RESEARCH ARTICLE

The Interaction Effects of Frequency and Specificity of Feedback on Work Performance

Jin-A Park\textsuperscript{a}, Douglas A. Johnson\textsuperscript{b}, Kwangsu Moon\textsuperscript{c}, and Jaehee Lee\textsuperscript{d}

\textsuperscript{a}Click Consulting, Seoul, Korea; \textsuperscript{b}Western Michigan University, Department of Psychology, Kalamazoo, Michigan, USA; \textsuperscript{c}Chung-Aug University, Department of Psychology, Seoul, Korea; \textsuperscript{d}Korea Institute of Child Care and Education, Fundamental Research and Statistics Team, Seoul, Korea

ABSTRACT

This study investigated the interaction effect between specificity (specific vs. global) and frequency of feedback (frequent vs. infrequent) on the quality of work performance. Eighty participants were recruited and randomly assigned to one of the four groups: specific and frequent feedback, global and frequent feedback, specific and infrequent feedback, and global and infrequent feedback. A 2 $\times$ 2 factorial design was adopted. Participants were asked to work on a simulated order-fulfilling task and attended 24 sessions. The dependent variable was the error rate of the completed tasks. The results showed that more frequent feedback was more effective and specific feedback was more effective than global feedback in improving the quality of performance. Furthermore, an interactive effect between feedback frequency and specificity was found. Specific feedback was more effective than global feedback when the feedback was infrequent, but global feedback was comparable to specific feedback when it was frequent.

ARTICLE HISTORY

Received 24 Jul 2018
Accepted 16 Dec 2018
Revised 18 Nov 2018

KEYWORDS

Specific feedback; global feedback; feedback frequency; feedback

Feedback has a long history of successful applications in the literature of Organizational Behavior Management (Balcazar, Shupert, Daniels, Mawhinney, & Hopkins, 1989; Bucklin, Alvero, Dickinson, Austin, & Jackson, 2000; Nolan, Jarema, & Austin, 1999; VanStelle et al., 2012). Feedback has been applied as an intervention component or as the primary independent variable itself in various organizational settings and types of performance, such as sales (Loughrey, Marshall, Bellizzi, & Wilder, 2013; Tilka & Johnson, 2018), safety (Hagge, McGee, Matthews, & Aberle, 2017; Lee, Shon, & Oah, 2014; Myers, McSween, Medina, Rost, & Alvero, 2010), quality control (Berglund & Ludwig, 2009; Choi, Johnson, Moon, & Oah, 2018; Goomas & Ludwig, 2017), and customer service (Eikenhout & Austin, 2005; Reetz, Whiting, & Dixon, 2016; So, Lee, & Oah, 2013).

Despite this distinguished history, it would be inappropriate to elevate feedback to the status of a principle of behavior (Peterson, 1982). Feedback is simply a stimulus and like any stimulus it can come to serve any number of...
functions as an antecedent (e.g., discriminative stimulus, conditioned motivating operation) or consequence (e.g., conditioned reinforcer, conditioned punisher) for behavior (Alvero, Bucklin, & Austin, 2001; Johnson, 2013; Johnson, Rocheleau, & Tilka, 2015; Prue & Fairbank, 1981). Just as there is great variation to the potential functions of feedback, there is great variation to the potential forms feedback, which in turn has led to a great variation in the recommendations for the delivery of feedback. Among the recommendations is that the more specific feedback is, the more effective it will be (Braksick, 2007; Daniels & Bailey, 2014; Kopelman, 1986). Williams and Geller (2000) classified feedback based on specificity into two types: specific and global. Specific feedback is defined as “the percentage of safe behavior occurrence over a given period of time for a certain target behavior … whereas global feedback is defined as an overall safety score based on the percentage of safe work practices over a given time period across a certain number of behaviors” (p. 136). Although these classifications are tied to safety performance, the concepts of specific and global feedback can be broadened to apply to non-safety related workplace implementations as well. Specific feedback can instruct workers on how to execute a task and provide helpful information in order to correct inappropriate task performance. In addition, specific feedback can reduce uncertainty about how well or poorly an individual is performing on a task (Ashford, 1986; Ashford & Cummings, 1983).

However, there have only been two published studies (Lee et al., 2014; Williams & Geller, 2000) that compare the relative effects of specific and global feedback on performance. Lee, Shon, and Oah examined the relative effects of two types of feedback on safety performance and compared the generalization of such feedback to non-targeted safety items at a construction site. The results from this study revealed comparable safety performance levels under the two feedback conditions. However, global feedback produced more of a generalized effect than specific feedback on non-targeted items. The results from Williams and Geller’s study indicated that when social comparison feedback was provided, safety performance was comparable for both specific and global feedback; whereas specific feedback was more effective in improving performance than global feedback when social comparison feedback was not provided. These results suggest that specific feedback is not always more effective than global feedback. The contradictions between the results of empirical studies and common recommendations, as well as the fact that few studies have investigated this topic, warrants further research.

Furthermore, it may the case that other variables may modulate the effectiveness of specific feedback. For example, Daniels and Bailey (2014) suggested that specific information is not always necessary because the “more immediate the reinforcer, the less specific it has to be” (p. 198). Immediate
feedback may be effective even though it does not include specific information because the performer can easily rely on a history of successful performance without having to identify behavior that needs change. Increasing the frequency with which feedback is delivered will typically reduce the temporal gap between performance and feedback regarding that performance. This relates to another general recommendation regarding the delivery of feedback: The more frequent the delivery of feedback, the more effective it will be. However, once again, the empirical evidence does not always cleanly align with conventional wisdom. For example, Balcazar, Hopkins, and Suarez (1986) found similar results comparing the effectiveness of daily and weekly feedback (42% and 41%, respectively), yet these frequencies yielded more consistent effects than monthly feedback (13%). Alvero et al. (2001) reported different results. They found that the percentage of consistent effects for monthly feedback (80%) was higher than the percentages for both daily and weekly feedback (71% and 52%, respectively). In addition, the results of previous studies directly comparing the relative effectiveness of different frequencies of feedback on performance have been mixed (Alavosius & Sulzer-Azaroff, 1990; Chhokar & Wallin, 1984; Kang, Oah, & Dickinson, 2003; Leivo, 2001; Mason & Redmon, 1992; Pampino, MacDonald, Mullin, & Wilder, 2003; So et al., 2013; Van Houten, Nau, & Marini, 1980).

For example, So et al. (2013) compared the relative effects of daily and weekly feedback on customer service at a gas station. Their results indicated that service behaviors improved when weekly feedback was introduced and improved further still when daily feedback was introduced. In contrast, Van Houten et al. (1980), Chhokar and Wallin (1984), and Leivo (2001) all found that varied frequency of feedback did not result in a difference in performance. Specifically, Chhokar and Wallin examined the relative effects of weekly and biweekly feedback, along with training and goal setting, on safety behavior in a manufacturing setting. The results indicated that weekly feedback did not produce better safety performance than biweekly feedback.

In other research (Alavosius & Sulzer-Azaroff, 1990; Kang et al., 2003), the effect frequency of feedback had on performance depended on other experimental conditions. For example, Alavosius and Sulzer-Azaroff compared the relative effects of continuous feedback, intermittent feedback, and no feedback on the acquisition and maintenance of safety behavior for healthcare staff. The results indicated that continuous feedback increased the acquisition of behavior more than intermittent feedback. However, there was no difference in the maintenance of safety behavior with either type of feedback. Kang et al. examined the effects of two frequencies of feedback on work performance under hourly and incentive pay conditions. Their results indicated that more-frequent feedback produced better performance than less-frequent feedback for the incentive pay groups. However, no performance difference was found for the hourly pay groups.
The inconsistency of results within the feedback literature highlights the need for component analyses to isolate and compare the various options for the delivery and composition of feedback (Johnson, 2013). Different implementations of feedback may alter the function of feedback and evoke different kinds and levels of performance. Regarding specific feedback, it is possible that corrective feedback may function as a reflexive conditioned motivating operation (CMO-R) in that the feedback establishes conditions that evoke behavior to remove such conditions (Johnson & Akpapuna, 2018; McGee & Johnson, 2015; Michael, 2004). For example, corrective feedback may establish a threat condition that is correlated with a progressive worsening if performance does not improve. However, global feedback may involve a different type of CMO-R than specific feedback and thus evoke different behaviors. To put it loosely, global feedback may involve a notification that something is wrong and as a result an employee engages in a wide variety of general improvements (which may or may not involve the performance of interest to the supervisor), whereas specific feedback may involve a notification that this is wrong and as a result the employee focuses their performance improvement efforts on a specific set of behaviors. In this manner, specific feedback fosters a precise calibration of performance whereas global feedback fosters a broad calibration of performance. This may account for the greater generalized effects of global feedback but implies a risk that the performance that is of primary interest may be neglected. As a consequence for behavior, specific feedback may serve to reinforce a narrower class of behaviors than global feedback. Of course, this can be problematic if the breadth of global feedback leads to a strengthening of irrelevant or undesirable behaviors.

Feedback delivered with great frequency may operate in a somewhat similar manner. As noted earlier, corrective feedback could potentially operate as a CMO-R. One feature of a CMO-R is that it includes a self-terminating feature (Johnson & Akpapuna, 2018), in that it evokes behaviors that will lessen or remove the motivating operation (i.e., the employee works to eliminate the warning and the correlated threat). Once the value-altering properties of the CMO-R are weakened, the behavior-altering properties will also cease until the CMO-R is re-established (i.e., the threat is gone until a new threat is delivered). Frequent corrective feedback may involve a more frequent establishment of the CMO-R, resulting in an overall higher frequency of desired performance. As a consequence for behavior, frequent feedback will result in a denser schedule of reinforcement than infrequent feedback, which may be an important consideration if one is dealing with a well-learned or new performance given the differences in effectiveness of continuous reinforcement and intermittent reinforcement for the development and maintenance of behavior.

Of course, many other potential behavioral functions of specific or frequent feedback should not be ruled out (SD for reinforcement or
punishment, CMO-T, punisher, etc.). Not only are component analyses important for resolving inconsistencies, they may be important for practical reasons as well. There are often time and financial costs involved in the implementation of feedback, whether it is frequent or specific in nature. Even though evolving technology has mitigated some of the costs in delivering feedback (Goomas & Ludwig, 2009, 2017), organizations still need to invest resources when they commit to implementing systems that standardize the delivery of feedback. These investments may involve the purchase of equipment to facilitate assessment and communication, managerial time for monitoring and evaluation, etc. If the components of feedback have an additive effect without placing an undue burden on the organization, such investments may be justified. Otherwise, a simpler intervention package may be warranted. The present study seeks to illuminate these considerations by examining the interaction effect between the specificity and frequency of feedback on the quality of work performance.

**Method**

**Participants and setting**

A total of 80 undergraduate students (47 males and 33 females) with a mean age of 21.5 ($SD = 1.58$) from a university in South Korea participated in this study. They were recruited via advertisements posted on the online university bulletin board. The experiment was conducted in campus computer laboratories containing 60 personal computers with the same specifications and operating system.

**Experimental task**

An order fulfillment task for a computer-simulated distribution center was developed for this study. As seen in Figure 1, this software would display icons representing stock to select from, shipping pallets, and order sheets listing the specifications required for the current order.

As part of completing the displayed order, participants would drag and drop stock items to a pallet. Stock items were organized in groups of 10 items that were placed in one of 10 district locations for a total of 100 stock items to be selected from. An order sheet was placed in the center of the screen, with odd-numbered districts located on the left and even-numbered districts located on the right. The sheet included the information that participants needed to perform the tasks: location number where the product belongs (LOC), the name of the product stacked in each item (PROD), the number of products that must be loaded onto the pallet (QTY TO PICK), the product weight (WEIGHT), and the fragility mark of the product (CAUTION). When
the mouse pointer was hovered above the box of products, it was possible to view the full product name.

Successful completion of this task required three important steps. The first step was to read the information on the order sheet. The second step was to decide on the order in which the items should be loaded onto the pallet, considering the CAUTION and WEIGHT labels. The five items listed on the order sheet had to be stacked according to the following rules in order to avoid damage:

1. An item labelled CAUTION (O) must be stacked last, no matter what its weight.
2. Items are to be stacked in descending order according to weight.
3. If two items have the same weight, they are to be stacked in ascending order of the item number.

The third step was to load the items onto the pallet. The loading method was to drag the item onto the shipping pallet icon on the bottom center of the screen and then click the “Complete an item selection” button. This button was activated if any five items were loaded. New tasks were started after clicking the “Complete an order” button. Participants were asked to complete five order sheets per session, thus processing 25 items and five pallets. The participants’ performance results were saved automatically by the simulation program.
**Dependent variables**

The dependent variable was the average rate of six errors that could occur during task performance. Similar to the loading errors measured in Berger and Ludwig (2007), the current study defined errors according to the following criteria:

1. **Omit**: Failing to load listed items onto the pallet
2. **Mispick**: Loading unlisted items onto the pallet
3. **Shortage**: Loading fewer products of each item onto the pallet than were listed on the order sheet
4. **Over**: Loading more products of each item onto the pallet than were listed on the order sheet
5. **Weight damage**: Loading items onto the pallet without complying with stacking rule
6. **Drop damage**: Dropping product outside the pallet area

The above six errors were automatically recorded by the simulation program. The proportion of each error was calculated as (number of items with each error(s)/total number of items) x 100. The average rate of error was calculated as \( \frac{\sum \text{Proportion of each error}}{6} \).

**Independent variables and experimental design**

The independent variables of this experiment were frequency (Frequent/Infrequent) and specificity (Specific/Global) of feedback. Similar to the definition used by Kang et al. (2003), one feedback delivery at the end of each experimental session was defined as frequent feedback, whereas one feedback delivery at the end of every fourth experimental sessions was defined as infrequent feedback. Using Williams and Geller (2000) as a model, specific feedback was defined as offering feedback about each of six recorded errors (e.g., a total of 25 items are executed, with error rates as follows: Omit-40%, Mispick-35%, Shortage-20%, Over-0%, Weight Damage-10%, Drop Damage-0%), whereas giving the average rate of six errors (e.g., a total of 25 items are executed, and the average rate of six errors is 13.33%) was defined as global feedback. A 2 × 2 factorial design was adopted and composed of the four feedback conditions: frequent & specific feedback, frequent & global feedback, infrequent & specific feedback, and infrequent & global feedback. Feedback was presented by the computer automatically.

**Procedure**

All participants attended a 30-minute orientation prior to the experiment. During this orientation, participation consent forms were collected and the
experimenter explained how to operate the simulated work task. The participants could review the instructions by examining printouts that were provided in advance. Participants were also given a chance to ask any questions. After the verbal explanation, a visual demonstration was shown through a projector. Subsequently, the participants were given 10 minutes to practice the assignments themselves with another opportunity to ask questions. This orientation concluded when participants could correctly complete 4 out of 5 assignments. All participants were offered an equal participation fee of 20 and informed that the fee would be provided at the completion of the experiment.

After the orientation session, participants took part in a total of 24 experimental sessions, with 8 sessions each day according to the participants’ individual schedules. The length of each experimental session was defined by the completion of five pallets, rather than by a fixed length of time, in order to make the task more similar to order fulfillment work in the field. Although there was no time limit for conducting the task, the participants were still asked to complete the task as quickly as possible. All participants performed work tasks in the same computer laboratory, but there was no interaction between the participants, even when working simultaneously, due to being placed far apart. The computer programs were set to present feedback automatically in accordance with the experimental conditions for each group. When each participant had completed eight experimental sessions, the researcher gathered the information saved on the computer. The participants were permitted to take a break, use the lavatory, and use mobile phones, messengers, and computer games whenever they pleased during the experiment.

Results

Table 1 shows means and standard deviations of the six errors across each experimental group. Regardless of the error type, the error rate was the highest for the infrequent and global group than for the other three groups. The mean of the six error types for participants in the frequent/global feedback, frequent/specific feedback, infrequent/global feedback, and infrequent/specific feedback, was 5.74 (SD = 3.16), 5.67 (SD = 3.03), 9.94 (SD = 6.02), and 5.75 (SD = 3.28), respectively (see Table 1). To decide whether the error rate differed across feedback frequency and specificity, a two-way analysis of variance (ANOVA) was conducted. The main effect of feedback frequency

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Specificity</th>
<th>Omit (SD)</th>
<th>Mispick (SD)</th>
<th>Shortage (SD)</th>
<th>Over (SD)</th>
<th>Weight Damage (SD)</th>
<th>Drop Damage (SD)</th>
<th>Overall (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequent</td>
<td>Global</td>
<td>9.22 (5.04)</td>
<td>7.86 (4.46)</td>
<td>10.67 (6.08)</td>
<td>.29 (.23)</td>
<td>3.01 (1.48)</td>
<td>3.37 (2.32)</td>
<td>5.74 (3.16)</td>
</tr>
<tr>
<td></td>
<td>Specific</td>
<td>9.37 (5.22)</td>
<td>8.32 (3.89)</td>
<td>10.33 (5.68)</td>
<td>.19 (.15)</td>
<td>3.30 (1.52)</td>
<td>2.51 (1.86)</td>
<td>5.67 (3.03)</td>
</tr>
<tr>
<td>Infrequent</td>
<td>Global</td>
<td>17.97 (15.25)</td>
<td>10.94 (4.07)</td>
<td>20.09 (16.68)</td>
<td>.42 (.38)</td>
<td>3.91 (1.54)</td>
<td>7.20 (13.68)</td>
<td>9.94 (6.02)</td>
</tr>
<tr>
<td></td>
<td>Specific</td>
<td>10.58 (8.91)</td>
<td>8.53 (5.47)</td>
<td>11.95 (9.63)</td>
<td>.23 (.30)</td>
<td>2.72 (1.90)</td>
<td>1.46 (1.38)</td>
<td>5.75 (3.28)</td>
</tr>
</tbody>
</table>
and specificity was statistically significant ($F(1, 76) = 5.52, p = .02, \eta^2 = .07$; $F(1, 76) = 5.48, p = .02, \eta^2 = .07$, respectively). The interaction effect was also statistically significant, $F(1, 76) = 5.15, p = .03, \eta^2 = .06$. Figure 2 displays the pattern of the interaction effect.

An analysis of the simple main effect was conducted to measure differences in error rates between global and specific feedback for both the frequent and infrequent feedback groups. For the frequent feedback group, the effect of feedback specificity was not statistically significant ($F(1, 76) = .00, p = .96, \eta^2 = .00$). For the infrequent feedback group, however, a significant difference between global and specific feedback was found ($F(1, 76) = 10.63, p = .00, \eta^2 = .12$).

Figure 3 displays the average error rates for specific and global feedback over time in the frequent feedback group. Participants with the two types of feedback showed similar error rates during all experimental sessions. However, in the infrequent feedback group, participants given specific feedback consistently showed a lower error rate than those given global feedback (see Figure 4).

**Discussion**

This study was designed to examine the possible interaction effects of feedback specificity and frequency on the quality of performance. The results of a two-way ANOVA indicated that frequent feedback was more effective for improving performance than infrequent feedback. This result corresponds with previous studies which found frequent feedback improves work performance (Mason & Redmon, 1992; Pampino et al., 2003; So et al., 2013).

For feedback specificity, specific feedback was more effective than global feedback, as also corresponds with the general assumption that specific feedback is more effective. Specific feedback provides a variety of information about what the performer is doing well and what needs improvement. Thus, specific feedback may have been much more helpful than global feedback in

![Figure 2](image-url). Means and standard deviations of error rates across feedback frequency and specificity.
identifying the behaviors the performer needs to change. Due to the unfamiliar nature of the experimental tasks for the participants, specific feedback would have been more effective than global feedback in mastering this task.

Above all, the interaction effect is noteworthy. The frequent feedback group showed no difference in the quality of performance for either type of feedback specificity, but the infrequent feedback group performed better when provided with specific feedback. In the infrequent feedback group, participants who received specific feedback performed consistently better during all 24 sessions than did participants who received global feedback. These findings, along with the previous findings from Williams and Geller

Figure 3. Average error rate by feedback specificity across sessions in the frequent-feedback groups.

Figure 4. Average error rate by feedback specificity across sessions in the infrequent-feedback groups.
(2000) and Lee et al. (2014), clearly indicate that specific feedback is not always more effective than global feedback. That is, specific feedback may be unnecessary when feedback can be delivered frequently.

Taken together, these results have important practical implications, particularly from a performance management perspective. If managers find it difficult to monitor employee performance frequently or to automate observations, they should observe specific behaviors when possible and provide specific performance feedback. If specificity proves difficult, then frequent feedback may achieve the same effect. In short, it is critical that feedback be specific or frequent, but one should not expect additional performance gains from embedding both elements into feedback.

One particularly interesting finding from this study is that the error rate tended to increase slightly in the later sessions for the frequent feedback groups, regardless of feedback specificity. This could suggest that frequent feedback cannot be expected to indefinitely maintain performance and may need to be combined with other components over time (e.g., evaluative feedback, praise, tangible reinforcers, etc.). Alternatively, it may also suggest that there could be a benefit to eventually changing from frequent to infrequent feedback as employee proficiency improves, much as there is often a benefit to switching from continuous to intermittent reinforcement with new learners.

The results of this study suggest several opportunities for future research. First, different specificities of feedback should be examined with well-learned behaviors as opposed to relatively new behaviors. Alavosius and Sulzer-Azaroff (1990), Chhokar and Wallin (1984), and Leivo (2001) all found that different feedback frequencies maintained behaviors equally well once the behaviors had been acquired and mastered. Although these results are consistent with statements that frequent feedback is less important for maintenance than for the acquisition of behaviors (e.g., Alavosius & Sulzer-Azaroff; Fairbank & Prue, 1982), it is not known whether results would differ if feedback with different specificities were provided for well-learned behaviors.

Similarly, feedback frequency or specificity may interact with high and low performance levels. For example, Moon, Lee, Lee, and Oah (2017) found that the interaction effect between feedback content and performance level affected performance. Their results indicated that social-comparison feedback was more effective than objective feedback for the high performers but was less effective for the low performers. High performers already know the rules and behaviors that are necessary for their tasks, so it is less likely that their performance will change because of the feedback frequency or specificity. However, specific and frequent feedback can help improve performance for low performers who are not familiar with the rules or what behaviors are needed for performing a task.

Second, task difficulty may also interact with feedback specificity. For difficult and complex tasks, specific feedback can be effective in improving
performance, but with easy tasks, global feedback will be sufficient to improve performance. Finally, in this study, feedback on incorrect responses (error rate) was provided. Daniels and Bailey (2014) argued that feedback should be based on correct responses rather than incorrect responses. Sigurdsson and Ring (2013) evaluated the preferences of undergraduate students for graphic feedback on percentage of incorrect performance or on percentage of correct performance. Their results indicated that, although the type of feedback students received for the first four quizzes did not affect subsequent quiz performance, participants preferred feedback on correct performance rather than on incorrect response. However, in the OBM literature, feedback delivered on incorrect responses (e.g., Bateman & Ludwig, 2003; Berglund & Ludwig, 2009; Sasson, Alvero, & Austin, 2006) has been found to be effective for improving performance. Therefore, more research is needed on the relative effects of correct- and incorrect-response feedback on performance and preferences.

Although these results demonstrate the effectiveness of feedback frequency, specificity, and the interaction between these two characteristics of feedback on work performance, there were limitations concerning the implementation of this study. One limitation was defining a session as completing 25 items rather than as a fixed amount of time in order to make it similar to the performance in an actual logistics center. Average duration time per session was 5.71 min. \((SD = .50)\), and there was no difference between the four experimental groups \((F(3, 76) = .276, p = .843)\). However, there was a difference in the average duration of sessions for each individual (Range: 4.88–7.08) that would likely generate a difference in productivity in the actual field. Another limitation is that the task was a simulation of a work environment and therefore may not fully capture the actual implementation of tasks in a modern distribution center or the subtleties associated with real work performance.

Despite such limitations, the current study provides evidence regarding the impact feedback has upon performance when considering characteristics such as the frequency and specificity with which the feedback is delivered. Although feedback in general has been studied often, the finer details of feedback implementation are less well understood but may be just as important. To establish a more in-depth understanding of the precise effects of feedback on performance, it is recommended that future studies be conducted with greater frequency and with more specific components under investigation.

**Disclosure statement**

No potential conflict of interest was reported by the authors.
References


