies that allowed employees to see their peers' performance and incentives gained.

VARIATIONS OF GOALS WITHIN INCENTIVE PROGRAMS

Goals are also requisite components of incentive programs because they specify the level of performance required to receive the incentive pay (Abernathy et al., 1982). When setting goals, it is important to ensure that they are difficult, yet attainable (Locke&Latham, 1990). If goals are too difficult to reach, they will not be effective in producing changes in performance. One potentially effective method of goal setting is to set goals that gradually increase in difficulty. This technique of gradually changing the goal allows more participants to sample the reinforcer available for goal attainment (Kazdin, 1994). If participants are given the opportunity to sample the reinforcer, they may be more likely to engage in the desired performance and more likely to reach subsequent goals.

Goals can be adapted over time in an incentive program in order to provide the participants with an initial opportunity to sample the reinforcer before becoming more stringent to shape acceptable levels of performance. For example, Ludwig and Geller (2000) described a study that provided some participants with goals set near their baseline rate. Then goals were raised by an incremental increase each week until reaching a final goal (i.e., "tiering"). Likewise, Wilk and Redmon (1990) adjusted productivity goals daily for university admissions employees. This intervention not only increased employee productivity but also decreased absenteeism. The present study used the tiered goal technique whereby workers were assigned a new goal every two weeks that increased in difficulty over time.

STATEMENT OF THE PROBLEM

The present study was initiated at the request of management in a food distribution warehouse to improve the quality of performance (i.e., decreasing errors) among selector employees. The company wanted to decrease errors because they were costly both immediately while correcting the mistake, and in the future through potential lost customers. This study focused on two types of errors: Shorts and mispicks. Before the present study, the food distributor already had an incentive program in place. It consisted of incentive pay for productivity and disincentives for errors. This study examined the effectiveness of an adapted incentive program that allowed employees the opportunity to earn back disincentive money if they reached the goal. In addition, the adapted incentive program included tiered goals and individualized normative feedback. The quality of selectors' performance was expected to improve substantially following the introduction of the adapted incentive program.

METHOD

Participants and Setting

The study was conducted at a foodservice distributor in the southeastern U.S. that caters to restaurants, schools, and other businesses. This company purchases and receives cases of food items from manufacturers

and stores these items in its warehouse. Orders are received

from customers and filled by selectors who pick food items and assemble

a delivery pallet. After the orders are prepared, delivery drivers

transport the food items to the customers.

The participants at the experimental site were selectors (n = 23), responsible

for assembling food orders to be distributed. The average age

of the selectors was 23 (range: 19 to 35), they were all males, and, on average,

were employed less than one year on the job (range 9 months to

22 months). Participants with less than one year of experience earned

$10.90 per hour plus incentive pay. Those who worked for more than

one year but less than two years earned $12.45 per hour plus incentive

pay. Shifts were eight hours long but an employee could leave when the

designated orders were completed. All selectors worked in the same

area of the warehouse and performed the same duties.

A warehouse located in a neighboring state and owned by the same

foodservice distributor served as a comparison site. The selectors (n =

14) included in the comparison group were all males with an average

age of 22 (range: 20 to 31). On average, these selectors had 5 months of

experience (range: 1 to 13 months). The comparison selectors earned

the same pay and worked the same number of hours as the selectors in

the experimental warehouse.

The duties of a selector consist of selecting food items for distribution.

Selectors first receive an order list for as many as three customers.

Referring to this list of food items, selectors drive pallet jacks in the

warehouse to the designated locations of the food items. Cases of food items from the order list are then stacked onto pallets on a pallet jack in order of the list. Items are listed in a particular order based on weight of the item and location of the item in the warehouse. Items that are adjacent to one another in the warehouse are listed together. When all items have been collected, the selector transports the pallets to the loading docks. When an order is complete, selectors begin picking from another order list. Order lists are divided among and assigned to 5-7-member teams of selectors. When all of the items on the order lists of the team for a particular day have been selected, the team members may leave. There are, on average, 60 cases per order. Selectors typically pick 20-24 orders per day.

Dependent Variable

Many errors can occur throughout the process of distribution. Errors in the form of shorts and mispicks can occur when a selector fills an order. A "short" is defined as a desired product being absent from the filled order. A "mispick" is defined as an incorrect item being placed within the filled order.

The shorts and mispicks data were recorded when delivery drivers unloaded the order at the customer site. After unloading the food items, the driver inspected the delivery and recorded whether there was a short or mispick on the invoice. If the mistake was determined after the driver left, the customer contacted a sales representative who recorded the error.

Errors were then recorded into the company database for each order. An order number was tied to the selector who filled the particular order. In this manner individual selectors' error rates were tracked across the study.

Baseline

The study included a changing criterion design with a non-treatment comparison group. Baseline data were collected for a total of eight weeks. During the Baseline period the company maintained its normal disincentive program. In this program, selectors were given feedback regarding their combined shorts and mispicks per 1000 cases. They received access to this feedback on computers located at their workstations and on a sheet attached to their weekly paychecks. Measures were also taken for case volume, the total number of cases per week. Feedback on case volume was also available on the company computers and on their weekly paychecks during the entire study.

Selectors received incentive money for productivity and lost money for their errors. The incentive pay for productivity was 1.6 cents per case selected. As a disincentive, selectors had deducted from their earned incentive money $.50 for every short and $1.50 for every mispick. On average, selectors were receiving $67.20 for productivity and losing $13.29 for errors per week.

Feedback, Tiered Goals, and Adapted Incentive Program

The intervention consisted of feedback, a changing criterion of tiered

goals, and an adapted incentive program. It continued for a total of 13

weeks. The selectors were informed of the intervention in a staff meeting

at the beginning of the intervention. The Warehouse Manager informed

the selectors that they would receive weekly graphic feedback

of their errors and that there would be a new goal for every two weeks

until reaching a final goal of 1.5 errors per 1000 cases. They were in-

formed that every time they reached the week's goal, they would earn

back all of their disincentive money for that week.

Feedback. Individualized graphic feedback was publicly posted to

show each selector's average combined number of shorts and mispicks

per 1000 cases. Individual data for all selectors and the weekly goal

were represented on the graphs. Data points represented the average

number of errors per 1000 cases per individual per week. The names of

the selectors were written next to their corresponding data points each

week. The weekly goal was represented by a horizontal line. These

graphs were updated weekly. Data for a particular week (Sunday

through Thursday) were displayed in a graph on the following Wednesday.

Each graph displayed the data for the prior two weeks. Please see

Figure 1 for a facsimile of the feedback graph.

Goals. The intervention also consisted of tiered goals set by the management

team to be achieved by each selector. An ultimate goal of 1.5

errors per 1000 cases was to be achieved by the 11th week. The initial

goal for the first two weeks was set at 2.0 errors per 1000 cases. Every

two weeks thereafter the new goal was reduced by 0.1 errors per 1000

cases. These weekly goals were represented on the weekly graph of selector

errors by a horizontal line (see Figure 1). By comparing their

feedback with the goal level, selectors were able to determine if they

had reached the goal for the week.

Earned-Back Disincentive Pay. At the end of a week, if the selectors'

average number of errors per 1000 cases was at or below the current

goal, that selector avoided losing $.50 for every short and $1.50 for every

mispick for that week. For example, if a selector's average number

of errors per 1000 cases was 1.9 for a week and this was below a goal of

2.0 errors per 1000 cases, he would not lose the money that the company

would normally charge. More specifically, if a selector reached the goal

and yet he still had 11 errors for the week, 6 mispicks and 5 shorts, he

would not lose the $11.50 from his incentive money for those errors in

his paycheck on the following Friday.

Manipulation Check and Social Validity Questionnaire

Two weeks after the program ended, selectors completed a short

questionnaire to determine the extent to which they were aware of the

program. Employees were asked to complete the questionnaire during

their daily meeting at the beginning of the shift. Items assessed whether

employees understood the program, had access to the posted feedback

FIGURE 1. Example posted graphic feedback. The shaded horizontal line represents the weekly goal.

Your Error Rate.
Did You Earn Your Disincentive Back?
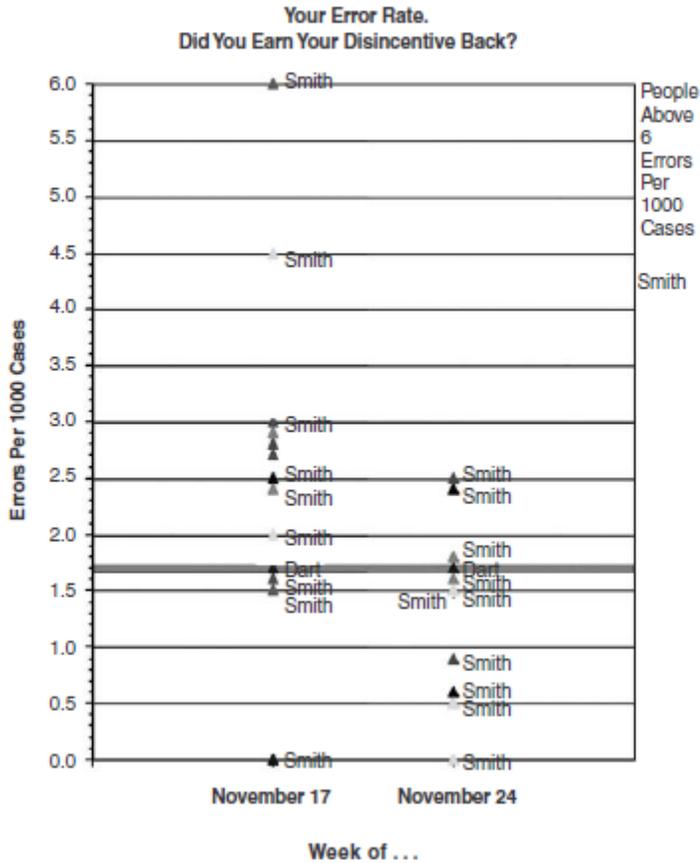
graphs, and whether each knew his error rate, the goals for each week, and the amount of money he had earned back during the program.

The questionnaire administered to the selectors also assessed social validity. The employees were asked whether they thought the program was fair, whether they were satisfied with their jobs, and if they wanted the program to continue.

RESULTS

Errors were measured on a weekly basis for 8 weeks during baseline
and 13 weeks during the intervention phase. To be included in the data
analysis, selectors must have worked a total of 18 out of the 21 weeks of
observation. Twelve selectors were eliminated using these criteria. This
left 23 selectors at the experimental site and 14 selectors at the comparison
site for the analysis in this study. Management at the comparison
group did not make individual-level data available to the investigators.

Manipulation Check

A questionnaire was administered to the selectors at the experimental
site to assess the integrity of the intervention. Forty-eight percent of the
workers who participated in the study filled out the questionnaire. One
hundred percent of these respondents attended the initial meeting that
introduced the program and reported being aware of the program. Most
of the respondents (91%) remembered when the program began. Sixtyfour
percent were aware of the first goal of 2.0 errors per 1000 cases
while 82% knew the final goal of 1.5 errors per 1000 cases. All respondents
reported that they had knowledge of their individual error rate
during the program. Eighty-two percent looked at the feedback graph at
least occasionally. Ninety-one percent reported comparing their error
rate to the weekly goal and 91% reported comparing their error rate to
the rates of others.

Error Rate

The selector error rates for the experimental group across the 21-week project are represented in Figure 2. The changing criterion for the experimental group is represented in two-week intervals by shortened vertical lines in Figure 2. The published goal for each two-week interval also appears in Figure 2.

The mean number of errors per 1000 cases for the experimental group during baseline was 3.16 (SD = .26; range: 2.26 to 3.41). This decreased to a mean of 2.54 (SD = .48; range: 1.83 to 3.40) errors per 1000 cases during the intervention phase. The start of the intervention was marked by a decrease in error rate at the experimental site during the first week followed by an immediate increase back to baseline levels. After this initial temporary impact there was a downward trend during the remainder of the intervention. All but 2 of the weekly error ratios during the 13-week intervention phase fell below the mean baseline level. Overall there were more mispicks per 1000 cases than shorts, however, the trends for both mispicks and shorts were very similar.

The error rate for the week of December 2nd (see circle "A" on Figure 2) was considerably higher than results from other weeks. This was due to a significant internal mishap that was out of the selectors' control. If the data point for the week of December 2nd is eliminated, error rate data for the week of October 28 through the final week suggests a notable downward trend consistent with the changing goal criterion.

The selector error rates for the comparison group are represented in

Figure 3. The error rate for the comparison group is markedly higher

and more variable than the error rate for the experimental group. The

mean number of errors per 1000 cases for the comparison group was

6.49 (SD = 1.14; range: 5.46 to 8.10) during the baseline phase at the experimental

site. During the intervention at the experimental site, the

mean error rate at the comparison site was 5.62 (SD = 1.0; range: 4.23 to

6.78). A downward trend similar to the experimental group was not observed

in the error rate of the comparison group (see Figure 3).

FIGURE 2. Mean number of errors (shorts and mispicks) per 1000 cases per week for selectors at the experimental site. The means during baseline and interventions are represented by dashed horizontal lines. The criteria are represented by dashed vertical lines with the errors per 1000 cases goal listed.
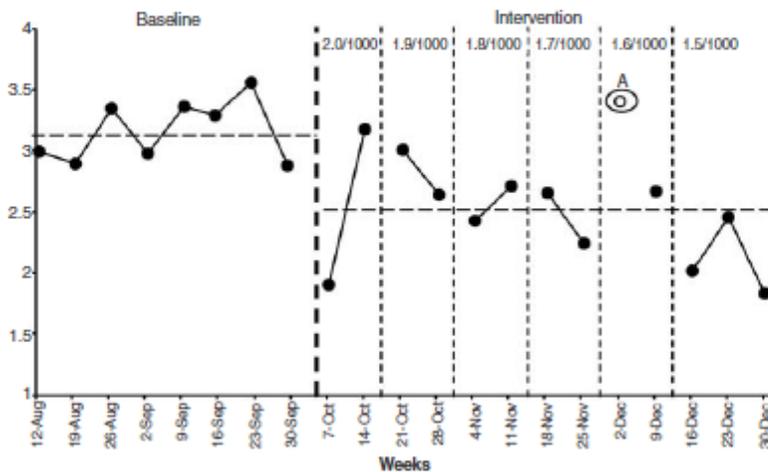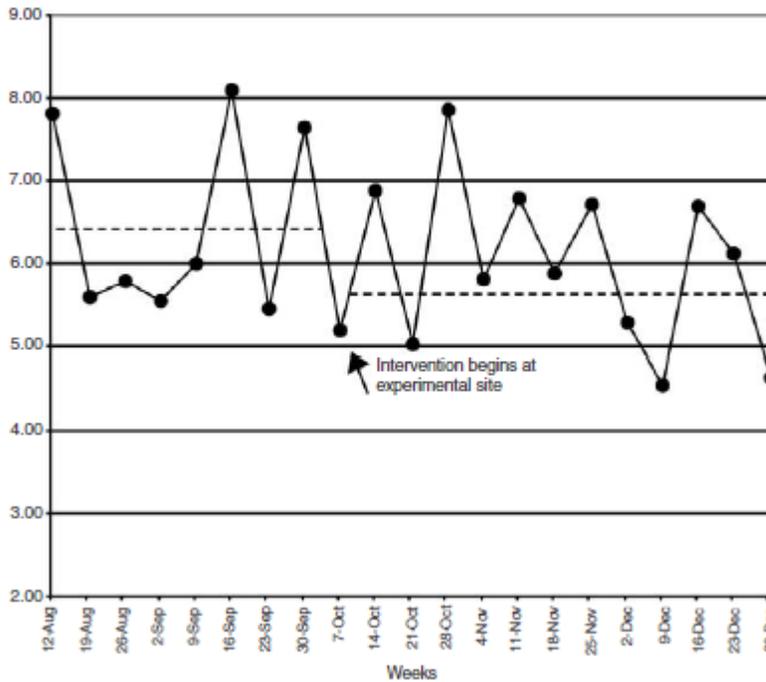
FIGURE 3. Mean number of errors (shorts and mispicks) per 1000 cases per week for selectors at the comparison site. The means for the comparison site during the experimental site's baseline and intervention are represented by dashed horizontal lines.

Percent of Selectors Achieving the Goal

The final goal during the intervention was 1.5 errors per 1000 cases.

Figure 4 displays the percentage of selectors who achieved the goal of 1.5 errors per 1000 cases per week. The mean percentage of selectors below the goal was 36.4% during baseline, and 42.0% during the intervention. During the last three weeks of the intervention, over 47% of the selectors were below this goal.

Error Rate by Baseline Responding

For further analysis, selectors in the experimental group were divided into groups based on their average error rates during baseline. Ten

FIGURE 4. Percentage of selectors below 1.5 errors per 1000 cases per week at the experimental site.
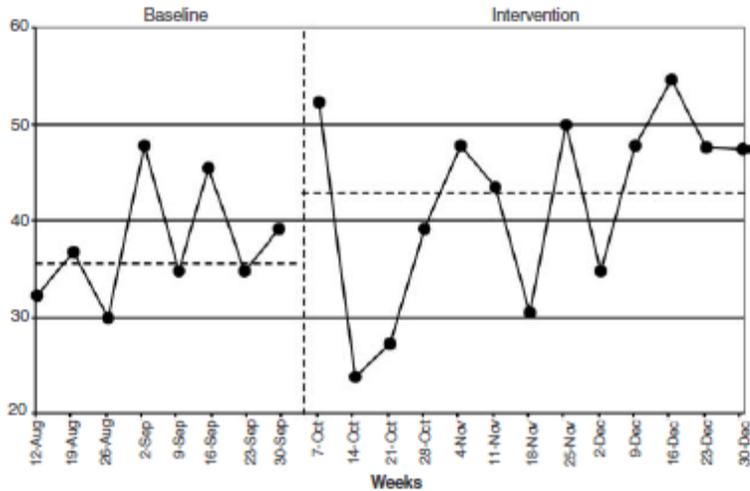

selectors whose error rate was below 2.00 errors per 1000 cases are represented as "Group 1." Seven selectors whose average error rates were between 2.00 and 3.50 errors per 1000 cases are represented as "Group 2." Finally, six selectors with average error rates above 3.5 errors per 1000 cases are represented as "Group 3." The data for all three groups are represented in Figure 5.

The average weekly error rate for selectors in Group 1 was 1.34 errors per 1000 cases (SD = .41; range: 0.60 to 1.87) during baseline. The average error rate during the intervention was 1.34 errors per 1000 cases (SD = .54; range: 0.73 to 2.51), which was equal to the average error rate during baseline.

The average weekly error rate for Group 2 decreased from 2.82 (SD = .57; range: 2.16 to 3.97) during baseline to 2.21 (SD = .84; range: .68 to 4.09) during the intervention. A large increase in error rate occurred during the week of December 2nd in the intervention. After December

2nd the error rate steadily decreased to a low of 0.68 errors per 1000 cases. All but one of the data points during intervention fell below the baseline mean. An example selector from this group appears in Figure 6. SelectorAhad an error rate in baseline just above the initial goal level of 2.0 errors per 1000 cases. He earned back his disincentive during the

FIGURE 5. Mean number of errors (shorts and mispicks) per 1000 cases for selectors by group. Group 1 included selectors with baseline means below 2.0, Group 2 included those with means between 2.0 and 3.5, and Group 3 included those with means above 3.5 errors per 1000 cases.
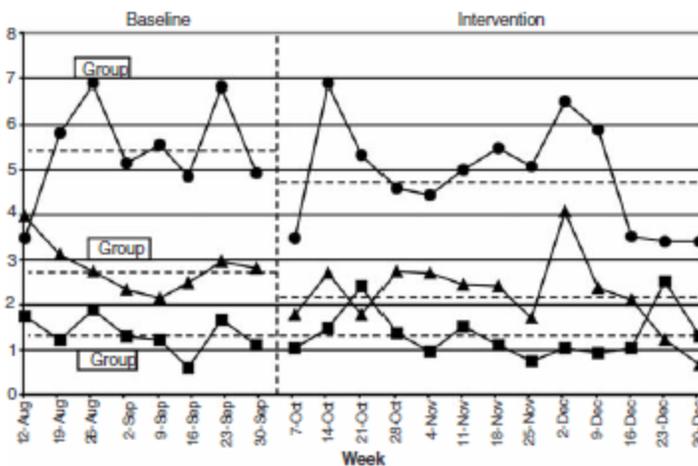

first week of the intervention and another 7 times over the intervention phase (highlighted by arrows in Figure 6). There is also a notable downward trend in his error rate over the course of the intervention phase.

The average error rate for Group 3 during baseline was 5.44 (SD = 1.13; range: 3.48 to 6.88). Mean selector error rate during the intervention phase was 4.85 (SD = 1.18; range: 3.39 to 6.90). A noticeable decrease occurred during the first week, followed by an increase back to baseline levels. The final three weeks remained below 3.5 errors per 1000 cases. An example selector from this group appears in Figure 7. Selector B had an exceptionally high error rate in baseline that decreased

substantially during the intervention. However, Selector B was never able to reduce his error rate below the goal level (2.0-1.5 errors per 1000 cases).

One selector, initially categorized in Group 1, demonstrated a contraindicative pattern of responding (see Figure 8). Selector C's baseline error rate was within the future goal level. In fact, he earned back his disincentive four times during the intervention (highlighted by arrows). However, after a month of the intervention contingencies his error rate trended upward, eventually increasing to twice his baseline level.

FIGURE 6. Mean number of errors (shorts and mispicks) per 1000 cases per week for Selector A. Arrows represent weeks where Selector A earned back his disincentive money.
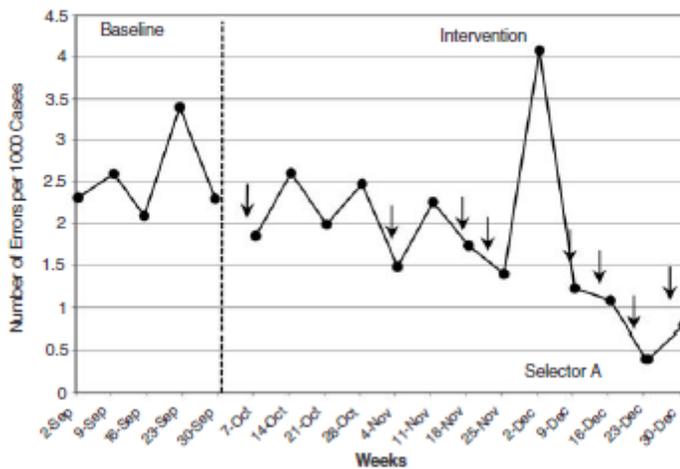
Social Validity

Social validity was assessed through a questionnaire administered two weeks after the study. When asked about the fairness of the program, 73% thought the program was "fair" or "very fair." The remaining 27% responded "neutral". Seventy-three percent reported being "satisfied" or

"very satisfied" with their jobs during the program. No one stated that they were "dissatisfied" during this time period. Ninety-one percent of the respondents believed that the program "should continue."

Cost Benefit Analysis

A cost benefit analysis was conducted to determine the value of the program to the company. There were 4141 total errors during baseline. If the errors during the intervention had continued steadily at the same rate as the rate during baseline, the total cumulative errors would have been 10,914. The actual total cumulative number of errors by the end of the intervention was 9,532. This results in a difference between projected cumulative errors and actual cumulative errors of 1,382. Each error was estimated to cost the company $7.50 by the manager of the

FIGURE 7. Mean number of errors (shorts and mispicks) per 1000 cases per week for Selector B.
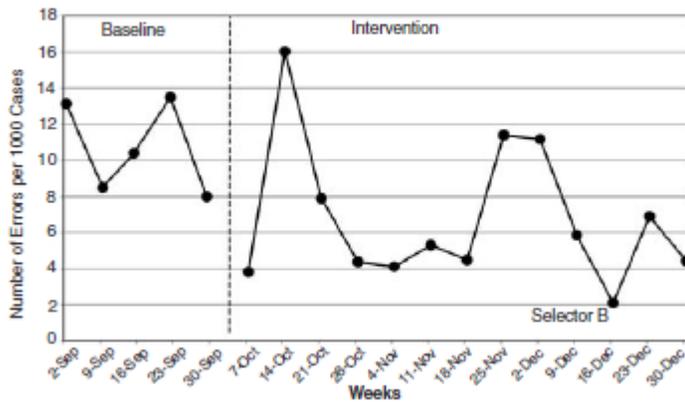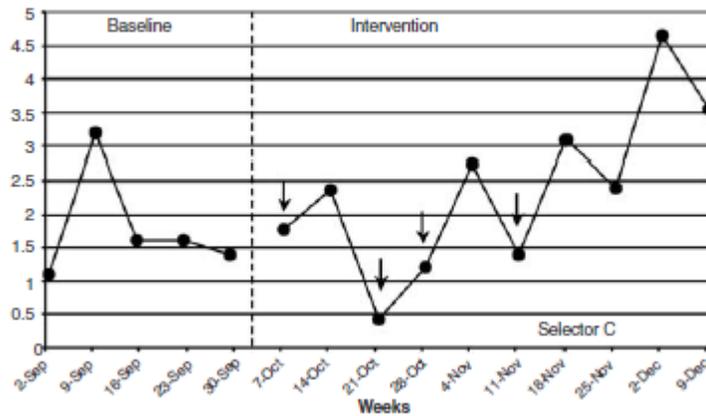

FIGURE 8. Mean number of errors (shorts and mispicks) per 1000 cases per week for Selector C. Arrows represent weeks where Selector C earned back his disincentive money.

warehouse based on benchmarks from other similar food distribution companies. The savings experienced from the avoided 1,382 errors during the 13-week program was therefore estimated to be $10,365.

The total amount that would have been collected from the selectors via disincentives during the intervention was $5,736.50. The amount of money that was given back to selectors when they met the weekly goal totaled $565.50. Thus, an average of $24 was returned to employees each week during this program. More accurately, there were 13 selectors who received returned money, these individuals received an average

of $43.50 per week during this program. After subtracting the total

amount repaid to selectors (the only costs incurred by the company in

this study), the amount saved to the company is estimated to be

$9,799.50.

DISCUSSION

The goal of this study was to evaluate the effects of an adapted incentive

program with tiered goals and normative feedback on the error rate

of selectors at a distribution warehouse. The intervention was associated

with a decrease in the selector error rate for the experimental group.

The start of the intervention was marked by a considerable decrease in

error rate at the experimental site. This may be the result of management's

strong initial support and communication of the program. The

second week of the intervention was marked by a large increase in error

rate back to baseline levels. A continuous downward trend was then observed

during the remaining 11 weeks whereby most of the data points

fell below the baseline mean.

The week of December 2nd was marked by a considerable increase

in error rate for the experimental group. According to management, numerous

cases of a popular item were labeled incorrectly when the shipment

was received during that week. Selectors picking that item in their

order inadvertently got the wrong case(s) or otherwise skipped the

case(s) causing a significant number of mispicks and shorts until the situation

was resolved later that week. If the data point for the week of December

2nd is eliminated, the remaining data show a notable downward

trend in the experimental group's error rate during the fourth week through the final week with all points below the baseline mean.

Such a downward trend may be expected when intervening with tiered goals within the changing criterion design of the study. Tiered goals are set at a beginning level suspected to be achievable by most participants. Yet initial behavioral change would be limited because of the lenient goal. However, as time goes by the goal becomes more and more difficult to achieve (i.e., the criterion changed) and the amount of behavior change required is greater. Thus, one would expect a trend in the same direction as the trend in the tiered goals.

Sampling the Contingency

Another method of demonstrating the effect of the intervention was to examine the percentage of selectors who achieved the goal on a weekly basis. The mean percentage of selectors achieving the goal increased during the intervention. In addition, there was an upward trend of selectors achieving the goal during the intervention. It is noteworthy that nearly half of the selectors were below the goal during the last three weeks.

Those individuals who have sampled the contingency by achieving the lenient goal early may be more likely to increase their efforts as the goals become more difficult. This was demonstrated by Selector A whose error rate during baseline was just above the goal level. He achieved the goal initially and continued to decrease his error rate as the goal level became more stringent. Other selectors whose baseline was

close to the goal also seemed to be in the best position to benefit from the intervention (i.e., Group 2). All of this group's weekly data points except one (December 2nd) during the intervention fell below the baseline mean. It is evident that the intervention had the most significant effect on error rate for this subsample of selectors.

In contrast, those individuals whose behaviors did not approximate the goal were not able to sample the contingency. These individuals with error rates considerably higher than the goal may not be able to change their behavior enough to "catch" up with the increasingly difficult goals. A good example of this was Selector B who decreased his error rate by 50% by the end of the intervention yet never achieved the goal or earned back his disincentive. Other selectors who had very high error rates during baseline decreased their error rate considerably, however it was still significantly higher than the goal. Like Selector B, these individuals were performing at a level that was so far from the goal, that it was difficult for them to change their behavior enough to reach the goal and sample the reinforcer.

This finding supports the suggestion by Ludwig et al. (2001) that it may not be necessary to sample the reinforcer in order to change behavior. Instead, the goal setting paired with the chance to earn back money may have resulted in rule-governed behavior (Blakely & Schlinger, 1987) by the selectors.

Another possible explanation for changes in behavior in those selectors who never earned back their disincentives may be the social contingencies presumably generated by other intervention operations (i.e.,

normative feedback and error rate goals). Social contingencies are

thought to be strongest when (a) others expect designated behaviors (as

implied by the management-mandated goals), (b) have opportunities to

observe whether the behaviors occur (via the feedback graphs with

names), and (c) when peers or superiors are willing to promote these behaviors

consistently with approving (rewarding) or disapproving (punishing)

consequences (Anderson, Chiricos, & Waldo, 1977; Sussman,

1989). Although the current research did not assess these variables directly,

we agree with Poling, Smith, and Braatz (1993) who called for

more systematic measures of social interaction resulting from an intervention

process (e.g., monitoring the verbal behavior of selectors interacting

with one another or with management).

Of course, those individuals already achieving the goal rate during

baseline would have no need to improve their behaviors since the current

rate would be reinforced. Indeed, those individuals whose baseline

behavior was already below the goal levels may not have been affected

by the intervention because they were already performing at the desired

levels and no further reinforcement was offered for lower rates of errors.

If additional goals below 1.5 errors per 1000 cases were presented

along with additional incentives the performance of these individuals

may continue to improve. Alternatively, the lack of significant improvement

in error rate may have been a function of a floor effect.

These selectors were already performing at a level close to zero, and

therefore may not have been able to reduce their error rate significantly.

It is notable that one individual increased his error rate substantially

during the intervention. Specifically, Selector C showed an increasing

trend in error rate to a level twice his baseline mean. This is similar to

examples reported in Ludwig and Geller (1991) where one pizza deliverer

decreased his safety-belt use from around 100% during baseline to

0% at one point during the intervention phase and another reported in

Ludwig and Geller (2000) who decreased his turn signal use from a

baseline mean of 83% to a low of 25% toward the end of the intervention.

Seemingly, a small number of individuals will respond negatively

to these types of intervention operations. These instances may be similar

to what Skinner (1971) referred to as countercontrol.

Social Validity and Cost-Benefit

The social validity questionnaire findings suggest that the adapted incentive

program was well received by the employees. Because most respondents

thought the program was fair, they were satisfied with their

jobs, and they wanted the program to continue, it seems that no harmful

effects were experienced. The program not only benefited management,

it also benefited the employees. On average, selectors who were successful

in earning back money saved an average of $43.50 during the

course of this study.

Another measure of social validity specific to business is the effect

on profitability. Even though the intervention produced a moderate effect

on selector error rate and employees approved of the program, the

company must save money in order for it to be considered by management

to be successful (i.e., practical significance). The cost-benefit

analysis suggests that the 13-week program saved the company money.

The estimated savings to the company was $9,799.50 for the entire 13-week period. If the host company would have kept this program in place for a year they may have realized an annualized savings of approximately $40,000. If the selectors maintained the same error rate as they accomplished in the last 4 weeks of the study, the estimated savings may indeed be greater.

Limitations and Suggestions for Future Studies

One limitation of the study was the lack of individual data for the selectors at the comparison site. The authors were unable to have access to individual data for the selectors in the comparison group. Inferential statistical analyses such as mixed design ANOVAs would have been possible if these data were available. In addition, the comparison group was not ideal for comparison with the experimental group. Even though the warehouse was similar and owned by the same company, the number of errors was considerably higher and more variable at the comparison site.

A longer baseline period would have also been preferable, however, another program targeting a number of performance factors was in place prior to August. A longer baseline was not used because the authors did not want the baseline to be contaminated by the previous program.

Additionally, the study would have been more rigorous if a return to baseline followed the intervention. Although the disincentive payback program was ceased at the end of the study, a true return to baseline was

impossible due to changes that management implemented immediately following the adapted incentive program. One of these changes included the change from 8 to 10-hour shifts. Such a change is likely to cause increases in error due to factors like fatigue, suggesting that the results of the study are somewhat conservative. Because of this and other changes in policy and staffing at the experimental site, management decided not to continue the intervention in the absence of the research team. Future researchers should consider designing a "transition plan" or other methods for indigenous intervention management to better ensure maintenance of successful intervention programs.

A factor that may have affected the reported error rates was a possible seasonal variation experienced at the distribution warehouses. The program occurred during two holidays that could have contributed to less demand. This makes sense because some of the customers of the warehouse include schools, which are not in session during the holiday season. When there is less demand, each selector receives a smaller caseload for the day. When this happens, selectors have the ability to work slower and consequently may commit fewer errors. However, selectors are allowed to leave when their work is finished and are paid for 8-hours of work regardless of when they finish. The ability to leave early may be more reinforcing and therefore counteract the tendency to slow down.

A suggestion for improving the program is to use an initial goal that is less difficult to achieve. This study used tiered goals so that employees would initially be able to sample the reinforcer. However, the initial

goal of 2.0 errors per 1000 cases may have been too difficult for the majority

of selectors to be able to attain it. It may be beneficial to set a more

lenient initial goal (e.g., 3.0 errors per 1000 cases) and decrease it by

larger increments than 0.1 errors per 1000 cases. This study was not

able to implement a lenient goal due to management's disapproval of

rewarding employees at a level higher than 2.0. Another suggestion is to

use individually-set goals. Rather than utilizing one company-wide set

of goals, management could set goals for each selector based on their

previous individual performance.

The data demonstrate that there were some selectors whose error

rates were extremely high (i.e., a baseline mean of 5.44 errors per 1000

cases). Clearly this level is unacceptable to management. A suggestion

is to provide training for the selectors with high error rates. It is possible

that these selectors were unaware of how to avoid errors. If employees

lack the skills, reinforcement will not contribute to improvements in

their performance.

A final suggestion for improving the intervention would be to use incentives

rather than disincentives to increase quality. When negative reinforcers,

such as disincentives, are used, employees may only do the

minimum to get by (Daniels, 1989). (We view disincentives in this context

as reinforcers because most of the errors in this study's context

were due to the lack of correct behaviors. A mispick is not a behavior;

instead it is the absence of a correct behavior. Therefore, the [dis]incentives

are considered to increase correct behavior). In addition, negative

reinforcement is associated with many negative side effects, such as aggression

and fear (Martin & Pear, 1996) or even countercontrol (Ludwig & Geller, 2000). Therefore, it may be more beneficial to use positive reinforcement, such as dropping the disincentive system altogether and focus on the use of incentives, to improve the quality of selectors' performance.

This study was unique in that it was the first to examine earned-back disincentive pay as part of an incentive program. In addition, it examined both normative feedback and tiered (i.e., changing) goals in combination with the adapted incentive program. Both normative feedback and tiered goals have been studied in the literature, but not in combination with an incentive program. The results of the present study support the previous findings that these independent variables produce significant improvements in performance. In addition, the present study supports the use of programs that utilize the repayment of disincentives to improve quality. Future research should continue to examine the effects of the repayment of disincentives on quality, as well as on other dependent variables.

REFERENCES

Abernathy, W. B., Duffy, E. M., & O'Brien, R. M. (1982). Multi-branch, multi-systems programs in banking: An organization-wide intervention. In R. M. O'Brien, A. M. Dickinson, & M.P. Rosow (Eds.), Industrial behavior modification: A management handbook (pp. 370-382). New York: Pergamon.

Alvero, A., Bucklin, B., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998). Journal of Organizational Behavior Management, 21(1), 3-30.

Anderson, L. S., Chiricos, T. G., & Waldo, G. P. (1977). Formal and informal sanctions: A comparison of deterrent effects. Social Problems, 25, 103-114.

Balcazar, F., Hopkins, B. L., & Suarez, Y. (1986). A critical, objective review of performance feedback. Journal of Organizational Behavior Management, 7(3-4), 65-89.

Blakely, E., & Schlinger, H. (1987). Rules: Function-altering contingency-specifying stimuli. The Behavior Analyst, 10, 183-187.

Daniels, A. C. (1989). Performance management: Improving quality productivity through positive reinforcement. Tucker, GA: Performance Management Publications.

Gaetani, J. J., Hoxeng, D. D., & Austin, J. T. (1985). Engineering compensation systems: Effects of commissioned versus wage payment. Journal of Organizational Behavior Management, 7 (1-2), 51-63.

Honeywell, J. A., Dickinson, A. M., & Poling, A. (1997). Individual performance as a function of individual and group pay contingencies. The Psychological Record, 47, 261-274.

Kazdin, A. E. (1994). Behavior modification in applied settings. Pacific Grove, CA: Brooks/Cole Publishing.

LaMere, J. M., Dickinson, A. M., Henry, M., Henry, G.,&Poling, A. (1996). Effects of a multicomponent monetary incentive program on the performance of truck drivers.

Behavior Modification, 20, 385-405.

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting and task performance. Englewood Cliffs, NJ: Prentice-Hall.

Ludwig, T.D., Biggs, J., Wagner, S., & Geller, E.S. (2001). Using public feedback and competitive rewards to increase the safe driving behaviors of pizza deliverers. Journal of Organizational Behavior Management, 21 (4), 75-104.

Ludwig, T. D.,&Geller, E. S. (1991). Improving the driving practices of pizza deliverers: Response generalization and moderating effects of driving history. Journal of Applied Behavior Analysis, 24(1), 31-44.

Ludwig, T. D., & Geller, E. S. (2000). Intervening to improve the safety of delivery drivers: A systematic behavioral approach [Monograph]. Journal of Organizational Behavior Management, 19 (4), 60-67.

Martin, G.,&Pear, J. (1996). Behavior modification: What it is and how to do it. Upper Saddle River, NJ: Prentice Hall.

Orpen, C. (1982). The effects of contingent and noncontingent rewards on employee satisfaction and performance. Journal of Psychology, 110, 145-150.

Poling, A., Smith, J. F.,&Braatz, D. (1993). Data sets in organizational behavior management: Do we measure enough? Journal of Organizational Behavior Management, 14 (1), 99-116.

Redmon, W. K.,&Agnew, J. L. (1991). Organizational behavior analysis in the United States: A view from the private sector. In P.A. Lamal (Ed.), Behavior analysis of societies and cultural practices (pp. 125-139). Washington, D.C.: Hemisphere.

Skinner, B. F. (1971). Beyond freedom and dignity. New York: Alfred A. Knopf.

Sulzer-Azaroff, B., & Mayer, G. R. (1991). Behavior analysis for lasting change. Fort Worth: Holt, Rinehart, and Winston.

Sussman, S. (1989). Two social influence perspectives of tobacco use development and prevention. Health Education Research, 4, 213-223.

Wilk, L.A., & Redmon, W.K. (1990). A daily-adjusted goal-setting and feedback procedure for improving productivity in a university admissions department. Journal of Organizational Behavior Management, 11(1), 55-75.