Berger, S. M., & Ludwig, T. D. (2007). Reducing Warehouse Employee Errors Using Voice-Assisted Technology That Provided Immediate Feedback. **Journal of Organizational Behavior Management**, 27(1), 1-31. DOI: 10.1300/J075v27n01_01 Version off record published by Published by Haworth / Taylor & Francis and is available online at: http://www.informaworld.com/ (ISSN: 0160-8061)

KEYWORDS: Employee errors | computer-based instruction | Feedback | voice picking

Reducing Warehouse Employee Errors Using Voice-Assisted Technology That Provided Immediate Feedback

Samuel M. Berger and Timothy D. Ludwig

ABSTRACT

A foodservice distributor in the southeastern United States implemented a voice assisted selecting tool to reduce selector errors by providing immediate feedback when errors occurred. An AB design with a nonequivalent comparison group was used to examine the effects of the voice technology on 132 selectors whose mispicks and shorts were collected over 6 weeks of baseline and 8 weeks of the intervention phase. Selector errors were reduced from 2.44 errors per 1,000 cases picked to 0.94 errors per 1,000 cases when voice technology was implemented. Further analysis indicated that the immediate feedback provided by voice had a greater impact on employees who were making the most errors during baseline.

ARTICLE

In a high-volume warehouse environment employee errors add up quickly and become quite costly. Employees called “selectors” account for nearly 40-60% of warehouses’ direct labor budgets (Miller, 2004). Selectors perform a key function within distribution warehouses by picking the items needed to fill an order that is then shipped directly to the retail customer. Selector accuracy has a direct impact on the quality of the product that the customer receives. In general, when a selector damages a product, picks the wrong product (a mispick), or leaves a product out of the order (a short) the customer is refunded at a financial loss to the warehouse. Any changes to the selection process that can reduce errors or increase productivity have a profound impact on the distribution centers' profitability.

Technology has played an important role in warehouse management systems (WMS), such as the use of wireless handheld and vehicle-mounted (e.g., fork-lifts and stock trucks) computers (Goomas&Ludwig, 2007) or voice-directed technology to manage and direct the workforce (Hill, 1996). Early warehouse management improvements involved real-time management of resources within a warehouse via terminals communicating radio frequency data throughout warehouses (Hill, 1996). The introduction of technology in warehouse and distribution management effectively changed how many employees conducted their jobs. Owing to the high-degree of automation, many information collection and distribution tasks that were once necessary were now under total automation of the WMS. Nevertheless, the effectiveness that many technological innovations offer in modifying employee behavior for performance gains have rarely been scientifically evaluated (Oran, 1991; Totty, 2005).

Voice-directed employee aids are becoming a common technology solution for productivity and error problems in these work environments (Lacefield, 2004). A voice-directed employee aid is a wireless, wearable system that allows employees to perform their job functions “hands free and eyes free.” The system consists of a battery-powered waist unit and a headset with an attached microphone that connects via radio frequency to a warehouse management system. This type of voice-directed system is increasingly becoming a tool used at large distribution centers for order picking.

In the 1990s large distribution centers, such as Wal-Mart, began to adopt voice-recognition technology (Lacefield, 2004). Providers of voice-recognition technology (e.g., Vocollect & Voxware) have reported a compounded annual growth rate of 70% over the last 4 years. These companies also make staggering claims regarding the effects of voice technology. Some of these claims include increasing accuracy up to 99.99% correct picking, productivity up to 50%, and reduction of training time by 50%.

Miller (2004) published a case study that compared the effects of voice-directed selection technology versus a bar code scanning selecting tool or traditional paper-based systems. He found that the implementation of voice technology was associated with an increase in overall accuracy from 99.52 to 99.64% resulting in an estimated 1.5 million dollars savings in a year. Similarly, in departments previously using bar-scanning devices, the switch to voice increased productivity 8-15% over previous levels. In the two departments still using the paper-based system productivity only increased 3-4%. The number of items left off an order (shorts) decreased by 11% and incorrect selections (mispicks) decreased by 25%. However, Miller did not provide any time table over which these changes occurred nor were specific error rates reported. Thus, the functional control (Kazdin, 1973) of the voice technology over performance could not be adequately assessed.

Immediate Feedback

One of the things that may contribute to the apparent success of voice technology is its ability to deliver immediate feedback to the user. Most definitions of feedback specify that it is information received by a responder regarding their performance on a task or set of tasks (Alvero, Bucklin, & Austin, 2001; Prue & Fairbank, 1981; Rummler & Brache, 1995; Sulzer-Azaroff & Mayer, 1991). There is a long-standing and robust set of evidence showing feedback as a means of increasing performance (Alvero et al., 2001; Ammons, 1956; Ashford & Cummings, 1983; Greller, 1980; Ilgen, Fisher, & Taylor, 1979; Ludwig & Geller, 1997). Yet feedback alone does not always improve performance. In Alvero et al.'s (2001) review of feedback literature in four prominent journals, they note seven different combinations of feedback with other antecedents and consequences that are commonly used as well as six different methods of delivering the feedback itself. Alvero et al. (2001) found that the results of feedback were most robust when paired with antecedents and consequences.

Feedback has been used successfully to reduce selector errors in warehouse settings. Bateman and Ludwig (2003) demonstrated that when feedback is paired with tiered goals, and an adapted incentive program, selector error rates in a distribution warehouse can be improved. However, the feedback provided by Bateman and Ludwig (2003) was only delivered once a week, when the individualized graphic feedback was posted.

Most of the research in organizational settings is lacking when it comes to the proximity of feedback to employee behaviors. Alvero et al.'s (2001) review provided evidence of this lack of research. The most frequent feedback in all the literature they reviewed occurred only twice a day (Poterfield, Evans, & Blunden, 1985; Richman et al., 1988; Wilk & Redmon, 1990, 1998) and most studies provided feedback only weekly (e.g., Nordstrom et al., 1988).

It has long been known that immediate reinforcement is more effective than delayed reinforcement (Skinner, 1938). It has been argued that reinforcement or feedback is best when provided immediately (e.g., Daniels, 2004; Sulzer-Azaroff & Mayer, 1991). Even short delays in the delivery of feedback have been shown to adversely impact learning (Aiken, 1968; Beeson, 1973; Gaynor, 1981).

Dihoff, Brosvic, Epstein, and Cook (2004) created an Immediate Feedback Assessment Technique (IF AT) that was used in classrooms to provide students with immediate feedback on practice tests. Performance on examinations was substantially better when students used the IF AT on practice tests rather than getting delayed feedback regarding the results of their practice tests (cf., Epstein et al., 2002).

Mason and Redmon (1993) studied the effects of immediate versus delayed feedback on error detection accuracy. They created a quality control situation in which participants were taught to identify different types of errors in pictures of a hard disk drive on a computer screen. Participants received feedback about their performance either immediately or after a delay. The immediate feedback group got to see their cumulative percentage of correct responses after each response while the delayed group only saw their percentage upon completion of the task. A machine-paced group received the stimuli at a pace determined by the computer program while a self-paced group got to control the pace at which the stimuli were presented. Mason and Redmon found that participants who worked at their own pace and received immediate feedback accurately detected more errors than the other groups.

In the workplace environment many authors have noted that providing immediate consequences to employees is often impractical and nearly impossible (Sulzer-Azaroff & Mayer, 1991). Dihoff et al. (2004) note that, unlike in the laboratory, feedback in applied settings is very difficult to deliver immediately. One possible reason for the lack of field research on the effects of immediate feedback is that until now researchers lacked the means to provide reliable immediate feedback. For example, supervisors attempting to provide immediate feedback would have to be present during each of the large number of selections made each day. A visit by the supervisor is generally a certain outcome when an excessive number of errors are made, but very few consequences are received by those employees whose selection errors do not draw a red flag. Indeed, most warehouse managers are content with a certain error threshold.

Recently experimenters and practitioners alike have begun to apply technology as a mode of delivering feedback in applied settings (Dihoff et al., 2004; Epstein et al., 2002; Goomas & Ludwig, in press; Terrel, 1990). These studies suggest that using technology to provide frequent feedback immediately following a behavior may have a substantial impact on performance. Although this may be a seemingly new line of research in OBM, made possible by new technological innovations, this research does have precedent in computer-based instruction (CBI).

Voice and Computer-Based Instruction

CBI, previously known as computer-assisted instruction (CAI), involves the use of computer software to aid an individual in learning a particular task. CBI has been used to train mathematical knowledge (Ku et al., 2004), languages (De Haan & Oppenhuizen, 1994), and employee behaviors (Eckerman et al., 2004).

The concepts of CBI are based largely on the works of B. F. Skinner and consist of methods of shaping behavior via operant conditioning

(Skinner, 1968). Responses are required to be made at the termination of each small set of instructions. This allows for the more immediate feedback to follow after each occurring behavior. CBI is often used to teach more complex processes on a molecular basis by breaking down complex behaviors into numerous smaller frames of necessary behaviors. This method of breaking instruction down into small individual steps accompanied by immediate feedback translated well into computer technologies.

One common structure of CBI provides standardized information and tests while the users control the pace (Anger et al., 2001). Information is delivered through a computer, and is followed by quizzes that assess the individual's retention of the material presented. Incorrect answers to questions result in the individual having to complete the item correctly before being able to continue.

The voice-directed technology evaluated in this study engaged selectors with a situation analogous to CBI. The voice system asked order selectors to respond with the correct check digits at each slot similar to the CBI scenario of presenting them with a "test." One of the main sources of selector errors is taking items from the wrong location (i.e., mispick). The voice system first directs selectors to a location by "speaking" a specific location to users through the headset. When selectors arrive at the designated location they read a "check-string" posted at the location into the microphone. The system then verifies whether the users are at the correct location. Selectors get instant verification of the correctness of their response, another CBI feature.

If selectors arrived at an incorrect location, they were not able to continue creating the pallet until appropriate arrival at the correct destination. When an incorrect selection is made the system will not move on to the next item. Instead it will repeat the command for that particular item. This not only lets users know immediately that the selection they made was incorrect but it also provides information that will help them get to the correct location.

Research in computer-based training has shown that it is beneficial to repeat or review problems that are answered incorrectly (Alessi & Trollip, 1985; Jonassen, Tessmer, & Hannum, 1999). In 1957, Wright noted that a "forced correct response" method of guidance facilitated learning, but only when paired with corrective feedback. In 1990, Terrel compared standard navigation training for Army aviation students to two experimental groups that received supplemental computer-based training. One of the experimental groups was unable to proceed in training until they provided a correct response. This group committed fewer errors in a post-test than the group that received standard training.

The voice technology evaluated in the present study provided selectors with immediate corrective feedback and required them to make a correct selection before it automatically proceeded to the next task. Based on the extensive pedigree from CBI and the robust behavioral changes associated with immediate feedback, the new voice technology system was predicted to decrease the number of selector errors made, including shorts and mispicks.

METHODS

Participants and Setting

This study was conducted at a foodservice distributor in the southeastern United States. The company's warehouse stores large quantities of food products, receives orders from grocery stores, and then ships the food products to the customer. Three departments from the warehouse participated in this study. The grocery department stored and shipped packaged dry food products. The produce department stored and shipped perishable food products. The mezzanine department stored and shipped nongrocery items (such as cigarettes). The grocery and produce departments received the voice technology intervention. The mezzanine department was used as a comparison group since they did not implement the voice-selection technology along with the other two departments.

The warehouse was organized into numerous aisles that were identified by two-digit numbers. The aisles were surrounded by large shelving called "slots" on either side of the aisle. Within each aisle there were numerous slots of different sizes, though most were approximately 6_ by 6_ slots and were filled with product cases. These slots were identified by the aisle letter and a specific 4-digit slot number.

The participants in this study were 471 "selectors" whose job was to pick out appropriate cases of food items and assemble orders to be shipped. The host company provided their written consent to allow the research team access to their employees, employee data, and warehouse methods. Additionally, selectors completed an informed consent form when asked for individualized data.

Work Process

When selectors began their task they received a page of stickers which listed the orders of food cases that were to be stacked on a pallet. There were two types of stickers. *Order Labels* described the order as a whole and contained the customer number, truck number, door number, pallet number, and the number of total pieces on each pallet. Order labels also had a bar code which could be scanned to bring up computerized information on that order. Alternatively, a *Case Label* contained information for each item in the pallet. These case labels provided selectors with the case's aisle number, slot number, quantity, and the product description.

The selector, using a pallet jack truck, drove around the warehouse retrieving each case based on the sequence in the order list. The stickers were arranged on a 4-inch-wide continuous sheet in an order that would lead selectors efficiently through the warehouse. Selectors did not necessarily have to follow this route. Selectors could make the decision to choose their own route through the warehouse. One reason they may do this would be to insure that they have an “ergonomically” correct pallet by picking heavier products first, followed by lighter products to reduce damage due to crushing.

After selecting a case, the selector was then required to stick the corresponding case label on to the item. The case label they affixed on the case had the product description printed on the label. This served as a way for the selectors to check and make sure they had picked up the correct item. Selectors were finished with their order when they were out of the case labels. They then took the pallets to be wrapped in plastic, which kept the order intact during transport on the truck. Finally, they put on the single order label with the order information so as to identify the pallet(s) for unloading.

Dependent Variables

Three common forms of errors can occur within this process. A “short” occurred if the selector neglected to pick up a case listed on the order list or if the location did not have sufficient quantity to fill out the order. A short typically occurred when the selector skipped over an item on the order list while attempting to build an “ergonomic” pallet and subsequently did not go back and pick the skipped item. Mispicks occurred when the selector picked up the wrong case. Mispicks occurred if selectors read the wrong case label off the order list or picked the wrong case and then failed to notice that the item description on the case label did not match the item description of the product on which they had placed a sticker.

The third type of error that occurred was damaging an item in the process of picking an order or delivering it to the shipping dock. Damage also occurred when a selector incorrectly built a pallet and crushed a lighter product stacked on the bottom of the pallet. Damage also occurred when a product was dropped or fell off the pallet. The incidents of damaged product were not predicted to be impacted by the implementation of voice technology. Therefore, damages were considered a constant series comparison behavior and were assessed alongside shorts and mispicks over the course of the study.

Internal audits were conducted on 1-3% of the outgoing orders to keep track of shorts, mispicks, and damages. During internal audits, after the selector assembled the pallet and delivered the pallet to the shipping dock, a Loss Prevention auditor disassembled the pallet before it was put on an outgoing truck. First, auditors looked at each case label to verify that it matched the case. Auditors also looked for damaged cases. Next auditors counted the number of cases in the order to make sure the pallet was not short any cases.

If there were no damages or mispicks, auditors wrote down the pallet number and the number of cases they counted in the pallet. Auditors then accessed the order information on a computer terminal and checked to make sure the number of cases counted on the pallet matched the number of items ordered for that pallet. If the pallet was short or if there were too many items (i.e., an “over”), they matched every item on the order list with the actual cases in the pallet until they determined which case(s) were missing or which case(s) didn’t belong.

Auditors completed a “quality audit” report. They filled in values for shorts, mispicks, damages, and overs (picking too many of a desired product) along with information about the employee, customer, date, and department. The “quality audit” reports were then entered into a spreadsheet. The Loss Prevention auditor then re-wrapped the pallet and transported it to the appropriate door number to be shipped. If there were errors found during the auditing process either the original selector or the auditor replaced the missing or damaged items.

Data on employee errors were listed by their clock number and electronically transferred to the research team. Data were then compiled onto a single database tracking individual clock numbers (i.e., employees) over the course of the study.

Design

An AB design with a nonequivalent comparison group and a constant series comparison behavior was used to evaluate the effects of the voice-assisted system. The grocery department and produce department received the voice technology while the mezzanine department was used as a comparison group. Voice was targeted to reduce shorts and mispicks yet it was unlikely to have any effect on damages. However, damages were measured as a potential control for other events in the warehouse that may have impacted quality.

Six weeks of baseline data were collected. Afterward, selectors were trained to use the voice technology to select orders over a 2-week period. After training, 6 weeks of data were collected while all members of the grocery and produce departments used voice technology. The timing of the intervention implementation was determined by the host company. Therefore, common behavior analytic methodologies, such as assuring a stable baseline before implementation or pursuing a multiple baseline, could not always be followed.

Voice Technology

When using voice technology, each selector received a headset with an earpiece and a microphone. These headsets were connected to a waist unit that contained volume and voice speed controls, as well as a small computer to translate data into spoken voice. The waist unit was linked via radio-frequency receiver to a computer system that relayed order information back to the waist unit. These data were translated into a spoken voice that prompted the selector with the slot location and quantity of the item needed.

Once employees reached the specific item they spoke into the headset microphone by reading a three digit “check string” off of a sticker beside the slot to ensure that the employee was at the proper slot. Once the system cross-checked the quantity and the 3-digit “check string,” it prompted employees with another location and pick quantity. If the employee selected the wrong item and thus read the wrong 3-digit check string the voice system would not move on to the next item. Instead, it continued to prompt selectors with the correct location and “check string.” When the correct “check string” was properly read, selectors were then instructed to pick the number of cases requested by the customer. They then spoke the quantity they picked into the microphone.

If the quantity spoken was correct then the system directed the selector to the next slot to pick the next item.
If selectors read an incorrect “check string” but wished to continue to the next item on the order they could command the system to “skip” that item. Most “skip” commands occurred when employees preferred stacking their pallet differently than the order of the lots to avoid damages (e.g., from placing a heavy item on top of a crushable item). Presumably the employee would skip the item but then go back and pick the item when they wanted it on the pallet. This ability to skip, however, could result in additional shorts under the voice system.

Training

Prior to training, selectors had to create their own personal voice template so that the system would recognize their voice and understand their commands. While creating templates selectors were shown how to correctly put on the waist unit and headset. Training began once templates were created for all selectors. Selectors were trained individually and training took approximately 1 hour. The trainers showed the selectors how to turn on the system and select their personal template. The trainer simply started on an order and walked trainee selectors through the process, telling them when to respond to the system and correcting any errors made. Trainers had on a headset and waist unit that was connected to the trainee’s so they could listen to the employee-computer interactions. Trainers remained with their trainee until the selector was comfortable with the process and could demonstrate an understanding of how to use the system. Once the employee was familiar with the voice system they could turn up the speed on the system allowing them to work faster. Posters that contained instructions on what to do when selectors started a shift, finished a shift, changed batteries, or needed help, were posted in centralized locations.

Employee Satisfaction

A 13-item survey, developed by the authors, was designed to assess work satisfaction along with satisfaction with feedback, training, interaction, job tools, and technology. The survey was developed with 5-point scales and anchors unique to each question. The employee satisfaction survey was administered to a random group of 44 selectors 1 month prior to the implementation of voice picking and re-administered (i.e., post-test) to these same individuals 3 months after the implementation of voice technology.

Officers of the host company agreed to pay overtime for some employees to complete these surveys. Because of cost issues the company did not want to pay overtime to all their selectors to complete the survey. The survey was administered to randomly selected participants in 4-group meetings by a member of the research team and a representative from the company's Human Resources Department. Survey participants were told the meaning of the survey, offered informed consent with assurances of confidentiality, and were presented with the survey. Participants were asked to put their clock number on the survey along with their responses. The clock number allowed researchers to compare participants' responses across two time periods (pre and post voice) and link responses with participants' error data, which were also linked to their clock number. When finished, participants put their survey in a large envelop, sealed it, and signed the seal.

Three months after the implementation of voice picking an identical employee satisfaction survey was again administered. Of the 44 selectors who completed the first survey, a total of 30 selectors took it again. This represented an attrition of 14 original participants (31%) most of whom were no longer working with the company.

RESULTS

Audits were done daily for 6 weeks during baseline, 2 weeks during training, and for 6 additional weeks after the voice technology were implemented to measure shorts, damages, mispicks, and the total number of cases in each order. There were 471 selectors in these three departments who were observed over the course of the study. To be included in the data analysis, selectors must have had at least 300 cases audited during baseline and at least 300 cases audited during intervention (including the training period). This inclusion criterion was set to assure that the data were analyzed using the same participants throughout. Because the dependent variable (i.e., errors per 1,000 cases) was a ratio instead of an absolute number, inclusion into the data analysis also had to be based on a large enough number of observations (i.e., the denominator) so as not to invalidate the data with undue variance. In short, we needed to have enough audits of the participant to get an adequate sample of their behavior.

The host organization suggested that employees have a minimum of 300 audits before reviewing individual employee errors. A visual inspection of the participant pool suggested that using this number would eliminate the majority of selectors who were part-time employees, absent for all or some of a phase, or who for some reason had insufficient audits to draw valid conclusions.

Using this criterion, 339 selectors were eliminated and a total of 132 selectors were used in the data analysis. A total of 85 selectors were used in the experimental group including 38 in grocery and 47 in produce. A total of 47 selectors from mezzanine comprised the comparison group.

A total of 268,110 cases were audited during baseline and 260,413 cases were audited during the intervention period. After the selectors whose data did not surpass the inclusion criterion were removed, 202,887 cases were audited and used for baseline data analysis and 207,843 cases were audited and used for intervention data analysis (i.e., training and post-implementation). So while only 28% of selectors' data were included in the data analysis, roughly 77% of the cases audited were included in the data analysis.

For data analysis purposes, 28,515 cases were audited during baseline in grocery compared with 27,487 during intervention. In the produce department 66,186 cases were audited during baseline compared with 44,271 during intervention. In the comparison group, the mezzanine, 108,186 cases were audited during baseline compared with 135,725 during the intervention occurring in the treatment departments. The average number of cases audited per person was 1,532 during baseline and 1,576 during the intervention.

For each type of error (i.e., shorts, mispicks, and damages) an average for each selector was obtained by summing the number of errors made during each period analyzed (i.e., week or phase) and dividing by the sum of all the cases audited. This product was then multiplied by 1,000 to reflect the estimated errors per 1,000 cases selected.

The shorts, mispicks, and damages per 1,000 cases selected for each participating department during the baseline and intervention phases are reported in Table 1. In the grocery department shorts decreased from 1.33 per 1,000 cases during baseline to 0.32 per 1,000 cases during intervention. Mispicks decreased from 3.93 per 1,000 cases during baseline to 1.19 per 1,000 cases during intervention. The combination of shorts and mispicks (the typical metric used in this industry) decreased from 5.26 per 1,000 cases during baseline to 1.51 per 1,000 cases during intervention. Damages (i.e., constant series comparison variable)

TABLE 1. Shorts, Mispicks, and Damages Per 1,000 Cases for the Grocery, Produce, and Mezzanine Departments Across Baseline and Intervention (Training and Post Implementation) Audits

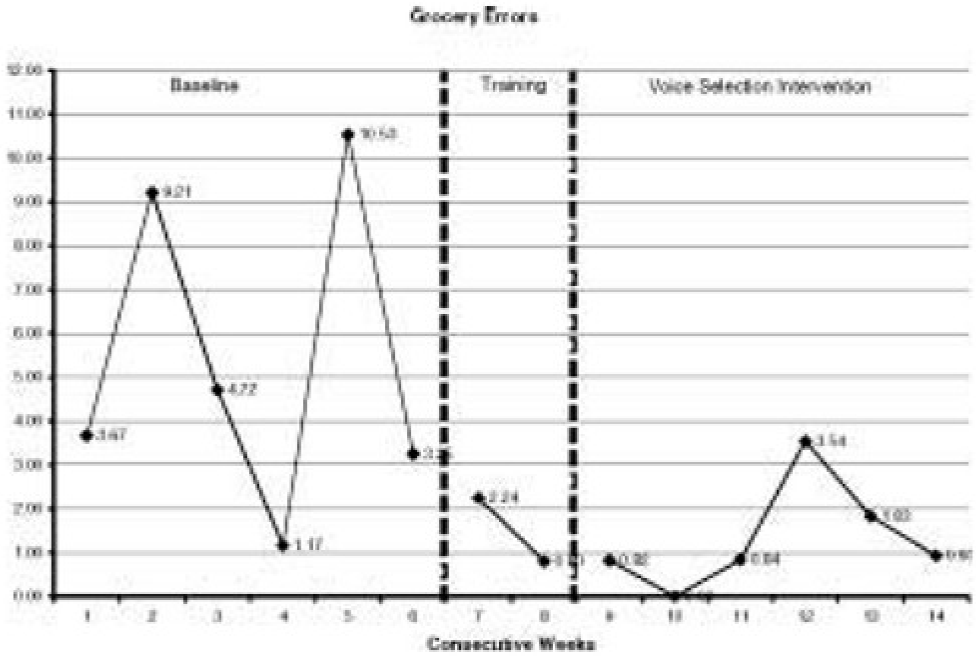
Department	Errors	Baseline M (SD)	Intervention M (SD)
Grocery	Shorts*	1.33	0.32
	Mispicks*	3.93	1.19
	Shorts + Mispicks*	5.26 (10.55)	1.51 (3.40)
	Damages	0.28	0.29
Produce	Shorts*	0.18	0.29
	Mispicks*	1.04	0.29
	Shorts + Mispicks*	1.22 (2.34)	0.59 (1.91)
	Damages	0.09	0.16
Mezzanine	Shorts*	2.14	2.03
	Mispicks*	0.77	1.03
	Shorts + Mispicks*	2.90 (2.88)	3.06 (6.36)
	Damages	0.01	0.01

Note: Asterisks appear next to targeted behaviors (Shorts and Mispicks in Grocery and Produce). Standard deviations appear in parentheses.

stayed fairly constant going from 0.28 per 1,000 cases during baseline to 0.29 per 1,000 cases during intervention. Figure 1 shows grocery's weekly error rate (shorts + mispicks) throughout the course of the study. The grocery department's error rate dropped after the implementation of voice. A decrease in the variability of the weekly errors was also observed after the voice technology was implemented.

In the produce department shorts increased slightly from 0.18 per 1,000 cases during baseline to 0.29 per 1,000 cases during intervention. Mispicks decreased from 1.04 per 1,000 cases during baseline to 0.29 per 1,000 cases during intervention. The combination of shorts and mispicks decreased from 1.22 per 1,000 cases during baseline to 0.59 per 1,000 cases during intervention. Damages (constant series comparison variable) stayed fairly constant going from 0.09 per 1000 cases during

FIGURE 1. Average Number of Errors for Selectors in the Grocery Department by Week

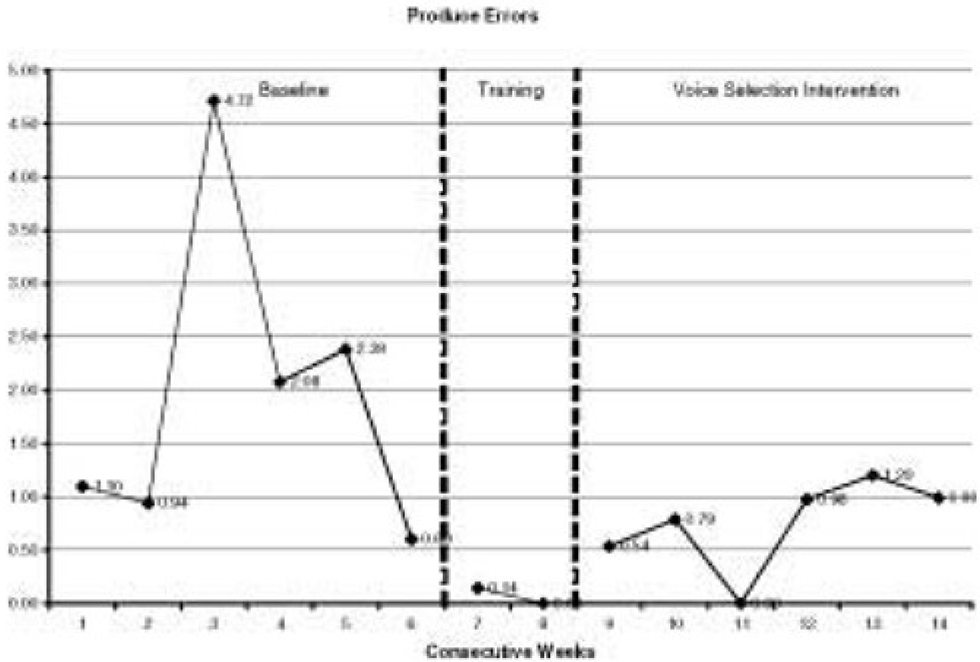


Note: The Vertical Lines Separate Phases and the Horizontal Lines Represent Means Within the Given Phase.

baseline to 0.16 per 1,000 cases during intervention. Figure 2 shows produce's weekly error rate (shorts + mispicks) throughout the course of the study. The produce department's error rate also dropped after the implementation of voice. A decrease in the variability of the weekly errors was also observed after the voice technology was implemented.

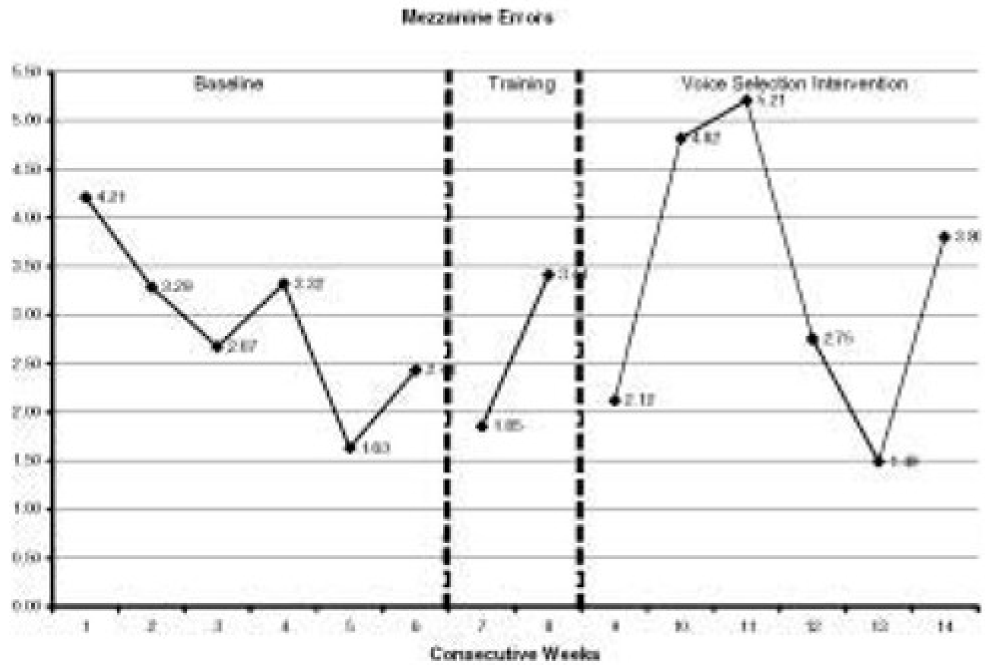
In the mezzanine department, where the voice technology was not implemented, shorts per thousand decreased just slightly from 2.14 per 1,000 cases during baseline to 2.03 per 1,000 cases during intervention. Mispicks increased slightly from 0.77 per 1,000 cases during baseline to 1.03 per 1,000 cases during intervention. The combination of shorts and mispicks increased from 2.90 per 1,000 cases during baseline to 3.06 per 1,000 cases during intervention. Damages stayed relatively constant at 0.01 per 1,000 cases for both baseline and intervention phases. Figure 3 shows mezzanine's (comparison group) weekly error

FIGURE 2. Average Number of Errors for Selectors in the Produce Department by Week



Note: The Vertical Dotted Lines Separate Phases and the Horizontal Dotted Lines Represent Means Within the Given Phase.

FIGURE 3. Average Number of Errors for Selectors in the Mezzanine Department by Week



Note: The Vertical Lines Separate Phases and the Horizontal Lines Represent Means Within the Given Phase.

rate (shorts + mispicks) throughout the course of the study. The mezzanine department's error rate stayed relatively constant before and after voice implementation. However, the variability of weekly errors did increase after the implementation of voice.

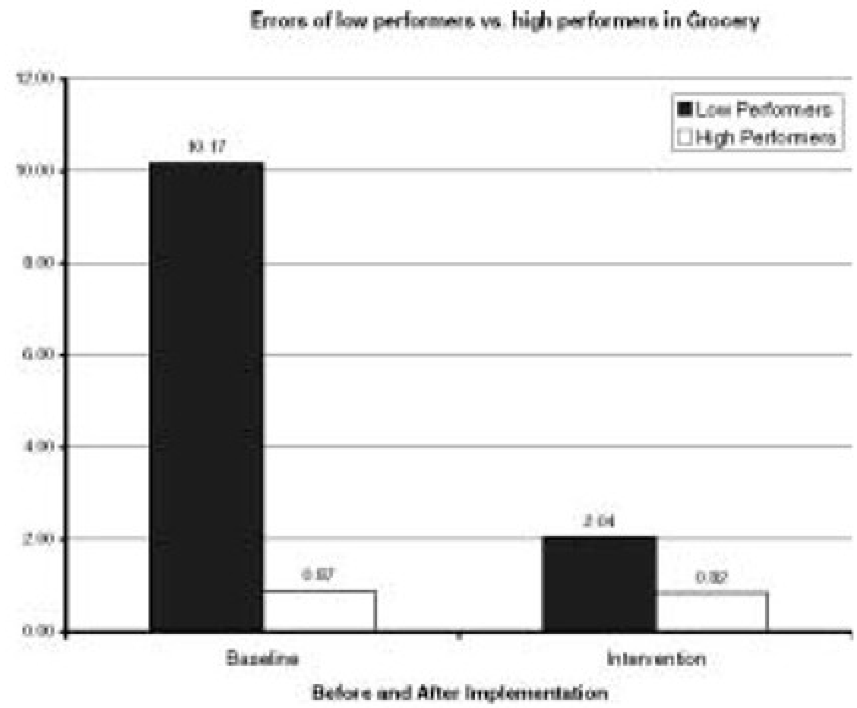
High versus Low Performers

Selectors were then divided into two categories by taking the median number of baseline errors made by all the selectors in that department. All those selectors above the median were classified "low performers" and all those below the median were "high performers."

The biggest impact of the voice technology on the error rate (i.e., mispicks + shorts) was for the "low performers" as shown in Figure 4. In the grocery department the "low performers" (n = 19) error rate dropped from 10.17 per 1,000 cases during baseline to 2.04 per 1,000 cases during intervention. The error rate of the "high performers" (n = 19) was relatively unaffected, changing from 0.87 per 1,000 cases during baseline to 0.82 per 1,000 cases during intervention.

As shown in Figure 5, the produce department's error rate (mispicks + shorts) of the "low performers" (n = 23) decreased from 2.79 per 1,000

FIGURE 4. Errors Per 1,000 Cases Selected for “High” and “Low” Performers in the Grocery Department During the Baseline and the Intervention Phase

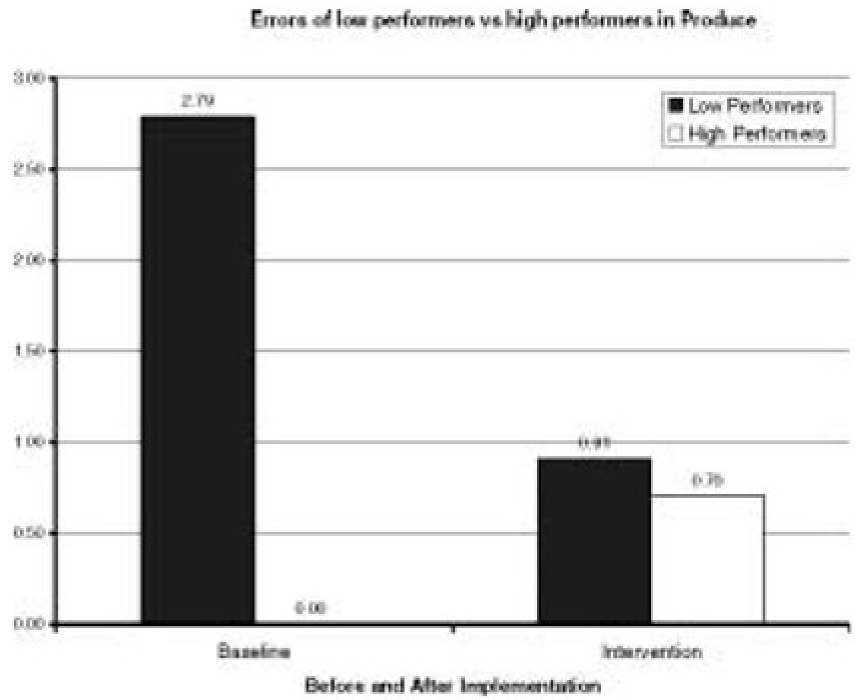


cases during baseline to 0.91 per 1,000 cases during intervention. The “high performers” (n = 24) error rate increased from 0.0 per 1,000 cases during baseline to 0.70 per 1,000 cases intervention.

Employee Satisfaction

The results of the work satisfaction survey administered to a random group of selectors prior to and after the implementation of voice technology are presented in Table 2. The survey average (on a 5-point

FIGURE 5. Errors Per 1,000 Cases Selected for "High" and "Low" Performers in the Produce Department During the Baseline and the Intervention Phase



scale) was 3.47 prior to voice and 3.30 after the implementation of voice (t = 2.73, $p < 0.05$). Most of the items on the survey did not result in significant changes across time. However, employees reported feeling significantly less motivated to come to work each day (t = 2.22, $p < 0.05$) and receiving significantly less constructive feedback (t = 2.25, $p < 0.05$).

Two additional items that were specific to the new voice process also showed significant differences between administrations. Respondents reported they had significantly less opportunity to pick cases in the order they choose (t = 4.97, $p < 0.05$), and less opportunity to interact with peers (t = 2.18, $p < 0.05$).

To explore the survey data further, correlations were conducted to compare survey responses with the participating selectors' error data during baseline. This analysis revealed significant positive correlations between selector errors and reports of receiving feedback about their

TABLE 2. Means and Standard Deviations (in Parentheses) of Employee Satisfaction Survey Items Administered Before and After Voice Implementation

Question	Pre-Voice	Post-Voice	*t*-Value
I feel motivated to come to work each day.	3.72 (0.78)	3.38 (0.96)	2.22*
I feel that my job as a selector is boring (negatively scaled).	2.58 (0.97)	2.76 (0.93)	−1.27
I feel that the feedback I receive regarding my performance is constructive.	3.77 (0.93)	3.41 (1.03)	2.25*
I receive feedback regarding my performance.	3.18 (1.05)	3.07 (1.14)	0.46
I get the opportunity to pick cases in the order I choose.	3.55 (1.13)	2.0 (1.14)	4.97*
I feel that I have received adequate training to perform my job effectively.	3.8 (0.96)	4.0 (0.87)	−1.13
I feel I have all the necessary tools to perform my job effectively.	3.79 (0.82)	3.66 (9.96)	1.38
I feel that I have the opportunity to interact with peers while at work.	3.54 (0.89)	3.14 (0.97)	2.18*
I'm open to new ideas to be implemented in the workplace.	3.65 (1.12)	3.48 (0.89)	0.88
I feel that, overall, (this company) is a good place to work.	3.86 (0.73)	4.07 (0.74)	−1.6
I tend to look ahead on the case labels to better pack my pallets.	3.77 (1.13)	4.14 (0.97)	−1.5
I think Voice will make my work easier.	3.07 (0.91)	3.28 (1.08)	−1.1
I'd rather not use computers and technology to perform my job as a selector (negatively scaled).	2.63 (0.81)	2.46 (1.12)	0.926
Survey Average	3.47 (0.39)	3.30 (0.46)	2.73*

Note: All questions on the survey were anchored with a 5-point scale with "5" representing the highest satisfaction with the area except for the two items noted as "negatively scaled." *t*-Test coefficients are denoted with an asterisk if $p < 0.05$.

performance ($r = 0.32$), reports of interacting with peers ($r = 0.34$), and reports of being open to new ideas ($r = 0.32$). No other significant correlations between baseline errors and survey results were noted.

Cost-Benefit Analysis

Each error was estimated to cost the warehouse an average of $15.75 (Garvey, personal communication, February 11, 2005). This is a conservative estimate compared with Miller's (2004) estimate of $20 an error. A reduction of 163 errors occurred in the two experimental departments (grocery and produce) during the last 8 weeks of the study. Given that audits were conducted on approximately 3% of orders in those departments, the implementation of voice technology eliminated an estimated 5,433 errors in the experimental departments over the

8weeks following implementation. This resulted in an estimated $85,569 in savings over an 8-week period in two departments. Had voice been implemented in the mezzanine department and cut the error rate in half (error rates were dropped over 50% in both experimental groups), it would have saved an estimated 7,166 errors worth an additional $112,864. Summing the savings in all three departments amounts to $198,443 saved in 8 weeks. This amounted to roughly 1.28 million dollars in savings a year. This estimate only accounts for audits included in the data analysis. If the reduced errors of all 339 selectors are included the estimated savings is approximately $1.48 million a year.

The total cost of implementing voice included purchasing the technology hardware and software, employee training time, and vendor support for the design and set-up of the technology. These costs were estimated at approximately $1.3 dollars (Garvey, personal communication, February 11, 2005). Therefore, it is probable that the cost of implementing the voice technology would be recovered in cost-savings owing to reduced selector errors in about 1 year. Thereafter, the warehouse would realize over a million dollars in savings annually, minus ongoing technology service and upgrade costs.

DISCUSSION

The findings of this study tentatively supported the hypothesis that voice technology can decrease the incidents of shorts and mispicks among selectors in a high-volume distribution warehouse environment. The largest decrease in errors occurred in the grocery department, which saw a substantial drop in the number of mispicks that occurred per thousand cases audited during the implementation of the voice technology. The variability in the number of errors made per thousand cases also narrowed substantially.

The produce department also saw decreases in mispicks after the implementation of voice, yet their results were not as substantial as the grocery department. One reason for this may be owing to the already rather low error rate, especially for shorts, in the produce department during baseline. However, the overall rate of shorts and mispicks did drop after implementation, as did the variance in the number of errors committed.

The claims of some of the voice-recognition providers of 99.99% accurate picking (vocollect.com) were not realized. The produce department did reach this level of accuracy although they were performing close to that level before the implementation of voice technology. Though the grocery department may have decreased its error rate it was unable to achieve a 99.99% accuracy level. However, there were weeks when the grocery department reached 99% accuracy.

The mezzanine department functioned as a comparison group and its mean number of shorts + mispicks per thousand cases audited increased slightly over the same time period. Damages also functioned as a constant series comparison measure. Over the course of the study, the number of damages in each department remained relatively stable before and after implementation of voice technology.

When employees were divided into “high performers” and “low performers,” the impact of voice technology was made clearer. Voice technology drastically helped employees who were performing at a low level. However, it did not help those employees who were already performing well. It is noteworthy that at another warehouse within the corporate structure of the host warehouse, voice technology is used only for those selectors whose error rate dropped below acceptable levels. The finding that voice technology may reduce low performers’ error rates suggests that the use of voice technology can be an effective training tool to improve unacceptable performance. Further analyses should be conducted, such as studies with much longer periods of post-implementation audits, to determine whether voice is a cost-effective tool for a warehouse’s highest performers.

These general findings are consistent with the case study conducted by Miller (2004) who found substantial decreases in employee errors when voice technology was implemented. This study also supports the assertion that immediate feedback can be beneficial in reducing employee errors (Dihoff et al., 2004; Epstein et al., 2002; Goomas & Ludwig, in press; Mason & Redmon, 1993; Terrel, 1990). It could be argued that voice technology is yet another, more sophisticated iteration of CBI, although in the present study the company used the technology within its daily work processes. The voice-directed selection technology in its current form uses some of the basic behavioral principles that are effective in CBI.

Three-Term Contingency After Implementation of Voice

Following implementation of the voice-directed system, the work environment changed substantially. For example, the antecedents of behaviors associated with selectors’ jobs changed after the implementation of voice technology. Prior to implementation, antecedents took the form of tickets specifying the locations and quantities of each individual product that needed to be added to pallet. After implementation, antecedents were delivered orally, as the system would read aloud the information regarding the location and identification numbers to the selector. With this change the selector was no longer required to look at a ticket and read the necessary information. The information was instead read to the selector. This may have eliminated reading-related errors and also allowed the selector to spend more time reading row and slot identifiers. Finally, the selector could have, at any time, requested verbally to have the system repeat any information.

Once the selector arrived at the destination to retrieve the cases, a new behavior was introduced to the work process. A number was read by the selector back to the system in order to verify the successful arrival at the location. Previously, it was the responsibility of the selector to inspect the digits on the order ticket with those on the slot. There was no way to verify that this was being done until errors were made. The voice technology process which was implemented ensured that this "checking" behavior was completed.

Voice also effectively changed the consequences associated with making a selection. Using the old paper-/sticker-based selection process selectors moved onto the next case on the order list as soon as they made a selection, no matter if the selection was incorrect or not. There was no feedback regarding the accuracy of any one selection unless a selector was audited (a 3% chance) and only then did they receive feedback regarding errors and that feedback was delivered approximately 1 hour after the particular selection had occurred.

In the voice system, if the check code read by the employee was incorrect, then the employee experienced an immediate consequence. Employees were given feedback that their location was incorrect and were generally not allowed to progress further unless they decided to "skip" the item with a verbal command. Similarly, if the code was correct, then immediate reinforcing consequences followed because selectors got to move on to the next item and thus be one step closer to task-completion.

According to Sulzer-Azaroff and Mayer (1991), effective feedback must be an antecedent of behavior. Feedback should exert stimulus control over the employee's behavior and function as a discriminative stimulus (SD) that describes the consequences of the (in)correct behaviors. Under these circumstances feedback can also serve as a reinforcer (Sulzer-Azaroff & Mayer, 1991).

In this intervention, immediate feedback was delivered to the user. The fact that feedback occurred immediately after the behavior allowed the user to repeat the behavior correctly within the work process. Both behavioral functions of feedback are evident here (Sulzer-Azaroff & Mayer, 1991). The feedback occurred immediately after the behavior, presumably negatively reinforcing correct behavior (avoiding the corrective voice statement by picking the right item). Additionally, corrective feedback also served to prompt the correct behavior to mitigate the error.

Empirical evidence from the literature on feedback (Alvero et al., 2001; Balcazar, Hopkins, & Suarez, 1986) suggests that the mere existence of feedback contingencies in the work setting do not result in adequate performance. In the present study, contingencies prior to the intervention were inadequate to maintain a low number of errors. The proximal contingencies put in place with the voice system may have impacted errors in ways desirable to employees (avoiding mistakes and their work consequences) and company alike.

Cost-Benefit Analysis

The cost/benefit analysis of voice-recognition technology is striking. By saving approximately $1.28-1.48 million in the first year alone the voice technology will have paid for itself. Although this study did not examine the long-term effects of voice technology, Miller's (2004) case study suggests that the long-term effects may be better than those found in this study.

In the current study only the internal audits were linked with the selector who picked the order. The foodservice distributor also kept track of complaints from customers when orders were incorrect. If the delivered order contained shorts, mispicks or damaged goods, customers would request credits for the errors. Although these credits were a good composite of selector errors, customer complaints did not always accurately reflect the performance of a selector because sometimes customers may have been wrong, lied, or the goods were damaged in transit.

In fact, when customers have to inspect the delivered orders at the retail store many additional costs are incurred. The distributor has to pay the truck driver while the customer counts the orders, which reduces the number of deliveries a driver makes in a day and ends up being very expensive. Additional system costs are incurred through the system when the retail customers have to pay employees to check the orders and do the clerical work to claim their credits for the distribution errors discovered.

If the distributor can reduce errors throughout the warehouse to a consistent rate of less than one per thousand then many of these additional costs may be avoided. In fact, at the conclusion of this study, the host distribution center was working with their largest grocery stores to automatically give the store one credit per thousand cases ordered and otherwise stop the costly warehouse audits and store inspections. Therefore the cost-benefit analysis reported in this study may not have captured all of the cost-savings realized throughout the distribution-grocery system.

Employee Satisfaction

An employee work satisfaction survey was administered to a random group of selectors before and after the implementation of voice. The results suggest that the selectors surveyed were generally less satisfied with their work, reporting being less motivated and not getting constructive feedback. However, many of the satisfaction questions showed no changes between administrations and employees rated the company as a good place to work (4.07 on a 5-point scale). It should be noted that the first administration of the survey occurred in November and the second in January. Historical artifacts such as managerial changes, work process changes, and weather may have influenced the results. Indeed, these artifacts cannot be ruled out without a control sample taking the survey twice without experiencing the implementation of voice. Additionally, the psychometric quality of the survey itself had not been assessed, so no assurances of reliability or validity can be made.

Two items, specific to the new work process adopted by selectors when working with voice, were included in the survey. Specifically, selectors reported that they were no longer able to control the sequence in which they progressed through their order. Second, the headphones used in voice may reduce the ability of selectors to speak to one another while working. In both of these cases we might find some further insight into why voice worked in this study. Limiting variation in the work process by not allowing the selectors to work out of sequence decreased one of the behaviors responsible for shorts (skipping an item and not going back to get it). Also, warehouse managers often blamed employee interactions taking place mid-task as a distraction that caused errors in selecting the right cases.

Limitations and Future Research

The major limitation of this study was that a quasi-experimental design was employed; that is, random selection or assignment to groups was not (and could not be) conducted. Therefore, the data may have been subject to inherent differences between departments as can be seen in the difference in error rates during the baseline observations. Also, the mezzanine may not have been an adequate comparison to contrast with the experimental groups because of its slightly different work processes and supervision. However, it was beneficial to have a constant series comparison throughout the course of the study.

Additionally, this study employed AB designs (i.e., Baseline and Intervention with no reversal) in two departments. Owing to the eagerness of the host company’s management to implement the expensive and potentially beneficial voice technology, the interventions in both departments began at the same time. Therefore we could not take advantage of a multiple baseline to evaluate the results. However, the effects of voice technology were replicated across three departments.

AB designs alone are not sufficiently adequate to demonstrate functional control of the intervention over the targeted behaviors. This should be considered another major limitation of the study. Future research could use a multiple baseline design by staggering voice technology implementation randomly throughout selectors or randomly selecting and assigning employees into implementation and control groups. In a large distribution warehouse that employs many different people complete control of all variables is unlikely. Therefore, this study should be replicated in a more controlled laboratory setting.

Nevertheless, within the host organization of this study, the adoption of voice technology seemed to have practical cost-saving benefits at the time of implementation. Executives in this company and those like it that adopt these types of technologies will constantly be trying to improve upon these results to get further returns on their investments. Organizational Behavior Management (OBM) can play a large role in adapting technologies like voice-selection so they maximize the effective, immediate, and personalized delivery of antecedents and consequences (Ludwig, 2003). For example, incentive and goal programs could increase the effect of voice-recognition technology whereby employees can request real-time feedback towards their incentive goals while in the middle of their tasks. Indeed, Goomas and Ludwig (2007) already have reported using technology in a warehouse setting to provide immediate feedback that was paired with tangible rewards and goal setting to increase productivity.

REFERENCES

Aiken, E. G. (1968). Delayed feedback effects on learning and retention of Morse code symbols. *Psychological Reports*, *23*, 723-730.

Allessi, S. M. & Trollip, S. R. (1985). *Computer-based instruction: Methods and development.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

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

Ammons, R. B. (1956). Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology, 54*, 279-299.
Anger, W. K., Rohlman, D. S., Kirkpatrick, J., Reed, R. R., Lundeen, C. A., &

Eckerman, D. A. (2001). cTRAIN: A computer-aided training system developed in SuperCard for teaching skills using behavioral education principles. *Behavior Research Methods, Instruments, and Computers*, *33*(2), 277-281.

Asford, S. F. & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies for creating information. *Organizational Behavior and Human Performance*, *32*, 370-398.

Bateman, M. J. & Ludwig, T. D. (2003). Managing distribution quality through an adapted incentive program with tiered goals and feedback. *Journal of Organizational Behavior Management*, *23*, 33-55.

Beeson, R. O. (1973). Immediate knowledge of results and test performance. *Journal of Educational Research*, *66*, 224-226.

Daniels, A. C. (1994). *Bringing out the best in people.* New York: McGraw-Hill.

Daniels, A. C. & Daniels, J. A. (2004). *Performance management: Changing behavior that drives organizational effectiveness.* Atlanta, GA: Aubrey Daniels International.

De Haan, A. & Oppenhuizen, T. (1994). A reflexive ITS to support the learning of second language spelling. *Computers in Human Behavior*, *10*, 21-31.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, *54*, 207-231.

Eckerman, D. A., Abrahamson, K., Ammerman, T., Fercho, H., Rohlman, D. S., & Anger, W. K. (2004). Computer-based training for food services at a hospital. *Journal of Safety Research*, *35*, 317-327.

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A.,

Epstein, B. B.,& Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, *52*, 187-201.

Garvey, D. (2005, February 11). Personal communication with the Vice President of Operations at the host organization.

Gaynor, P. (1981). The effect of feedback delay on retention of computer-based mathematical material. *Journal of Computer-Based Instruction*, *8*, 28-34.

Goomas, D. T. & Ludwig, T. D. (2007). Enhancing incentive programs with proximal goals and immediate feedback: Engineered labor standards and technology enhancements in stocker replenishment. *Journal of Organizational Behavior Management, 27*(1), 33-68.

Greller, M. M. (1980) Evaluation of feedback sources as a function of role and organizational level. *Journal of Applied Psychology*, *65*, 24-27.

Hill, J. M. (1996). Warehouse management systems: The competitive tool for the 1990s. *Modern Materials Handling*, *51*, 3-12.

Ilgen, D. R., Fisher, C. D.,&Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*, 349-371.

Jonassen, D. H., Tessmer, M.,& Hannum, W. H. (1999). *Task analysis methods for instructional design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kang, K., Oah, S.,&Dickinson, A. M. (2003). The relative effects of different frequencies of feedback on work performance: A simulation. *Journal of Organizational Behavior Management*, *23*, 21-53.

Kazdin, A. E. (1973). Methodological and assessment considerations in evaluating reinforcement programs in applied settings. *Journal of Applied Behavior Analysis*, *6*, 517-531.

Ku, H. Y., Harter, C. A., Liu, P. L., Thompson, L., & Cheng, Y. C. (2007). The effects of individually personalized computer-based instructional programs on solving mathematics problems. *Computers in Human Behavior, 23*(3), 1195-1210.

Lacefield, S. (2004, October 1). I can hear you now . . . as the technology matures and ROI improves, the adoption rate for voice-recognition technology is rising. *Logistics Management.* Reed Business Information, a division of Reed Elsevier Inc. Last retrieved from http://www.logisticsmgmt.com/article/CA470802.html, April 30, 2007.

Ludwig, T. D. (2003, January). *Skinner box on wheels: Today's short haul trucking fleets.* Invited paper presented at the Organizational Behavior Management Network Meetings, St. Pete Beach, FL.

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

Ludwig, T. D.&Geller, E. S. (2001). *Intervening to improve the safety of occupational driving: A behavior change model and review of empirical evidence*. Binghamton, NY: Haworth.

Mason, M. A.&Redmon, W. K. (1992). Effects of immediate versus delayed feedback on error detection accuracy in a quality control situation. *Journal of Organizational Behavior Management*, *13*, 49-78.

Miller, A. (2004). Order picking for the 21st Century. Unpublished manuscript Tompkins Associates, Raleigh, NC.

Nordstrom, R., Hall, R. V., Lorenzi, P., & Delquadri, J. (1988). Organizational behavior modification in the public sector: Three field experiments. *Journal of Organizational Behavior Management*, *9*, 91-112.

Oran, D. P. (1991). Apply psychology to product design. *Machine Design*, *63*, 37-41.
Porterfield, J., Evans, G.,& Blunden, R. (1985). Involving families and staff in service improvement. *Journal of Organizational Behavior Management*, *7*, 117-133.

Prue, D. M. & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management*, *3*, 1-16.

Rummler, G. A. & Brache, A. P. (1995). *Improving performance: Managing the white space on the organizational chart.* San Francisco, CA: Jossey-Bass.

Schaffer, L. C. & Hannafin, M. J. (1986). The effects of progressive interactivity on learning from interactive video. *Educational Communication and Technology Journal*, *2*, 89-96.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York, NY: Appleton-Century-Crofts.

Skinner, B. F. (1968). *The technology of teaching*. New York, NY: Appleton-Century-Crofts.

Sulzer-Azaroff, B. & Mayer, G. R. (1991). *Behavior analysis for lasting change*. Fort Worth, TX: Harcourt Brace College.

Terrell, D. J. (1990). A comparison of two procedures for remediating errors during computer-based instruction. *Journal of Computer-Based Instruction*, *17*, 91-96.

Totty, M. (2005, January 7). Technology, a special report; Business solutions. *Wall Street Journal*, Section R, pp. 4.

Waldrop, P. B., Justen, J. E. III, & Thomas, M. A. II. (1986). A comparison of three types of feedback in a computer-assisted instruction task. *Educational Technology*, *26*, 43-45.

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*, 55-75.

Wilk, L. A. & Redmon, W. K. (1998). The effects of feedback and goal setting on productivity and satisfaction of university admission staff. *Journal of Organizational Behavior Management*, *18*, 45-68.

Wright, V. (1957). A note on the role of "guidance" in learning. *British Journal of Psychology*, *48*, 133-137.