

## **Objective Performance Feedback: Is Numerical Accuracy Necessary?**

MICHAEL G. PALMER and C. MERLE JOHNSON  
*Central Michigan University, Mount Pleasant, Michigan, USA*

DOUGLAS A. JOHNSON  
*Western Michigan University, Kalamazoo, Michigan, USA*

*Performance feedback is ubiquitous in Organizational Behavior Management (OBM); yet its essential components are still debated. It has been assumed that performance feedback must be accurate, but this assumption has not been well established. Two experiments were carried out to research feedback accuracy. Experiment 1 was a single-subject design where performance feedback accuracy was manipulated. Results from Experiment 1 suggested feedback may not need to be accurate to improve performance prompting a follow-up study. Experiment 2 was a repeated measures between-groups design with three types of objective feedback: accurate, high (triple) and low (1/3) inaccurate, and no feedback control. Both accurate and tripled feedback significantly improved performance over the control and low-inaccurate feedback groups. Performance feedback may have reduced time off-task across all three feedback conditions compared to the control. Data from performance feedback research need multi-faceted analysis to fully understand how and why performance feedback changes behavior.*

**KEYWORDS** *Performance Feedback, Feedback Accuracy, Objective Feedback*

---

Address correspondence to Michael G. Palmer, Department of Psychology, Central Michigan University, 101 Sloan Hall, Mount Pleasant, MI 48859, USA. E-mail: [palme1mg@cmich.edu](mailto:palme1mg@cmich.edu)

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/WORG](http://www.tandfonline.com/WORG).

Organizational behavior management (OBM) interventions utilize a variety of stimuli to change targeted behaviors in diverse settings. These stimuli have ranged from antecedents such as task clarification to consequences such as piece-rate pay (Frederiksen, 1982; C. M. Johnson, Redmon, & Mawhinney, 2001; Luthans & Kreitner, 1985; O'Brien, Dickinson, & Rosow, 1982). Other stimuli are harder to analyze in the three-term contingency, such as goal setting and feedback, because of varied usage and the manner in which they are presented. These stimuli are often combined into an intervention package in an attempt to raise the probability of success. Although successful packaged interventions are allowing the practice of OBM to gain in efficacy and social validity, they are also hindering the ability to do thorough analyses of the individual elements within the package (Filipkowski & Johnson, 2008).

Just as these intervention packages may be composed of many individual elements that merit investigations into the relative contribution of each element, the individual stimuli themselves may consist of many components also worth investigating individually, especially if that stimulus is potentially complex or broad. One such commonly used stimulus is performance feedback. The first issue of the *Journal of Organizational Behavior Management (JOBM)* contained an article looking at the effects of feedback (Kreitner, Reif, & Morris, 1977). For the next 25 years performance feedback was used in 65%–70% of all articles published in *JOBM*, and it was effective in nearly all of its applications (Alvero, Bucklin, & Austin, 2001). A recent review (VanStelle et al., 2012) found feedback to be the most frequent independent variable in OBM studies (p. 112). Performance feedback continues to be used in both lab and applied settings in OBM (Balcazar, Hopkins, & Suarez, 1985–1986), yet the essential characteristics of performance feedback remain elusive (D. A. Johnson, 2013; Peterson, 1982).

At the beginning of the 21st century, feedback had been used in 65 applications within 43 different studies; therefore, a solid definition of feedback seems plausible (Alvero et al., 2001). Unfortunately, performance feedback has been defined in multiple ways, all of which usually contain similar features. Prue and Fairbank (1981) defined it as information about the quality and/or quantity of past performance that is then provided to the behaving person. Daniels (2000) defined it as information about past behavior that allows a person to adjust his or her current performance, and Mayer, Sulzer-Azaroff, and Wallace (2013) defined it as information given to an individual after his or her behavior. These definitions contain fragments referring to some sort of information about past occurrences of a particular person's behavior. However, although the definitions of feedback may have been similar, in the 65 different applications the stimulus labeled as feedback was slightly different. Even slight differences in implementation could potentially have strong differences in outcomes.

These groups of stimuli are all classified as performance feedback, yet functionally and topographically they may be very different. For instance,

studies have used graphic, oral, and written displays of feedback that have been delivered on various schedules (weekly, daily, etc.) by both researchers and supervisors, about group and individual performance, both publicly and privately (Alvero et al., 2001; VanStelle et al., 2012). Even after these varied uses of performance feedback there continues to be a debate over what components are necessary for behavior change (D. A. Johnson, 2013). This problem was raised in a recent issue of *JOBM* in which focus was directed from a component analysis of feedback to its application (Houmanfar, 2013; D. A. Johnson, 2013). This type of analysis on stimuli, such as feedback, should be done more often. However, component analyses are rare, and the most recent one focused on whether feedback needs to be objective, evaluative, or both (D. A. Johnson, 2013).

Participants in that group design study received objective feedback, evaluative feedback, combined objective and evaluative feedback, or no feedback while completing a simulated bank check-processing task over four sessions. *Objective feedback* was defined as specific unbiased information about past behavior, such as “You processed 600 checks today.” *Evaluative feedback* was defined as subjective information based on past behavior, such as “You did a great job today.” Feedback on the previous session’s performance was delivered vocally by the researchers prior to the next session. Those who received both objective and evaluative feedback improved performance in an additive summation effect compared to those who received only one of those components, who performed similarly across sessions. All three feedback groups performed significantly better than the no feedback group. Results suggested that combined evaluative and objective feedback, which takes supervisors only a few more moments to administer, results in the largest increases in performance (D. A. Johnson, 2013). However there are other characteristics, in addition to the objective and evaluative nature of the feedback, that warrant attention. Clearly more research is needed to clarify the necessary and sufficient characteristics of performance feedback in organizational settings.

One such characteristic is accuracy. Inaccurate feedback has been used in research outside of behavior analysis for years. Most research that has used inaccurate feedback has been interested in participants’ reaction to the inaccurate feedback, not performance of the behavior itself (Beedie, Lane, & Wilson, 2012; Venables & Fairclough, 2009). Other researchers have been interested in more physiological responses to inaccurate feedback, such as heart rate, oxygen uptake, or tension headache reduction (Faulkner, Arnold, & Eston, 2011; Kondo & Canter, 1977). Many experiments have used inaccurate biofeedback as a control condition because inaccurate biofeedback does not change physiological responses in a desired direction (Kondo & Canter, 1977; Nestoriuc, Martin, Rief, & Andrasik, 2008).

The knowledge of results literature also has reference to inaccurate performance feedback. However, these studies usually involved a signal

detection/vigilance task rather than tasks encountered in typical workplaces and often used a light to deliver feedback as opposed to verbal responses used in most performance feedback research (Mackworth, 1964; Weidenfeller, Baker, & Ware, 1962). Moreover, this literature reported conflicting outcomes. Inaccurate feedback is detrimental to performance (Mackworth, 1964) or does not produce significantly different results from accurate feedback (Antonelli & Karas, 1967; Peretti, 1970; Weidenfeller et al., 1962). Although this line of research looked promising, knowledge of results did not provide strong evidence to conclude whether feedback should be accurate.

Performance feedback should be accurate or precise and objective (Prue & Fairbank, 1981); that is, it should reflect the actual performance of the individual (Daniels, 2000). All studies in reviews of performance feedback adhere to these guidelines (Alvero et al., 2001; Balcazar et al., 1985–1986; VanStelle et al., 2012). Moreover, as Daniels and Daniels (2004) noted, most performance feedback in motor skills acquisition is automatic (e.g., golf shot). Indeed, when one is teaching new behaviors immediate feedback should be accurate (Hirst, DiGennaro-Reed, & Reed, 2013; Mackworth, 1964).

Recently Hirst et al. (2013) manipulated performance feedback accuracy for participants engaged in a conditional discrimination task. Undergraduate students received immediate feedback as to whether their conditional discrimination was correct or incorrect; however, in some phases participants received inaccurate feedback that they were correct/incorrect in 25%, 50%, or 75% of trials. Those who received inaccurate feedback did not emit correct responses, which were a function of the proportion of inaccurate feedback they received. Hirst and DiGennaro-Reed (2015) replicated and then extended these results to preschool-age children in a school setting. These results suggest that immediate performance feedback should be accurate for proper task acquisition, supporting OBM recommendations over many decades.

In the task acquisition studies by Hirst et al. (2013) and Hirst and DiGennaro-Reed (2015), and in most of the knowledge of results literature (Antonelli & Karas, 1967; Mackworth, 1964; Peretti, 1970; Weidenfeller et al., 1962), feedback was provided immediately after every instance of the behavior. In these situations, a direct acting contingency is in place. However, in most workplace settings and many OBM applications feedback is often delayed, an aggregate of some kind, and not related to acquisition of a task (Alvero et al., 2001; VanStelle et al., 2012). Within a job training context, it is possible that feedback may be delivered as an immediate consequence to modify a trainee's behavior. However, during most on-the-job situations feedback is more likely to function as an antecedent or delayed consequence as part of an indirect acting contingency. This makes it difficult to apply the findings from Hirst et al. and Hirst and DiGennaro-Reed to employees and managers at work.

Malott (1992) and Weatherly and Malott (2008) hypothesized that in most workplaces performance feedback, because of its delayed and aggregate applications, sets up a rule-governed contingency that is mediated by verbal descriptions of the current contingencies. Performance feedback typically is mediated by managers and supervisors in organizations, suggesting that it should be treated as verbal stimuli for the performer of interest; without a verbal repertoire, feedback would likely be a stimulus not attended to by the individual. Malott and Shane (2013) argued that most instances of performance feedback are verbal stimuli made by the supervisor prior to the employee engaging in work.

Although feedback is provided about past behavior, it may serve as an antecedent for subsequent behavior. To illustrate how feedback might involve contingency-specifying stimuli and its possible role as an antecedent within organizational settings, some examples involving feedback regarding poor performance functioning as a conditioned motivating operation may be helpful. One type of conditioned motivating operation is the reflexive conditioned motivating operation (CMO-R), which is a stimulus that is reliably correlated with some form of worsening and will evoke behaviors that result in the removal of that stimulus (Michael, 2004). The removal of the CMO-R therefore functions as a form of reinforcement for behaviors that are successful in eliminating this stimulus. In organizational settings, feedback indicating insufficient performance is often correlated with the onset of aversive stimuli (e.g., social disapproval, loss of income, negative performance reviews). Whether this involves self-generated feedback (“My performance is looking bad”) or feedback supplied by others (“Your performance is looking bad”), such negative feedback is likely to function as a CMO-R and evoke behaviors to remove it, such as improving one’s performance, hiding evidence of insufficient performance, trying to prevent efforts at performance measurement, and so on. This negative feedback can be thought of as a warning stimulus for impending aversive consequences from one’s supervisor. It is a warning that is not necessarily a single event, as the performer can continually describe his or her performance (“I’m still clearly below the performance standards”). The removal of that warning is likely to function as a form of reinforcement for relevant behaviors.

Negative feedback could also function as a single component of a CMO-R, especially when work deadlines are involved. In this case, the compound stimulus condition of insufficient performance/no time left would function as the CMO-R because the joint components of (a) having failed to perform at acceptable standards and (b) no longer having time left are both needed to be correlated with aversive stimulation. This is because the conditions of sufficient performance/no time left and insufficient performance/plenty of time left are not correlated with the onset of some form of worsening in conditions, although they may retain enough common elements with the CMO-R

to produce a slight evocative effect. Thus, the initial setting of a deadline may produce only a low level of responding on the part of the employee. However, as time passes, the stimulus conditions begin to more closely resemble the CMO-R so that the evocative properties become progressively stronger (i.e., insufficient performance/little time left), resulting in increasing levels of performance (or hiding evidence of performance levels, requesting deadline extensions, etc.). This increasing similarity of stimulus conditions to the maintaining CMO-R would explain why performance rapidly increases as the deadline approaches, a phenomenon sometimes referred to as a *procrastination scallop* or *J-curve of performance* (Daniels, 2000; Michael, 2004).

For these contingencies to be successful, it may require that negative feedback be accurate. Negative feedback provision may result in a self-generated verbal description such as “My performance may or may not be insufficient” and some similar statement if the performer has learned that the negative feedback is historically untrue. Behavior may be less likely to occur because of a decrement in the similarity of stimulus conditions between current conditions and the CMO-R. If the CMO-R ultimately relies on contingency-specifying stimuli, such as “My performance is insufficient and therefore I’m in trouble” or “My performance is insufficient and I’m out of time so I’m going to be in trouble,” then verbal stimuli such as “My performance might or might not be insufficient because you can’t trust that feedback” are dissimilar enough that the evocative properties are weakened or eliminated. As others have pointed out (D. A. Johnson, 2013; D. A. Johnson, Rocheleau, & Tilka, 2015; Mangiapanello & Hemmes, 2015), feedback can take many other stimulus functions besides that of a CMO-R, such as a reinforcer, a punisher, a discriminative stimulus, another type of conditioned motivating operation, or a combination of functions, but further analyses involving these possible functions are beyond the scope of this article. Furthermore, the analyses relevant to positive feedback might involve different functions than the analyses relevant to negative feedback. The broader point is that introducing an element of salient inaccuracy to positive or negative feedback may change the verbal descriptions of workplace contingencies and therefore alter the strength of the antecedents and consequences involved in rule-governed behavior.

It should be no surprise to readers that verbal stimuli are sometimes untrue. Often, particularly in a setting where the verbal response comes from someone with more authority than the listener, the listener may not question the verbal stimuli from the supervisor. These verbal stimuli may intentionally or unintentionally be false. A manager may intentionally provide inaccurate feedback to avoid spending extra time and effort generating immediate frequent performance feedback (e.g., “You’re doing fine”) or deliver positive feedback regardless of performance in an attempt, albeit a misunderstood attempt, to deliver positive reinforcement to everyone. Unintentionally inaccurate feedback may be generated by faulty equipment, observer

drift, poorly trained observers, sloppy on-the-job training, coworkers, or a sampling procedure that does not accurately portray work behavior. Little or no research has been conducted to see what effect inaccurate feedback, intentional or unintentional, has on performance. Clearly, more research is needed to clarify the multiple and complex roles of performance feedback in organizational settings.

Because little research exploring delayed and aggregate accurate as well as inaccurate performance feedback has been carried out, the purpose of this first study was to determine whether a stimulus describing previous performance needs to be directly reflective of previous behavior for performance to improve. In other words, does feedback on the employee's previous behavior presented by the supervisor need to be accurate to improve performance?

## EXPERIMENT 1

### Method

#### PARTICIPANTS AND SETTING

Three females from the Psychology Department subject pool at a midwestern university in the United States were recruited. Participants were randomly assigned to one of three conditions and were compensated \$80 for participating (\$20 after every four sessions, with a \$15 bonus for completing the study). In addition, they earned 13 credits that could be exchanged for extra credit in classes for which they were registered. Monetary compensation was prorated because of institutional review board concern about possible coercion if compensation was withheld until the end of the study. Sessions lasted 45 min and were carried out in laboratory settings (approximately 10 × 15 feet) on campus containing one to four computer stations. Sessions were run 3 or 4 days a week, and it took approximately 3 weeks for each participant to complete all sessions.

#### EXPERIMENTAL TASK AND ALTERNATIVE ACTIVITIES

Participants calculated and entered latency data from slips of paper into an Excel spreadsheet. The numbers were data that the experimenter was using from another experiment. An example of the type of data participants entered was "start time 4:00, finish time 4:15." Participants were to enter "15" min in the appropriate row for a correct response. The slips of paper were grouped into separate days, and participants were instructed to enter the data points in rows, rather than in columns, to prevent easy counting of how many data points they entered. Counting of columns, labeled by letters rather than numbers, was unlikely, as they would have had to be concurrently adding

up columns while calculating latencies. Participants were not told of any alternative activities besides using the restroom and drinking fountain.

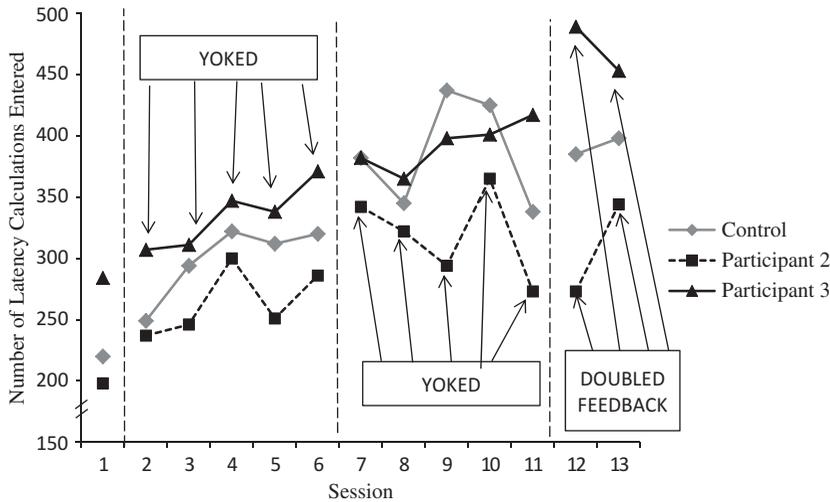
#### DEPENDENT AND INDEPENDENT VARIABLES

The number of completed latency calculations was the primary dependent variable. The control participant did not receive any feedback, whereas Participant 2 experienced accurate performance feedback for five sessions, followed by inaccurate feedback for five more sessions, followed by two sessions of doubled feedback. Participant 3 experienced inaccurate feedback, accurate feedback, followed by doubled feedback. The inaccurate feedback was generated by yoking the feedback to the control participant's performance. Therefore, if the control participant entered 250 latency calculations, the participant receiving inaccurate feedback was told that she had calculated and entered 250. The doubled feedback was the participant's performance multiplied by 2; therefore, if the participant entered 300 latency calculations, she was told that she had entered 600. Performance feedback consisted of a written number of latency calculations completed along with a graph depicting performance across sessions. Participants completed a survey after the sixth, 11th, and 13th sessions. The control participant was told that the survey had been used in a previous study and that the questions about feedback did not apply.

#### EXPERIMENTAL PROCEDURES

Prior to the first session informed consent was obtained from all participants. Following informed consent the researcher or assistant described both monetary and course credit compensation. The purpose of the data calculation and entry task was then explained. Participants were told to try to complete as many calculations as possible, accurately, and then they were told that they needed to enter at least 60 correct calculations to gain the monetary compensation for that session. The researcher then demonstrated how to enter the calculated results into the Excel spreadsheet.

The researcher then explained that the purpose of the study was to see how future pay affects the number of latency calculations processed. Participants were told not to engage in conversation with the experimenter and to focus only on their workstation. Following this introduction the participants had the opportunity to ask any questions. Then the researcher set a timer for 45 min, which participants could not see. The researcher stayed in the room with the timer. After 45 min elapsed the session was finished and the researcher thanked participants for attending and then dismissed them. The first session was identical for all participants, and the three were run separately. Participants engaged in the task for a total of 13 45-min sessions.



**FIGURE 1** Experiment 1 results. Session 1 was the baseline session. Accurate feedback was presented to Participant 2 after baseline and Participant 3 in the third phase. Data points that have arrows pointed at them indicate sessions in which inaccurate feedback was delivered.

Prior to the second session, feedback on first-session performance was delivered to participants in the two experimental conditions. Participant 2 received accurate feedback, and Participant 3 received inaccurate feedback (this participant received the performance data of the control). These conditions remained the same for five sessions. Then conditions were reversed for the last two participants; that is, the second participant received inaccurate feedback and the third participant received accurate performance feedback. The control continued to not receive feedback. These conditions remained for five more sessions. For the last two sessions Participants 2 and 3 received feedback indicating that they completed double their actual performance for the last two sessions. The control continued to not receive feedback.

## Results and Discussion

No obvious differences, based on visual inspection of the data, were found between or within participants between feedback conditions. [Figure 1](#) shows that the control participant gradually increased the number of latency calculations entered across the entire duration of the study. Both participants who received feedback performed similarly to the control participant. [Table 1](#) displays results from the survey.

Results from Experiment 1 suggested that feedback may not need to be accurate for performance gains to occur. However, it was unclear whether the inaccurate feedback was discernable to the two experimental participants. Neither participant suggested in surveys that she believed that any

**TABLE 1** Mean (*SD*) Scores on Experiment 1 Surveys Across All Participants and Administrations

Question	Survey			Grand mean
	One	Two	Three	
Payment fair	1.3 (0.6)	2.0 (1)	1.7 (0.6)	1.7 (0.7)
Set goals	1.7 (0.6)	2 (1)	1.7 (0.6)	1.6 (0.7)
Feedback accurate	1.7 (0.6)	2 (0)	1 (0)	1.6 (0.5)
Task boring	1.5 (0.7)	1.5 (0.7)	2 (0)	1.8 (0.5)
Feedback helpful	2 (1)	2 (0)	1.5 (0.7)	1.8 (0.6)
Money motivating	1.7 (0.6)	2 (0)	2 (0)	2.0 (0.3)
Tell friends	2.3 (0.6)	2 (0)	2.3 (0.6)	1.6 (0.9)
Enjoyed study			2.3 (1.2)	2.3 (1.2)
Aggregate	1.8 (0.7)	2.1 (0.7)	1.9 (0.6)	

*Note.*  $N = 3$ . Content in the "Question" column is abbreviated items from the survey. Question 8 ("enjoyed study") was not visible during the first and second administrations. Survey answers were on a 5-point Likert-type scale with 1 indicating *strongly agree* and 5 indicating *strongly disagree*, and questions were as follows: (1) "The amount I was paid to participate in the study was fair," (2) "I set goals for myself when entering checks each day," (3) "The performance feedback I received was accurate," (4) "I found the check entry task boring," (5) "The feedback on my previous performance was helpful," (6) "The compensation promised at the end of the study motivated me to continue," (7) "I would recommend participating in this study to a friend," and (8) "Overall, I enjoyed participating in this study."

of the feedback she had received was inaccurate. All three performed similarly across time. Because the feedback was inaccurate by 20–50 latency calculations from a base rate of at least 200, this may have been below the threshold to be detectable. This interpretation has validity from the survey data, in that both participants receiving feedback indicated that they believed it was accurate during the entire experiment.

This may support Ilgen, Fisher, and Taylor's (1979) suggestion that feedback must be accepted to be effective, as both participants indicated that the feedback they received was accurate. However, because neither indicated that the feedback was inaccurate, a thorough analysis could not be made. Moreover, the performance feedback may have been accepted by participants even though it was inaccurate. Because the control did not receive any feedback and had comparable performance, it may be that the other two did not attend to the feedback, and the results may be explained by task acquisition alone. Indeed, there were no measures of participant attending behavior to the feedback in this experiment.

The task also could have influenced the results; it had not been used in research before and may have been slightly more difficult than tasks previously used. This may have reduced or increased variability between participants' performance. More research with this task would clarify this. Researchers were not able to run phases out to steady state, leaving each phase with a slightly increasing trend that continued into the next phase. Having this increasing trend throughout the study makes it difficult, if not

impossible, to determine whether accurate and inaccurate feedback had any effect on how many latency calculations were entered.

Finally, it could have been possible that because the experimenter was in the room during sessions, a reactivity effect was influencing results. Participants were not told beforehand of any alternative activities besides using the restroom and drinking fountain. Although participants could have engaged in other activities, such as using an Internet browser or playing a game on the computer, none did. It is postulated that alternative activities were not engaged in because the experimenter was in the room and visible during each session. These weaknesses resulted in designing a second experiment.

## EXPERIMENT 2

To eliminate the need to carry phases out to steady state, a group design was used. Moreover, some OBM researchers claim that group designs are better for comparison research because they do not have problems of multiple treatment interference (Komaki & Goltz, 2001). Next, the lack of awareness of doubled feedback inaccuracy in Experiment 1 suggested that tripled feedback would be needed to amplify possible differences between conditions. In addition, a one-third feedback condition was added to see whether underreporting performance feedback would influence results in a different manner.

Experiment 1 results suggested that sessions should be completed with the experimenter outside of the room; the presence of a researcher can influence the behavior of participants (Lebbon & Austin, 2013). After completing Experiment 1 the researchers believed that participants should have something to do other than the experimental task. Perhaps high rates of on-task behavior were due, at least partially, to no explicit distractions being available. Although most managers would prefer that their employees work at high rates, there are usually other competing contingencies in many workplace settings, such as opportunities to use the Internet. Therefore, alternative activities may allow for better external generalization.

Accurate and inaccurate feedback could have influenced the number of errors made by the participants. However, researchers in Experiment 1 were unable to calculate easily the number of errors produced by participants. Thus, it was decided that a simpler task should be used that calculates accuracy automatically and potentially increases variability in performance between participants. Not having this level of analysis prevented a more molecular viewpoint. Data collected on a moment-to-moment basis are needed in feedback research (D. A. Johnson, 2013). A check-processing task used in previous studies (D. A. Johnson, 2013; D. A. Johnson, Dickinson, & Huitema, 2008; McGee, Dickinson, Huitema, & Culig, 2006; Slowiak,

Dickinson, & Huitema, 2011), which simulated the job of a proof operator in a bank, had the capacity to record these data automatically and had a solid record of ensuring variability in the performance data to run analyses. Therefore, this was used in Experiment 2.

In addition, many researchers have analyzed and published their data using a molar perspective, that is, based on aggregates or patterns of behaviors across large amounts of time (W. M. Baum, 2002). Although these data are what most managers are interested in, they may not give the clearest picture of what independent variables are doing to behavior (D. A. Johnson, 2013). For instance, it is fairly well established that providing performance feedback can increase behavior output (VanStelle et al., 2012), but is that because it increases the rate of responding, decreases errors, or keeps employees from taking breaks? A molecular analysis can provide data that can answer these questions. Indeed, D. A. Johnson (2013) suggested that future research evaluate the measures used in feedback research to account for time on task, accuracy of responding, and rates of responding. Experiment 1 did not allow for a molecular analysis of the participants' behavior; Experiment 2 was designed to explore this level of analysis both between and within sessions.

## Method

### PARTICIPANTS AND SETTING

Sixty participants (17 males and 43 females) from the Psychology Department subject pool at a midwestern university in the United States were recruited. Participants were randomly assigned to one of four groups and were compensated \$25 for participating (\$10 after the second session and \$15 after the last session). In addition, they earned 12 credits that could be exchanged on the Psychology Department subject pool website for extra credit in classes for which they were registered. Monetary compensation was prorated because of institutional review board concern about possible coercion if compensation was withheld until the end of the study. Sessions lasted 45 min and were carried out in laboratory settings on campus containing one to four computer stations, similar to Experiment 1. Rooms with multiple computers had computers separated by vertical dividers, ensuring that participants were unable to see others' computer screens. One to four participants were run at a time; the average was 2.7 ( $SD = 0.9$ ).

### EXPERIMENTAL TASK AND ALTERNATIVE ACTIVITIES

Participants engaged in a computerized data entry task similar to the job of a proof operator (check processor) in a bank. Similar to D. A. Johnson's (2013) experiment, the verbal feedback was not delivered by the computer.

The only feedback from the simulated task was the immediate change that a particular check was completed when it disappeared from the screen and a new check was shown. The data entry task presented checks with values ranging from \$10.00 to \$999.99. Participants entered the amount using a keyboard into a separate area of the program. The program automatically recorded the number of correct and incorrect check values, and each entry was time stamped; these data were not visible to participants.

Participants had access to other computer programs (an Internet-browsing program as well as games such as Minesweeper and Solitaire). The participants were able to engage in these programs during their session and were able to return to the data entry task at any time. These were available to give participants off-task activities to approximate many natural working environments. The software did not record what off-task activities participants engaged in but did record time spent off task, defined as at least 10 s not engaging the software.

#### DEPENDENT AND INDEPENDENT VARIABLES

The number of correctly completed checks and the total time off task were the primary dependent variables. Performance during the first session served as a covariate in the data analysis. The independent variable, the type of performance feedback presented, was divided into four categories: accurate feedback, one-third inaccurate feedback, tripled inaccurate feedback, and no feedback from the experimenter. A 4 (type of feedback received)  $\times$  5 (each 45-min-session in which feedback was presented) mixed factorial design with repeated measures on the second factor was used. Participants were randomly assigned to one of the four experimental conditions, with 15 per group, prior to the first session. A total of six sessions was required for each participant. Only one session was run per day for each participant. Sessions were carried out 3 days per week for each participant.

Control for possible experimenter bias was attempted by having research assistants run an equal number of participants from each group. The two research assistants were an undergraduate and a graduate student. Research assistants were required to have had at least two courses in research methods and one course in behavior analysis. Both research assistants were trained by the first author individually on how to conduct all six sessions.

Surveys also were administered at the end of the second and last sessions to gauge the participants' reaction to the experiment and to determine whether feedback was perceived as inaccurate. These were the same surveys as in Experiment 1, with the exception of Question 8 ("Overall, I enjoyed participating in this study") being present during each administration (see [Table 1](#) for survey items). Similar to Experiment 1, control participants were told that the survey was reused from a previous study and that some of the

questions on the survey were not relevant to them. No direction was given on whether to skip or answer the questions. The surveys served as secondary dependent variables.

#### EXPERIMENTAL PROCEDURES

Prior to the first session informed consent was obtained from all participants. Following this, the researcher or assistant described monetary and course credit compensation. Participants were then assigned a timecard with their subject number and upcoming sessions. The purpose of the check-proofing task was then explained, and participants were told to try to complete as many checks as possible, as accurately as possible, which simulated the job of a bank check processor. They were told they needed to enter at least 100 correct checks in each session to earn pay for that session (minimum work requirement). The researcher then demonstrated how to use the data entry program as well as how to engage in alternative activities. Participants were reminded that they were not to use their cell phones during the experiment but were not required to turn them off.

The researcher explained that the purpose of the study was to see how future pay affects the number of checks processed. Participants were told not to engage in conversation with others in the room and to focus only on their workstation. This was to prevent discussion on the number of checks entered. Following this introduction, participants had the opportunity to ask questions, and then researchers took each person into the hall to sign a sign-in sheet. Participants were then asked to confirm future session dates and times and then instructed to reenter the room. Researchers set the timer for 45 min, which participants were unable to see. Researchers left the room with the timer and stood outside. After 45 min elapsed, researchers reentered the room and told participants that the session was finished. Participants were thanked and then dismissed. This first session was identical for all four groups.

Prior to the beginning of the second session, participants were asked to confirm future session dates and times outside of the experimental room. Those in the control group then reentered the room. While still outside the room accurate feedback on individual performance was delivered to participants in the three experimental groups, far enough away that others could not hear or see. Those not in the control group were informed of their performance from the previous session in vocal and graphic form, both of which were purely objective. They were then instructed to sign the bottom of each feedback sheet with their subject ID number. After withdrawing this graphic feedback sheet, the researcher and participant reentered the room and started the task. The researcher then left, and after 45 min elapsed the researcher then reentered the room and asked participants to stop the task.

The researcher then asked them to complete a short survey. Afterward participants were given their first monetary compensation in cash, were thanked for attending the session, and were told that they may leave.

Sessions 3, 4, 5, and 6 were conducted similarly to this second session for all groups with three exceptions: the type of feedback that was administered to the two inaccurate feedback groups, a second survey was administered only after Session 6, and participants were paid the second time after the sixth session. For those in the tripled feedback group, feedback for each of these sessions was triple their actual performance during the previous session. For example, if participants entered 400 correct checks in their third session, they received feedback at the beginning of the fourth session indicating that they entered 1,200 correct checks. For those in the one-third feedback group, feedback for each of these sessions was one third their actual performance from the previous session. Therefore, if participants entered 400 correct checks in the third session, they were told that they entered 133 correct checks, shown the graph, and they signed it before beginning the fourth session. In summary, in the first session participants did not receive feedback; in the second session feedback was accurate for all three feedback conditions; and in the remaining sessions the different independent variables were manipulated, yielding four sessions of differential exposure to accurate and inaccurate feedback.

## Results and Discussion

Four of the 60 participants dropped out of the study. These were evenly distributed across all four groups. Each experimenter had at least one dropout, and the primary researcher had two. Of these four participants, one quit after the first session, two quit after the third session, and one quit after the fifth session. None of their data were included in the analyses.

### NUMBER OF CHECKS ENTERED CORRECTLY

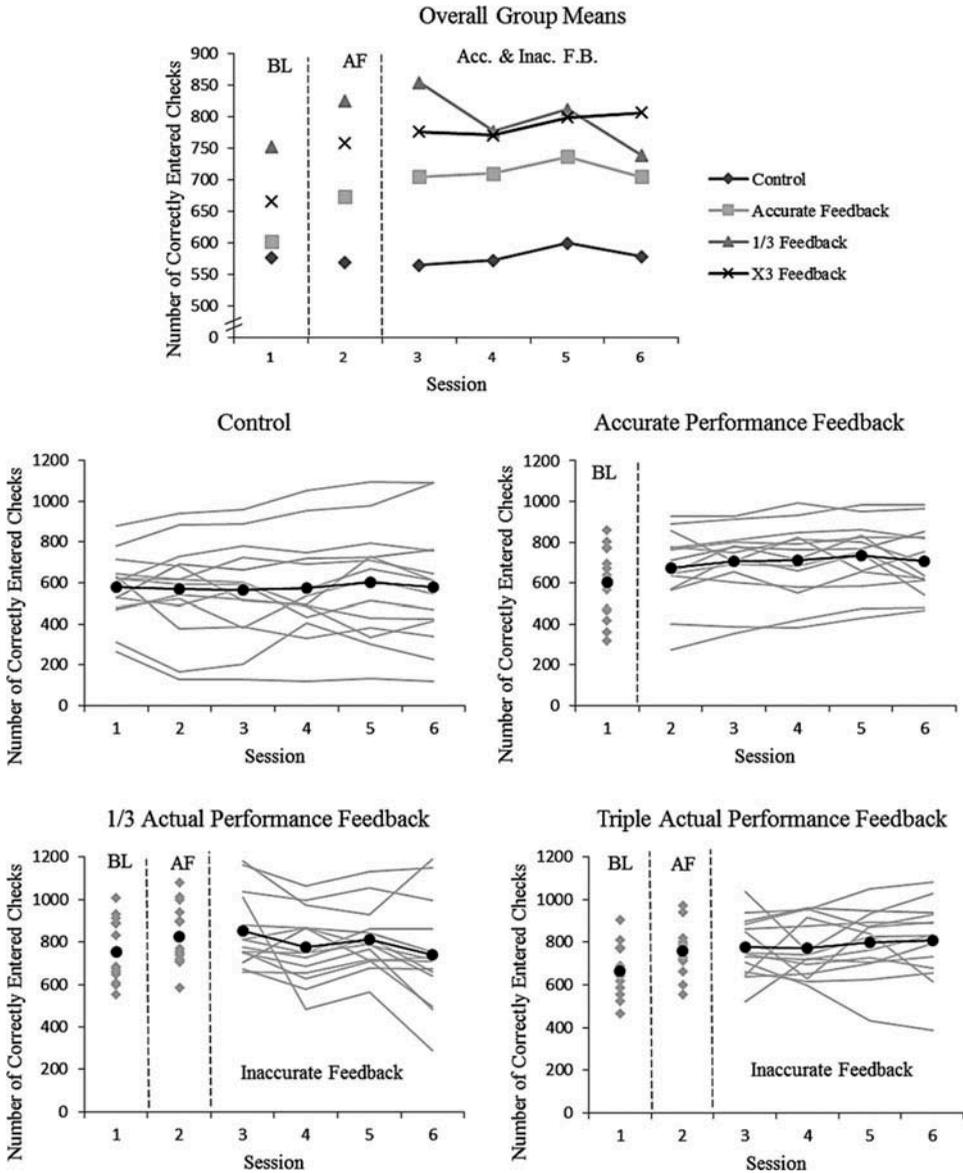
An analysis of variance was conducted on data from the first session to examine whether there were initial group differences in performance. The results showed a significant difference,  $F(3, 52) = 3.76, p = .016, \eta^2 = .178$ . Thus, an analysis of covariance (ANCOVA) was used to adjust for these differences in performance during the first session. The number of checks completed correctly during the five remaining sessions was analyzed using a repeated measures ANCOVA. The number of correctly completed checks during the first session served as the covariate. There was a significant effect of the different types of feedback on performance after initial first-session performance was controlled,  $F(3, 51) = 3.47, p = .02, \eta^2 = .17$ . There was no main effect of the sessions on performance,  $F(3.19, 162.59) = 0.88, p = .46$ ,

$\eta^2 = .02$ . There was no interaction between the sessions and the different types of feedback,  $F(9.56, 162.59) = 1.05$ ,  $p = .4$ ,  $\eta^2 = .06$ . Planned contrasts revealed that the accurate ( $p = .014$ , 95% confidence interval [CI] [23.36, 199.50]) and tripled ( $p = .006$ , 95% CI [38.73, 218.61]) feedback groups entered significantly more checks than the control group, and the one-third feedback group entered a similar amount as the control group.

Figure 2 displays the unadjusted mean number of correctly completed checks over time, with the unadjusted number of correctly completed checks for each participant in each group in grey and the group mean in black. Figure 3 displays the adjusted mean number of correctly completed checks over time. These results suggest that giving accurate or exaggerated feedback may be more beneficial than not giving any feedback or underreporting one's performance as feedback. It is important to note here that some degree of caution should be taken when interpreting the results from this experiment. There were participants in the control group who started the experiment entering fewer checks because of taking more breaks. These participants may have brought the control group's mean down throughout the experiment. However, the ANCOVA used in this study should have compensated for some of this initial difference.

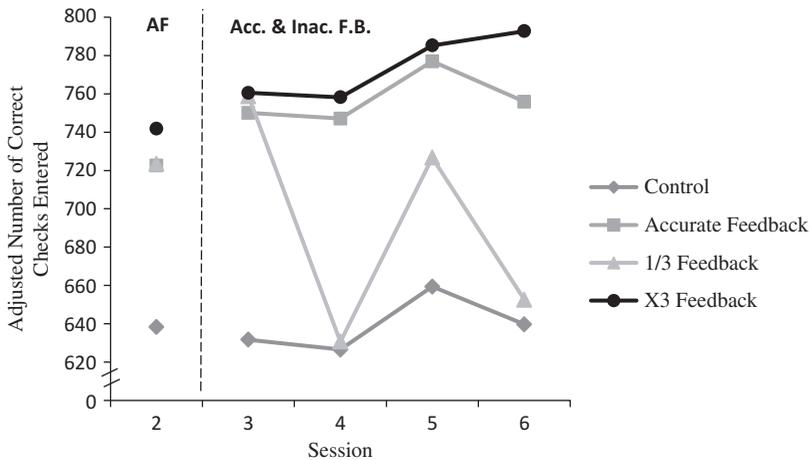
As can be seen in the bottom graphs of Figure 2, how individuals responded to the different types or lack of feedback was quite varied. In some cases, providing accurate feedback merely stabilized responding across sessions rather than increasing the number of correctly completed checks. Providing inaccurate feedback sometimes increased performance, whereas others in the same group decreased performance. In other cases in which no feedback was administered, responding increased to levels above those of any individuals who received any type of feedback. This may be due to whether participants could detect the inaccurate feedback, which may have been more difficult to detect by participants in the tripled feedback condition and slightly easier to detect by participants in the one-third feedback condition because of a contradiction between patterns of performance and patterns of feedback. It is also likely that the participants' history with performance feedback may have played into the variability seen between and within groups.

Visual analysis of the data paths of participants in the one-third feedback group shows that underperformance was due to some skewing the group mean. Coincidentally, about one third of the participants in this group demonstrated this downward trajectory. However, because the feedback was generated by dividing their total number of checks entered correctly by 3, a sizable drop in performance may not have been detectable in the feedback they received (e.g., if their actual performance dropped from 812 to 740, their feedback indicated they dropped from 271 to 247). This may



**FIGURE 2** Average number of correctly completed checks over time. In bottom graphs black data paths are the group means, whereas grey data paths are individual participant data paths. BL = baseline; AF = accurate feedback; Acc. & Inac. F.B. = accurate and inaccurate feedback.

account for some of the deterioration in performance started in later sessions. Participants could have been working to improve their performance, but because the feedback they received was only one third of their actual increases, it may have been insufficient to support additional effort.



**FIGURE 3** Adjusted average number of correctly completed checks over time. Adjustments were made based on first-session performance. AF = accurate feedback; Acc. & Inac. F.B. = accurate and inaccurate feedback.

It is also unclear why the tripled feedback group performed at similar levels to the accurate feedback group. Many in the tripled feedback group indicated that their performance feedback was accurate, so the researchers' verbal and graphic report at the beginning of a session might not have been detectable as inaccurate. In organizational settings inaccurate feedback is not likely to be triple, or one third, of employee performance and thus less likely to be detectable as inaccurate. Even so, the feedback participants received was numerically inaccurate, and performance gains in the tripled feedback group were clearly seen.

Both forms of inaccurate feedback used in this study were partially contingent on individual performance, and although the performance feedback reported prior to the second session was accurate, some participants may not have been able to detect the inaccurate feedback reported prior to the third and subsequent sessions. This may be due to whether the pattern of feedback matched the pattern of performance. Prior to the second session accurate feedback was delivered to all feedback conditions. The third session was the first time that half of the participants (triple and one-third feedback groups) shifted to the inaccurate feedback. In the accurate and tripled feedback conditions, participant performance improved in Session 2, and the feedback received before the next session indicated an improvement. However, in the one-third feedback condition, participant performance improved in the second session, and the feedback received before the next session indicated a worsening of performance. Therefore, the one-third feedback condition may be the only group with a salient inaccuracy (on average performance went up by about 30, but feedback went down by about 540).

Although the tripled feedback condition had a sudden shift in numbers, there was less of a shift in the pattern, which likely was not detectable. Participants may be insensitive to feedback inaccuracy if there is not a contradiction in trends in feedback. The deterioration in performance by some in the one third inaccurate group may have been a trend that would have become more pronounced over time. In addition, the reduced performance in the one-third feedback group was fairly immediate; however, a reduction in performance may have become evident after further repeated exposure to tripled feedback than that provided here.

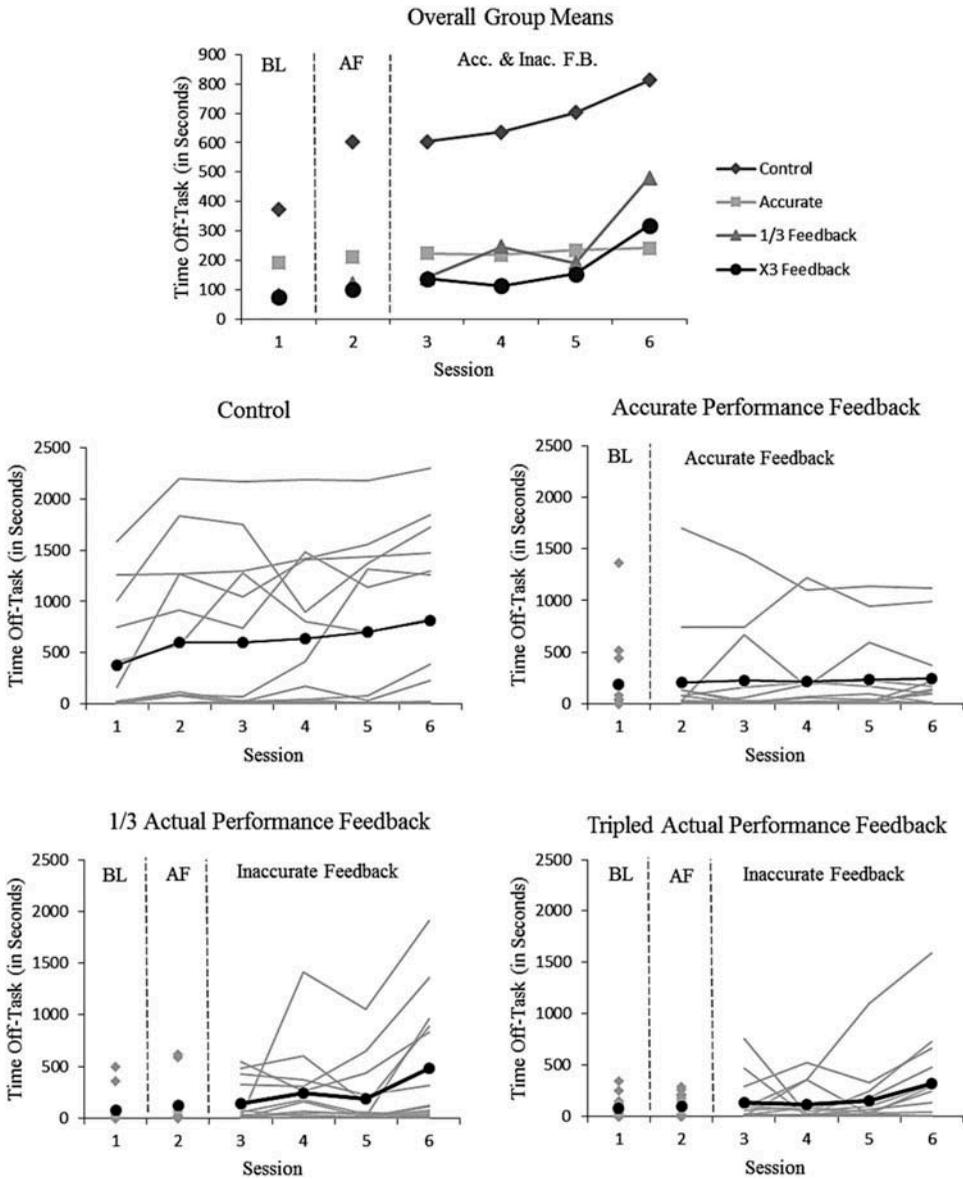
#### TIME OFF TASK, RUN RATES, AND INCORRECTLY ENTERED CHECKS

What is clearer from this experiment is that in the absence of feedback, performance did not improve. This replicates the results from D. A. Johnson (2013). As can be seen in Figure 4, this appears to be at least partially attributable to the control group taking more breaks. As noted earlier, it should be noted that despite random assignment there were more participants in the control group taking breaks during the first session compared to the other groups. Thus, results may be partially due to incoming differences rather than the task or manipulation of the variables used in this study.

Time off task during the five experimental sessions was analyzed using a repeated measures ANCOVA; data from the first session served as the covariate. There were no differences in time off task by the type of feedback after the first session was controlled,  $F(3, 51) = 2.03, p = .12, \eta^2 = .11$ . There was a main effect of the sessions,  $F(2.44, 124.32) = 7.28, p < .001, \eta^2 = .13$ . There were no significant interactions between the experimental sessions and the different types of feedback,  $F(7.31, 124.32) = 1.09, p = .38, \eta^2 = .06$ . Planned contrasts revealed that the accurate feedback group ( $p = .02, 95\% \text{ CI } [-477.06, -38.71]$ ) spent significantly less time off task than the control group, and the tripled feedback group approached significance ( $p = .08, 95\% \text{ CI } [-425.54, 25.84]$ ).

Run rates were calculated for each session by subtracting the total recorded no-activity time from 45 min and dividing the number of correctly completed checks by this number. These data can be seen in Figure 5. Visual analysis of these graphed data suggests that there were no differences of run rates between all four groups.

The mean number of incorrectly completed checks for each participant and each group can be seen in Figure 6. Visual analysis of this graph suggests no differences between the groups. The control group attained 99.98% accuracy on the checks entered, whereas the remaining groups—accurate, one-third, and tripled feedback—all attained 99.99% accuracy across all sessions. After these data were normalized to account for the differences in the total number of checks entered, no differences between the groups could be seen.

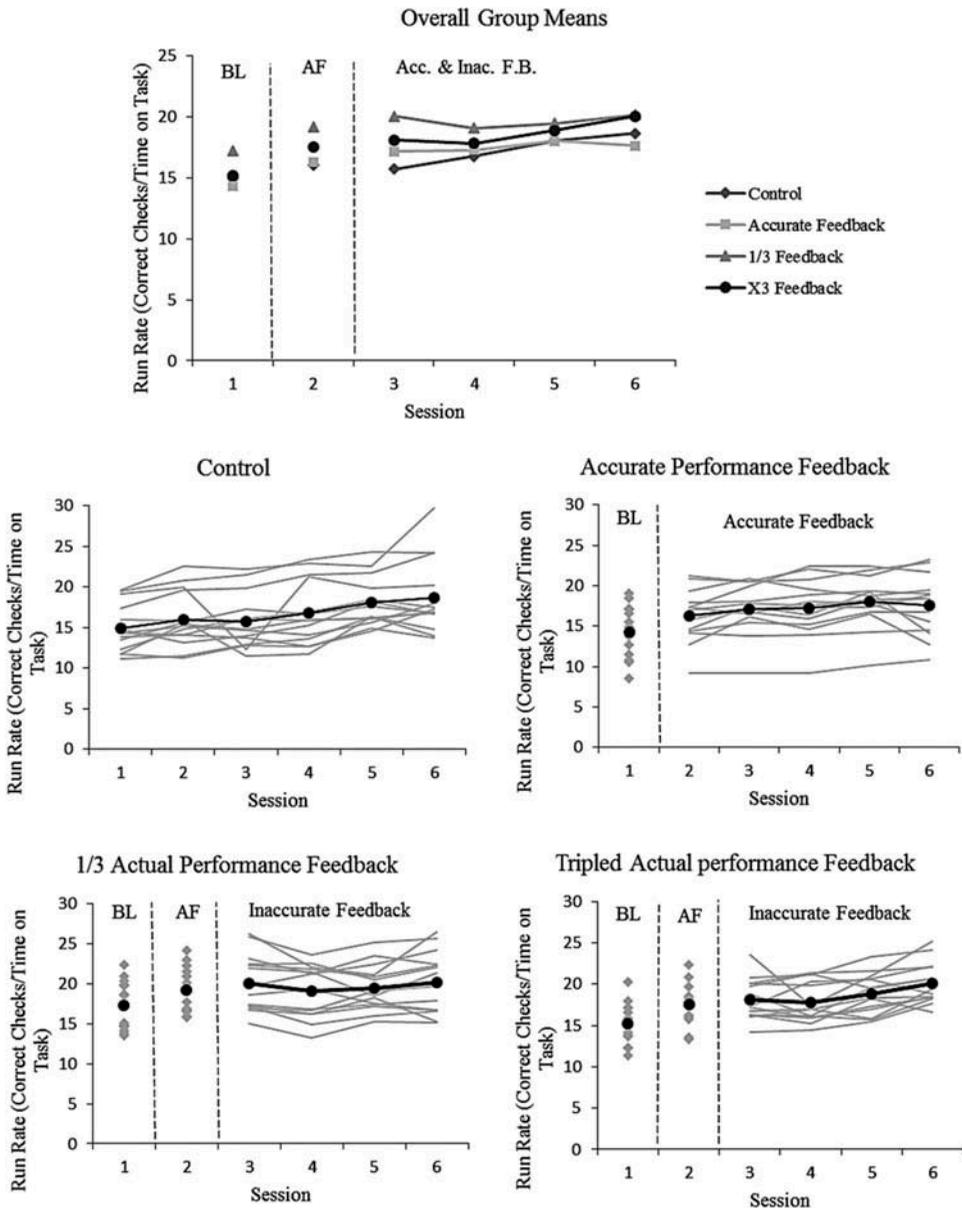


**FIGURE 4** Average time off task over time. In bottom graphs black data paths are the group means, whereas grey data paths are individual participant data paths. BL = baseline; AF = accurate feedback; Acc. & Inac. F.B. = accurate and inaccurate feedback.

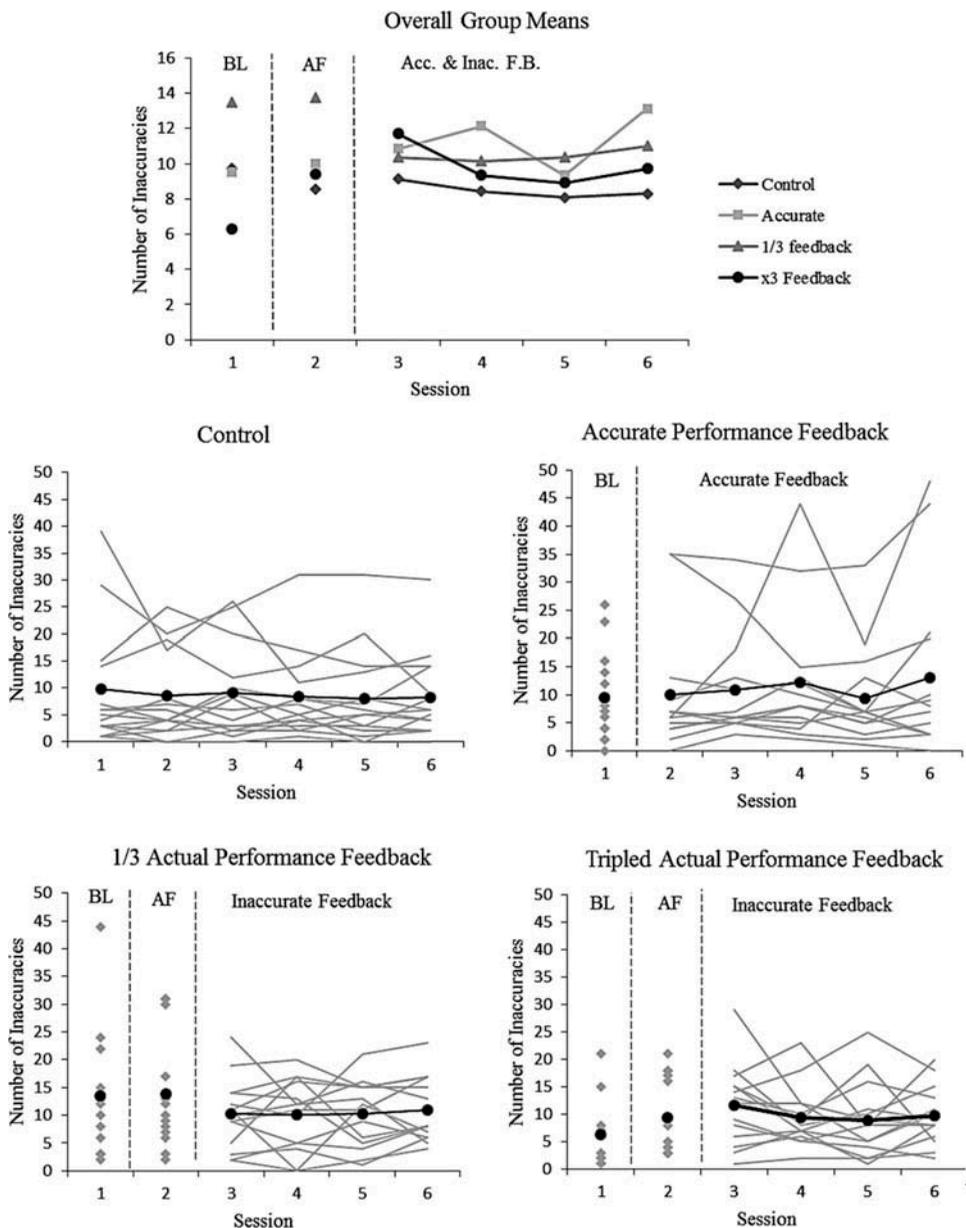
WITHIN-SESSION PATTERNS OF BEHAVIOR

Each session was then broken into 15 3-min bins, and the number of correct checks entered during each consecutive 3-min bin was calculated for each participant. No statistical differences between sessions existed, so the

Downloaded by [Western Michigan University] at 21:26 30 November 2015



**FIGURE 5** Average run rate over time. In bottom graphs black data paths are the group means, whereas grey data paths are individual participant data paths. Run rates were calculated by dividing the number of correct checks entered per session by the total time on task for that session. BL = baseline; AF = accurate feedback; Acc. & Inac. F.B. = accurate and inaccurate feedback.

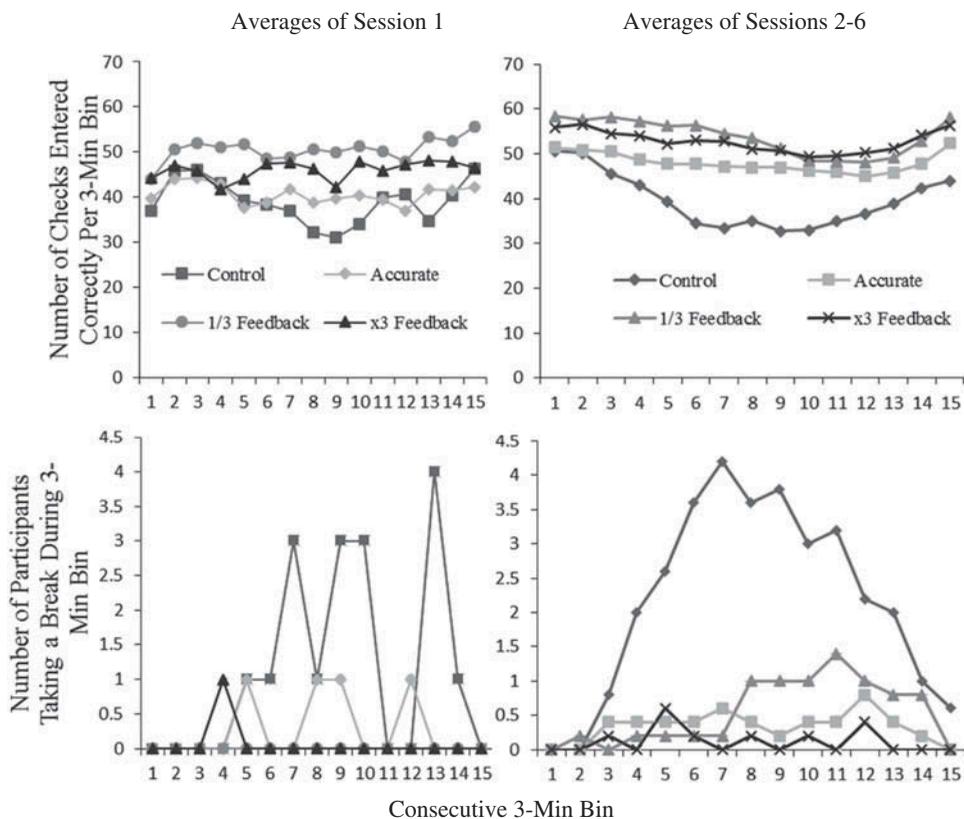


**FIGURE 6** Average number of inaccuracies over time. In bottom graphs black data paths are the group means, whereas grey data paths are individual participant data paths. BL = baseline; AF = accurate feedback; Acc. & Inac. F.B. = accurate and inaccurate feedback.

group averages across Sessions 2 through 6 were collapsed; because the first session served as baseline, it is displayed separately. The top graphs in Figure 7 display these group averages for each bin for Sessions 2 through 6 and Session 1. These data depict that participants in the control group took

more breaks, as indicated by a downward trend from the beginning until the middle, with a slight increase in performance at the end of sessions. These data also demonstrate that participants were either entering checks around the mean rate or taking breaks. Few entered checks below 10 per minute. However, the bottom graphs in Figure 7 display this effect more clearly: The control group had more constituents taking breaks compared to the other three groups.

The analysis of these data (within-session rates of responding), along with Figures 4 and 5, revealed that participants typically took a few longer breaks rather than many small breaks. This indicated that participants engaged in bout-like behavior (Shull, 2011). Shull, Gaynor, and Grimes (2001) noticed pause-and-run (bout) patterns in cumulative records produced by certain schedules of reinforcement and hypothesized that the overall response rates calculated for these data were likely a composite measure of performance. That is, response rate was a combination of the post-reinforcement pauses and actual lever pressing. Participants in this



**FIGURE 7** Within-session patterns of behavior. Graphs on the left are the average of five sessions and are representative of each session.

experiment were not presented with any overt reinforcers on different schedules of reinforcement within sessions and therefore may not have had true post-reinforcement pauses. In addition, interresponse time analyses were not completed.

However, when within-session data were compiled, a similar pattern of pause-and-run behavior became apparent. In this experiment pausing was conceptualized as taking a break, and the bouts were conceptualized as the time when participants were engaged in the task. Here the response rate can be viewed as a composite measure consisting of the time between entering checks (breaks) and time on task entering checks. This pattern of behavior becomes evident in [Figure 7](#) as participants were either not entering checks or entering checks at or around the mean.

This type of performance is why run rates are reported. To include the amount of time the participants were not engaging in bouts of responding in the calculation of response rates would suggest that the control group simply responded at a slower rate than the other groups. This would lead many readers to conclude that performance feedback increases the overall rate of responding. However, when these variables are analyzed in more detail a clearer picture can be portrayed of the participants' behavior. It is suggested that future researchers measure responding and time off task, within and between sessions, as performance may be a composite measure. Further analyses of these composite measures, although more time consuming, may be worth the gain in understanding of the variables used (Shull, 2011).

This type of analysis also reveals a strength of Experiment 2 over Experiment 1. In Experiment 1, with the constant presence of the researcher, participants did not engage in alternative activities. Because of this, any measure of response rate consisted only of responding. In Experiment 2, the researcher was not in the room and participants were more likely to engage in alternative activities. The researchers not being in the room allowed a more realistic simulation of job conditions and led to the run rate analysis. In the absence of this change in experimental method, the finding that response rate with this task is a composite of on-task and off-task behaviors would have been missed.

It should be noted that few participants were off task during the first and last 3 min of each session. This is likely because of the reactivity effects described earlier but should be explored in future research by systematically manipulating experimenter presence and absence. This is especially important because of differences between the control and experimental groups during the first session, a possible confound. Despite random assignment, the control group started lower and remained so throughout the six sessions. This effect size was as strong (.18) as group differences (.17) from the independent variable.

**TABLE 2** Mean (*SD*) Scores on Surveys in Experiment 2 by Group and Time of Administration

Question	Survey	Condition			
		Control	Accurate	One third	Triple
Payment fair	1	1.6 (0.5)	1.6 (0.8)	1.5 (0.9)	1.3 (0.6)
	2	1.6 (0.7)	1.8 (0.7)	1.6 (0.8)	2.0 (1.2)
Set goals	1	2.3 (1.2)	2.7 (1.3)	2.4 (0.9)	2.2 (0.7)
	2	2.1 (0.8)	2.4 (1.1)	2.6 (1.2)	2.7 (1.0)
Feedback accurate	1		1.7 (0.8)	1.9 (0.9)	1.6 (0.9)
	2		1.7 (1.2)	3.1 (1.3)	2.4 (1.5)
Task boring	1	1.9 (1.0)	1.9 (0.9)	1.9 (1.1)	1.6 (0.8)
	2	1.7 (0.8)	2.0 (1.1)	1.7 (0.9)	1.5 (0.9)
Feedback helpful	1		1.6 (0.9)	2.4 (1.0)	2.3 (1.0)
	2		2.0 (0.7)	2.8 (1.0)	2.5 (1.4)
Money motivating	1	1.4 (0.5)	2.2 (1.1)	2.1 (1.0)	1.6 (0.7)
	2	1.7 (0.8)	1.7 (1.2)	1.9 (1.0)	1.7 (0.7)
Tell friends	1	1.7 (0.8)	1.8 (0.7)	2.1 (0.7)	2.0 (0.9)
	2	1.6 (0.9)	1.7 (0.6)	2.1 (1.0)	2.5 (1.1)
Enjoyed study	1	1.9 (0.9)	1.9 (0.8)	2.4 (0.6)	2.5 (1.1)
	2	1.9 (1.0)	1.9 (0.8)	2.3 (0.8)	2.8 (1.2)
Aggregate	1	13.7	15.4	16.6	15.0
	2	14.1	15.3	18.1	18.1

*Note.*  $N = 14$  per group. Data are unavailable for Questions 3 and 5 (“feedback accurate” and “feedback helpful”) for the control group, as this group did not receive feedback. Content in the “Question” column is abbreviated items from the survey (see the note to Table 1 for full survey questions).

## SURVEY RESULTS AND OTHER SUBJECTIVE MEASURES

Survey data also were analyzed. There were no apparent differences in the response to Question 3 (“The performance feedback I received was accurate”) between the three feedback groups for the first survey; this is good, because all feedback thus far was accurate. However, on the second survey, the one-third feedback group responded more skeptically ( $M = 3.1$ ,  $SD = 1.3$ ), along with the tripled feedback group ( $M = 2.4$ ,  $SD = 1.7$ ), compared to the accurate feedback condition ( $M = 1.7$ ,  $SD = 1.2$ ). The one-third group may have responded more skeptically because of their feedback being more detectable. Participants in this condition were the only ones who experienced a contradiction between their performance and their feedback between Sessions 2 and 3. In addition, these participants had a lower correlation between performance and feedback in the remaining sessions. Some participants also self-reported setting goals for themselves on both surveys. Table 2 displays survey results.

After running the first 10 experimental participants and not witnessing participants reacting to the inaccurate feedback, and after the funding source recommended adding participant reactions to the feedback as a secondary measure, the researchers began measuring the participants’ reactions to the feedback. Researchers then rated 32 experimental participants regarding their response to the feedback immediately after starting Sessions 2 through 6.

What was unusual was that across sessions in both inaccurate feedback groups, participants displayed very little to no emotional responses to the feedback they received. When presenting the feedback the researcher and research assistants attempted not to deliver signs of approval or disapproval; however, as indicated by D. A. Johnson (2013), participants could have believed they were being evaluated regardless of this attempt.

As can be seen in Figures 3, 4, and 6, a high amount of variability can be seen in the data, both between and within participants. There are a few reasons why this variability may be present. There were differences both within and between participants in the amount of time between the end of the previous session and the presentation of the feedback at the next session. Research on the delay of feedback has been unclear as to whether a short or long delay is more beneficial (Bechtel, McGee, Huitema, & Dickinson, in press; Kang, Oah, & Dickinson, 2005; Krumhus & Malott, 1980; Mason & Redmon, 1993; Pampino, MacDonald, Mullin, & Wilder, 2004; So, Lee, & Oah, 2013). Therefore, the differences in feedback delay for individual participants in this study could have introduced some uncontrolled variation in the data. The average time between feedback presentations in this study was 1.8 days ( $SD = 1.5$ ), and this was consistent across groups. Nevertheless, the delay of feedback effects may have played a role.

None of the participants spoke to one another during the sessions, at least at audible levels from the researchers' position in the hallway. No objective data were gathered systematically to determine what they were doing when not entering checks. Participants could have been either using their cell phones, which the researcher and research assistants did not monitor; playing a game that was on the computer; browsing the Internet; or even sleeping. Because researchers were not in the room during the session, which is similar to many jobs that this task simulated, it is unknown what participants were doing when off task. This is another limitation to this study, as these data would allow for a better understanding of competing contingencies in the workplace.

#### COMPARISONS WITH PAST RESEARCH

D. A. Johnson et al. (2008) used the same task, coupled with performance feedback, and found that this objective feedback had no effect on the number of correct checks entered. A few variables differed between their experiment and the present one. First, D. A. Johnson et al. (2008) did not allow participants access to the Internet. This powerful competing contingency had the potential of causing more participants in the second study to take breaks. Indeed, businesses often monitor employee usage of the Internet and have suspended or fired employees for inappropriate use (Weiss, 2000). However, control participants in D. A. Johnson et al.'s (2008) study spent more time off task (adjusted  $M = 13.0$  min) than control

participants in this study (adjusted  $M = 7.9$  min). Clearly these differences in outcomes for relatively similar research protocols need additional investigation.

Second, their study provided objective performance feedback on screen, without being delivered by a person. It is possible that differences in performance may result when performance feedback is delivered by a researcher rather than a computer, as well as when it is delivered textually on a continuous basis rather than as a pre-session vocal and visual stimulus. This may support the hypothesis that feedback delivered by managers (as a verbal stimulus) sets up a rule-governed contingency similar to those stated previously, whereas feedback delivered by a computer may lead to contingency-shaped or rule-governed behavior that is governed by different contingencies (Malott, 1992). In the current study the overall mean number of correctly entered checks for the control group was 639.1, and in D. A. Johnson et al. (2008) the overall mean for the control group was 571.4. This can probably be attributed to time-off-task differences, as the run rates of entering checks were similar (17.2 and 16.7). However, additional differences exist between the current study and D. A. Johnson et al.'s (2008) study. In the current study the adjusted overall mean number of correctly entered checks in the accurate feedback group was 750.5; it was 567.4 in D. A. Johnson et al.'s (2008) study.

As mentioned earlier, these differences may have arisen because of the researcher delivering accurate feedback at the beginning of a session compared to the computer delivering feedback during the session. D. A. Johnson (2013) arrived at similar results to the present experiment in his objective feedback group. The overall mean score for his objective feedback group was 765.7, whereas the overall mean for the current study was 750.5. Moreover, the control group means were similar: 654.0 and 639.1. These numbers are quite remarkable, even with both studies using the same experimental task, considering that the studies were completed at different universities and by different researchers with slightly different methods. More differences were found within labs at the same university (D. A. Johnson, 2013; D. A. Johnson et al., 2008) than between this study and the D. A. Johnson study. This further suggests that computer-delivered objective feedback, as delivered by D. A. Johnson et al. (2008), is a different type of feedback than human-delivered objective feedback, as delivered by D. A. Johnson. Clearly the past history of a participant interacting with supervisors and computer-delivered systems would play an important role.

## GENERAL DISCUSSION

These two studies were designed to determine whether accurate objective feedback is necessary for performance to improve. Objective feedback may

not need to be numerically accurate to improve performance, at least to the levels studied. However, whether these levels are discernable to participants is still unknown. Experiment 2 data also provide some evidence for the recommendation of using accurate feedback. Compared to underreported feedback, accurate feedback does seem to help improve performance.

Overall, it seems that accurate and exaggerated objective performance feedback is better than no feedback and underreported feedback. However, small deviations from actual previous performance may not be detrimental to future performance. The question about exaggerated feedback is particularly worthwhile to explore. The tripled inaccurate feedback in Experiment 2 did not deteriorate performance, and participants did not question this on the survey or during face-to-face interactions before Sessions 2 through 6. Again, this may be because of there being no contradiction in direction between performance and feedback. Could this eventually cause problems or facilitate performance over longer periods of time? That is, are there caveats to honesty is the best policy?

### General Limitations and Future Research

It is possible that all groups receiving some type of feedback exhibited reactivity. Some control participants were off task more, which again may have been at least partially due to incoming differences in a few of the control participants. It has been proposed for many years that if participants know they are being watched, their behavior will change (C. G. Baum, Forehand, & Zegiob, 1979). It is hypothesized that the presence of the experimenter, coupled with the researchers' presented feedback at the beginning of sessions, may have caused participants to exhibit what they thought were socially desirable responses regardless of the type of feedback received (C. G. Baum et al., 1979). It is possible that the participant made a covert response such as "Oh, while the experimenter isn't in the room, he is still watching so I better do the task." However, although the reactivity hypothesis may account for some of the variance, this effect may not account for a lot of the variability. Indeed, research on reactivity has provided mixed conclusions about the size of its effect (C. G. Baum et al., 1979; Brackett, Reid, & Green, 2007; Hagen, Craighead, & Paul, 1975; Nelson, Kapust, & Dorsey, 1978).

A goal of these studies was to separate the presentation of objective performance feedback from other intervention components often used in OBM research, such as goal setting. Although this was attempted, participants from both experiments indicated on the surveys that they set goals for themselves (see the means for Question 2 in Tables 1 and 2 ). Surveys administered during Experiment 2 also had the option of explaining how these students set goals. Generally, if participants answered this question, there were two different responses. Participants either set the goal of entering more than the previous session or set the goal of staying on task. Ample research suggests

that goal setting can impact performance (Alvero et al., 2001; Balcazar et al., 1985–1986; Fellner & Sulzer-Azaroff, 1984; VanStelle et al., 2012). If some, but not all, participants in each group were self-setting goals, this may have added to the variability seen in the data. However, this may not be an experimental flaw but rather a natural part of any feedback intervention in both lab and field implementations.

Future research also should investigate whether presenting accurate feedback, followed by inaccurate feedback, impacts performance differently than immediately presenting inaccurate feedback. It may be that the presentation of accurate feedback prior to Session 2 was then used as a reference point during the remainder of the study. The tripled feedback group received a large increase in its reported performance compared to the accurate feedback reported in Session 2, whereas the one-third feedback group received a large decrease in their reported performance. Different covert responses likely ensued. What might happen to performance if this reference point is removed? Furthermore, participants in these studies were not told that feedback might be inaccurate. Would differences in performance ensue if participants are aware that feedback might be inaccurate?

Future research also should attempt to reduce possible sources of variability in the data. This may be accomplished by ensuring that delays to feedback are constant across participants. Researchers also should try to keep the number of participants running at the same time constant across all participants. Although there were no obvious differences between running participants alone versus in a group in these studies, this could add variability to the data obtained. However, although controlling for these variables may increase experimental control, it is likely to reduce external validity, as many workplaces have delayed and aperiodic delivery of feedback as well as employees working alone and in groups. Nevertheless, future research should investigate the impact these variables may have on performance.

In addition, more single-subject research designs should be carried out with many more sessions, thus approximating more typical working conditions. It may also be worthwhile to develop a new implementation of inaccurate feedback. The procedure used here involved inaccurate but positively correlated feedback. Perhaps having a consistent negative correlation between performance and feedback between sessions (performance goes up, feedback goes down, and vice versa) or zero correlation (feedback numbers are randomly generated) would increase the saliency of the inaccurate feedback. As Skinner (1957) noted, verbal stimuli, such as performance feedback, have complex qualities that might exert multiple controls on behavior. In this study, the feedback may have been partially inaccurate, so by definition false, but partially accurate as well. Clearly, performance feedback is not a simple stimulus; it is multidimensional in nature and extremely complex.

## General Conclusions

In summary, OBM has much more to learn about its most commonly used intervention component. More in-depth analyses are recommended to determine how performance feedback affects different dimensions of behavior. For instance, it was found here that performance feedback did not change the accuracy or rate at which participants entered data. However, it seems that performance feedback may influence how much time participants spend off task, and subsequently the number of checks entered. More research will need to be completed to determine why participants performed as they did when exposed to the inaccurate feedback. It is also suggested that these types of analyses be done when doing component analyses of performance feedback.

This set of experiments explored only one component, and results from these suggest that it may be beneficial to report accurate performance feedback, replicating results from years of experimenting and justifying recommendations made for decades. However, it may be just as beneficial to exaggerate performance feedback. Because more research will need to be done to determine the side effects, both short and long term, of underreported performance feedback, supervisors should be very careful if they are tempted to underreport performance. Managers who might be tempted to underreport performance to get employees to work harder may be causing damage rather than increasing organizational output (in addition to other legitimate concerns related to credibility and ethics), and these potential effects warrant future research. These cautionary recommendations may be particularly important to supervisors who do not carefully monitor employees under their direction. When guessing how well a worker is performing, first do no harm.

## ACKNOWLEDGMENTS

We wish to thank Mark P. Reilly and Kyunghye Han for their thoughts and suggestions regarding this report.

## FUNDING

We wish to thank the Chris Anderson Research Grant through the OBM Network and Carl Johnson's Central Michigan University President's Award for Research for funding portions of this research.

## REFERENCES

- Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998). *Journal of Organizational Behavior Management*, 21(1), 3-29. doi:10.1300/J075v21n01\_02
- Antonelli, D. C., & Karas, G. G. (1967). Performance on a vigilance task under conditions of true and false knowledge of results. *Perceptual and Motor Skills*, 25(1), 129-138. doi:10.2466/pms.1967.25.1.129
- Balcazar, F. E., Hopkins, B. L., & Suarez, Y. (1985-1986). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, 7(3-4), 65-89. doi:10.1300/J075v07n03\_05
- Baum, C. G., Forehand, R., & Zegiob, L. E. (1979). A review of observer reactivity in adult-child interactions. *Journal of Behavioral Assessment*, 1(2), 167-178. doi:10.1007/BF01322022
- Baum, W. M. (2002). From molecular to molar: A paradigm shift in behavior analysis. *Journal of the Experimental Analysis of Behavior*, 78(1), 95-116. doi:10.1901/jeab.2002.78-95
- Bechtel, N. T., McGee, H. M., Huitema, B. E., & Dickinson, A. M. (2015). The effects of the temporal placement of feedback on performance. *The Psychological Record*, 65(3), 425-434. doi:10.1007/s40732-015-0117-4
- Beedie, C. J., Lane, A. M., & Wilson, M. G. (2012). A possible role for emotion and emotion regulation in physiological responses to false performance feedback in 10 mile laboratory cycling. *Applied Psychophysiology and Biofeedback*, 37, 269-277. doi:10.1007/s10484-012-9200-7
- Brackett, L., Reid, D. H., & Green, C. W. (2007). Effects of reactivity to observations on staff performance. *Journal of Applied Behavior Analysis*, 40(1), 191-195. doi:10.1901/jaba.2007.112-05
- Daniels, A. C. (2000). *Bringing out the best in people* (2nd ed.). New York, NY: McGraw-Hill.
- Daniels, A. C., & Daniels, J. E. (2004). *Performance management: Changing behavior that drives organizational effectiveness* (4th ed.). Tucker, GA: Performance Management.
- Faulkner, J., Arnold, T., & Eston, R. (2011). Effect of accurate and inaccurate distance feedback on performance markers and pacing strategies during running. *Scandinavian Journal of Medicine & Science in Sports*, 21, e176-e183. doi:10.1111/j.1600-0838.2010.01233.x
- Fellner, D. J., & Sulzer-Azaroff, B. (1984). A behavioral analysis of goal setting. *Journal of Organizational Behavior Management*, 6(1), 33-51. doi:10.1300/J075v06n01\_03
- Filipkowski, M., & Johnson, C. M. (2008). Comparisons of performance and job insecurity in union and nonunion sites of a manufacturing company. *Journal of Organizational Behavior Management*, 28(4), 218-237. doi:10.1080/01608060802454437
- Frederiksen, L. W. (Ed.). (1982). *Handbook of organizational behavior management*. New York, NY: Wiley.

- Hagen, R. L., Craighead, W. E., & Paul, G. L. (1975). Staff reactivity to evaluative behavioral observations. *Behavior Therapy*, *6*, 201–205. doi:10.1016/S0005-7894(75)80141-9
- Hirst, J. M., & DiGennaro-Reed, F. D. (2015). An examination of the effects of feedback accuracy on academic task acquisition in analogue settings. *The Psychological Record*, *65*, 49–65. doi:10.1007/s40732-014-0087-y
- Hirst, J. M., DiGennaro-Reed, F. D., & Reed, D. D. (2013). Effects of varying feedback accuracy on task acquisition: A computerized translational study. *Journal of Behavioral Education*, *22*, 1–15. doi:10.1007/s10864-012-9162-0
- Houmanfar, R. (2013). Performance feedback: From component analysis to application. *Journal of Organizational Behavior Management*, *33*(2), 85–88. doi:10.1080/01608061.2013.787002
- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*(4), 349–371. doi:10.1037/0021-9010.64.4.349
- Johnson, C. M., Redmon, W. K., & Mawhinney, T. C. (Eds.). (2001). *Handbook of organizational performance: Behavior analysis and management*. Binghamton, NY: Haworth Press.
- Johnson, D. A. (2013). A component analysis of the impact of evaluative and objective feedback on performance. *Journal of Organizational Behavior Management*, *33*(2), 89–103. doi:10.1080/01608061.2013.785879
- Johnson, D. A., Dickinson, A. M., & Huitema, B. E. (2008). The effects of objective feedback on performance when individuals receive fixed and individual incentive pay. *Performance Improvement Quarterly*, *20*(3–4), 53–74. doi:10.1002/piq.20003
- Johnson, D. A., Rocheleau, J. M., & Tilka, R. E. (in press). Considerations in feedback delivery: The role of accuracy and type of evaluation. *Journal of Organizational Behavior Management* *35*(3–4).
- Kang, K., Oah, S., & Dickinson, A. M. (2005). The relative effects of different frequencies of feedback on work performance. *Journal of Organizational Behavior Management*, *23*(4), 21–53. doi:10.1300/J075v23n04\_02
- Komaki, J., & Goltz, S. M. (2001). Within-group research designs: Going beyond program evaluation questions. In C. M. Johnson, W. K. Redmon, & T. C. Mawhinney (Eds.), *Handbook of organizational performance: Behavior analysis and management* (pp. 81–123). Binghamton, NY: Haworth Press.
- Kondo, C., & Canter, A. (1977). Time and false electromyographic feedback: Effect on tension headache. *Journal of Abnormal Psychology*, *86*(1), 93–95. doi:10.1037/0021-843X.86.1.93
- Kreitner, R., Reif, W. E., & Morris, M. (1977). Measuring the impact of feedback on the performance of mental health technicians. *Journal of Organizational Behavior Management*, *1*(1), 105–109. doi:10.1300/J075v01n01\_08
- Krumhus, K. M., & Malott, R. W. (1980). The effects of modeling and immediate and delayed feedback in staff training. *Journal of Organizational Behavior Management*, *2*(4), 279–293. doi:10.1300/J075v02n04\_05
- Lebbon, A. R., & Austin, J. (2013). A preliminary examination of the effects of observer presence on work-related behavior in a simulated office.

- Journal of Organizational Behavior Management*, 33, 185–199. doi:10.1080/01608061.2013.815095
- Luthans, F., & Kreitner, R. (1985). *Organizational behavior modification and beyond*. Glenview, IL: Scott, Foresman.
- Mackworth, J. F. (1964). The effect of true and false knowledge of results on the detectability of signals in a vigilance task. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, 18(2), 106–117. doi:10.1037/h0083493
- Malott, R. W. (1992). A theory of rule-governed behavior and organizational behavior management. *Journal of Organizational Behavior Management*, 12(2), 45–65. doi:10.1300/J075v12n02\_03
- Malott, R. W. & Shane, J. T. (2013). *Principles of behavior*. Seventh Edition. New York, NY: Routledge.
- Mangiapanello, K. A., & Hemmes, N. S. (2015). An analysis of feedback from a behavior analytic perspective. *The Behavior Analyst*, 38(1), 51–75. doi:10.1007/s40614-014-0026-x
- Mason, M. A., & Redmon, W. K. (1993). Effects of immediate versus delayed feedback on error detection accuracy in a quality control simulation. *Journal of Organizational Behavior Management*, 13(1), 49–83. doi:10.1300/J075v13n01\_05
- Mayer, G. R., Sulzer-Azaroff, B., & Wallace, M. (2013). *Behavior analysis for lasting change*. Cornwall-on-Hudson, NY: Sloan.
- McGee, H. M., Dickinson, A. M., Huitema, B. E., & Culig, K. M. (2006). The effects of individual and group monetary incentives on high performance. *Performance Improvement Quarterly*, 19(4), 107–130. doi:10.1111/j.1937-8327.2006.tb00387.x
- Michael, J. (2004). *Concepts and principles of behavior analysis* (2nd ed.). Kalamazoo, MI: Association for Behavior Analysis International.
- Nelson, R. O., Kapust, J. A., & Dorsey, B. L. (1978). Minimal reactivity of overt classroom observations on student and teacher behaviors. *Behavior Therapy*, 9, 695–702. doi:10.1016/S0005-7894(78)80001-X
- Nestoriuc, Y., Martin, A., Rief, W., & Andrasik, F. (2008). Biofeedback treatment for headache disorders: A comprehensive efficacy review. *Applied Psychophysiology and Biofeedback*, 33(3), 125–140. doi:10.1007/s10484-008-9060-3
- O'Brien, R. M., Dickinson, A. M., & Rosow, M. P. (Eds.). (1982). *Industrial behavior modification: A management handbook*. New York, NY: Pergamon Press.
- Pampino, R. N., MacDonald, J. E., Mullin, J. E., & Wilder, D. A. (2004). Weekly feedback vs. daily feedback. *Journal of Organizational Behavior Management*, 23(2–3), 21–43. doi:10.1300/J075v23n02\_03
- Peretti, P. O. (1970). Effects on reaction time of random-false knowledge of results of performance. *Perceptual and Motor Skills*, 30(3), 838. doi:10.2466/pms.1970.30.3.838
- Peterson, N. (1982). Feedback is not a new principle of behavior. *The Behavior Analyst*, 5, 101–102.
- Prue, D. M., & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management*, 3(1), 1–16. doi:10.1300/J075v03n01\_01
- Shull, R. L. (2011). Bouts, changeovers, and units of operant behavior. *European Journal of Behavior Analysis*, 12(1), 49–72.

- Shull, R. L., Gaynor, S. T., & Grimes, J. A. (2001). Response rate viewed as engagement bouts: Effects of relative reinforcement and schedule type. *Journal of the Experimental Analysis of Behavior*, 75(3), 247–274. doi:10.1901/jeab.2001.75-247
- Skinner, B. F. (1957). *Verbal behavior*. New York, NY: Appleton-Century-Crofts.
- Slowiak, J. M., Dickinson, A. M., & Huitema, B. E. (2011). Self-solicited feedback: Effects of hourly pay and individual monetary incentive pay. *Journal of Organizational Behavior Management*, 31(1), 3–20. doi:10.1080/01608061.2011.541816
- So, Y., Lee, K., & Oah, S. (2013). Relative effects of daily feedback and weekly feedback on customer service behavior at a gas station. *Journal of Organizational Behavior Management*, 33(2), 137–151. doi:10.1080/01608061.2013.785898
- VanStelle, S. E., Vicars, S. M., Harr, V., Miguel, C. F., Koerber, J. L., Kazbour, R., & Austin, J. (2012). The publication history of the *Journal of Organizational Behavior Management*: An objective review and analysis: 1998-2009. *Journal of Organizational Behavior Management*, 32(2), 93–123. doi:10.1080/01608061.2012.675864
- Venables, L., & Fairclough, S. H. (2009). The influence of performance feedback on goal-setting and mental effort regulation. *Motivation and Emotion*, 33, 63–74. doi:10.1007/s11031-008-9116-y
- Weatherly, N. L., & Malott, R. W. (2008). An analysis of organizational behavior management research in terms of the three-contingency model of performance management. *Journal of Organizational Behavior Management*, 28(4), 260–285. doi:10.1080/01608060802454643
- Weidenfeller, E. W., Baker, R. A., & Ware, J. R. (1962). Effects of knowledge of results (true and false) on vigilance performance. *Perceptual and Motor Skills*, 14, 211–215. doi:10.2466/pms.1962.14.2.211
- Weiss, T. R. (2000). *Dow chemical fires 50 workers for sending pornographic e-mail*. Retrieved from the Computerworld website: <http://www.computerworld.com/article/2596891/enterprise-applications/dow-chemical-fires-50-workers-for-sending-pornographic-e-mail.html>