*AN ANALYSIS OF A PERFORMANCE FEEDBACK
SYSTEM: THE EFFECTS OF TIMING AND FEEDBACK,
PUBLIC POSTING, AND PRAISE UPON ACADEMIC
PERFORMANCE AND PEER INTERACTION¹*

RONALD VAN HOUTEN, SHARON HILL, AND MADELINE PARSONS
MOUNT SAINT VINCENT UNIVERSITY, ALEXANDER SCHOOL, HALIFAX, N.S.,
AND ATLANTIC INSTITUTE OF EDUCATION, HALIFAX, N.S.

In the first experiment, after establishing baseline composition rates in each classroom, timing (announcing time limits) and feedback (student self-scoring) were introduced followed by the introduction, removal, and reintroduction of public posting of highest scores. Timing and feedback improved story writing performance and public posting of highest scores improved performance even further in both classrooms. Teacher praise produced further improvement in one classroom but had no effect on performance in the other. Changes in on-task behavior paralleled changes in writing rate. Comments made by children concerning their own work or work of their peers were recorded throughout the experiment. Although the baseline rate of performance comments was almost zero, the introduction of each variable markedly increased the rate of performance comments. In the second experiment, baseline rates on reading and language exercises were established in a fifth-grade classroom. The entire performance feedback system was introduced on a multiple baseline across the two behaviors and then removed during the final phase of the experiment. Introducing the system improved performance on both tasks. These results further increased the generality of some of the findings of the previous experiment and of previous research on the efficacy of the experimental package of timing, feedback, public posting, and praise.

DESCRIPTORS: feedback, public posting, time limits, writing rates, academic performance, classroom research, self monitoring

The development of techniques that accelerate progress toward academic objectives is an important concern of educators working with culturally deprived children. One technique that might prove effective is a performance feedback system consisting of timing, feedback, and public posting of student scores.

Van Houten, Morrison, Jarvis, and McDonald (1974) improved children's composition

writing performance by introducing an experimental system consisting of timing, feedback (self-scoring), and public posting of student performance. One question raised by their results is how much each factor contributed to the overall effectiveness of the experimental system. Other research suggests that each of those elements can be effective when individually applied to non-academic behaviors. For example, feedback and self-scoring have proven effective in training teachers to use praise (Cooper, Thompson, and Baer, 1970), to train parents to use social consequences effectively (Herbert and Baer, 1972), to improve classroom behavior (Brodén, Hall, and Mitts, 1971), and in the treatment of phobias (Leitenberg, Agras, and Thompson, 1968). Hence, it is reasonable to suspect that feedback might be effective in increasing academic productivity.

Public posting of performance has also proven effective in a wide range of settings. Panyan, Boozer, and Morris (1970) found that publicly

¹This research was supported by Grant A 8352 from the National Research Council of Canada to the senior author. We would like to express appreciation to the Halifax School Board and to Keith Sullivan, Principal of Alexandra School, for their support and assistance. The authors would also like to thank Dr. John Wincze for his careful reading of the manuscript, and Dr. James Sherman for comments that greatly determined the direction of this research. The second experiment was submitted by the third author as partial fulfillment for the M.Ed. at the Atlantic Institute of Education, Halifax, N.S. Reprints may be obtained from Ronald Van Houten, Department of Psychology, Mount Saint Vincent University, Halifax, Nova Scotia, Canada B3M 2J6.

posting the percentage of sessions conducted by the staff of a training school for the retarded markedly improved staff performance. In another study, publicly posting the names of children who returned to class promptly after recess decreased the number of students late for class (Hall, Cristler, Cranston, and Tucker, 1970). McKenzie and Rushall (1974) found that having swimmers publicly post their cumulative attendance and the number of training units completed increased both attendance and work output. The results of these studies suggest that public posting might have also been an effective variable in the Van Houten *et al.* (1974) experiment.

The present experiments were designed first, to determine how much each element contributed to the effectiveness of the total package in the Van Houten *et al.* (1974) experiment; second, to document experimentally the casual observation that performance comments increase with public posting of student performance. In the Van Houten *et al.* (1974) experiment, the experimenter observed a large increase in social interactions involving comments on student performance after the experimental package was introduced similar to that observed by McKenzie and Rushall (1974) after they introduced public posting. These casual observations may be of importance because peer comments affect classroom behavior (Soloman and Wahler, 1973).

The final purpose of the present research was to determine whether the variables of timing, feedback, and public posting would be effective if applied to another task, such as reading.

EXPERIMENT I

METHOD

Subjects and Setting

Children in two fourth-grade classrooms, located in a low socioeconomic multiracial area of Halifax, Nova Scotia, served as subjects. The class sizes were 19 and 20 children for Classes A and B respectively. The second author taught language arts to both fourth-grade classes at

separate times of the day. These students were divided the previous year according to their reading ability, with the better readers placed in Class A and the poorer readers in Class B.

General Procedure

Before beginning the experiment, the teacher was instructed to make a list of composition topics of approximately equal difficulty (in her judgement). Topics were then randomly ordered and one was preassigned for each day of the program. Composition writing was then introduced into the class schedule every school day except Wednesdays. Each child received a sheet of lined paper at the beginning of composition time. The topic was then written on the board and read aloud by the teacher. The class was then told to write as much as they could on the topic. While the class worked, the teacher helped children with their spelling. If a child did not know how to spell a word, the teacher would write the word on a separate sheet of paper on the child's desk. Discipline was maintained throughout with the use of instructions and reprimands. At the end of 20 min for Class A and 10 min for Class B, the teacher asked the students to stop writing and collected the papers. The teacher then moved on to the next topic.

Measures

Writing rate. All papers were scored each day by a research assistant who counted the number of words in each composition. The number of words written was converted into a rate measure by dividing by the total available writing time. Nonsensical or repetitive sentences were not included. The definitions of these two classes were the same as used in the Van Houten *et al.* (1974) study. The number of words excluded during all phases was very low. However, repetition did increase for three children in Class A during experimental conditions. One reason for this may have been the length of the composition period (20 min), possibly too long for these children. A second scorer counted words on two randomly sampled days during each experimen-

tal phase. Each scorer's counts were totalled and compared (low total score/high total score). Mean rescoring agreement for Class A was 97%, with a range of 94% to 100%. The mean rescoring agreement for Class B was 98%, with a range of 96% to 100%.

On-task recording. On-task behavior consisted of children sitting in their seats with their pencils in hand, either writing or looking at the material they had written. All other behaviors, such as looking around the room, talking to another student, or drawing on the paper were considered off-task. Children with their hands raised were always considered on-task.

On-task behavior was observed for each child using a time-sampling technique. The classroom had five rows of desks. Each row was sampled for 1 min. Every 10 sec the observer swept down the row from front to back recording the number of children in the row and the number of children on-task in that row. After six recordings, the observer moved on to the next row for another minute until all five rows had been sampled. This procedure was then repeated until the end of the writing period.

A second observer independently recorded on-task behavior on two days during each experimental phase. Both observers began recording at the start of the composition period and stopped at the end. Occurrence reliability was calculated by dividing the total number of times that the two observers agreed by the total number of observations. Mean reliability for Class A was 98.4%, with a range of 93% to 100%. Mean reliability for Class B was 95.5%, with a range of 93% to 98%.

Comments on performance. Remarks made about the composition were recorded by an observer seated at the back of the class, just before, during, and after the composition period, by marking a separate box on the recording sheet. Only clearly audible comments such as: "How did you do?", "I beat my score!", "Did you beat your score?", "How much did you write?", were recorded. A second observer seated directly behind the first observer independently recorded

comments on two days during each phase. Inter-observer reliability was calculated by dividing the larger count into the smaller count. Mean reliability for Class A was 91%, with a range of 83% to 100%. Mean reliability for Class B was 93%, with a range of 79% to 100%.

Experimental Design

A reversal design was employed with both classrooms in order to assess the relative contribution of several different variables on writing performance. After stable baseline rates were obtained in both classes, the timing and feedback phase was introduced. This phase was followed by the introduction, removal, and reintroduction of public posting, with timing and feedback remaining in effect throughout. Finally, praise was added and removed to the previous phase (A — B — B + C — B — B + C — B + C + D — B + C).

Baseline. During this phase, the children were not told that they were being timed and they received no feedback on their compositions. They were instructed not to write repetitious sentences and were given several examples of this type of sentence. The children in Class A were given 20 min a day to write and the children in Class B were given 10 min. Class B wrote for a shorter time because the teacher felt that 20 min would be too long for this group. At the end of the allotted time, the teacher said that it was time to move on to other activities and the papers were handed in. This baseline phase was in effect for five sessions for both classes.

Timing + feedback 1. During the first day of this phase, children in both sections were told that they were being timed with a stopwatch for 20 and 10 min respectively and that this procedure would continue for all further compositions. At the end of the allotted time, they were instructed to stop writing and pick up a red pen. They were then to count all words and enter their total at the top of the paper. These counts were not used for the experiment but served to provide the children with immediate feedback. The children's count was compared for accuracy

with the experimenter's count by dividing the higher of the two scores into the lower. The mean reliability of the children's counts throughout the experiment was 87.3% for Class A and 81.1% for Class B. The timing and feedback phase lasted for five sessions.

Timing + feedback + public posting 1. During the first day of this phase, each child's name was written on the board on a chart showing the greatest number of words each child had written on any composition to date. They were then instructed to see if they could beat their score. On the following day, and for each successive day during this phase, the scores on the chart were changed each time a child exceeded his highest previous score. These changes were always based upon the experimenter's data, and scores were changed just before the start of each day's composition. In addition, a check was placed beside the child's name whenever his score was changed. Hence, it was possible for each child to determine how many times he had beaten his score by counting the number of checks beside his name. Just before composition period, the children were instructed to look at the chart to see whether or not they had beaten their score. In all other respects, this phase, which lasted five sessions, was identical to the timing and feedback phase that preceded it.

Timing + feedback 2. At the start of this phase, the chart was erased from the board and the teacher announced to the class that she was no longer going to put their scores on the chart on the board. In all other respects, this phase, which lasted five sessions, was identical to the first timing and feedback phase.

Timing + feedback + public posting 2. On the first day of this session, the chart again appeared on the board. The students were shown their highest score to date and were asked to try and beat this score. This phase was identical to the first timing, feedback, and public posting phase.

Timing + feedback + public posting + praise. This phase was similar to the previous phase, with the addition of teacher praise. Chil-

dren were praised by the teacher whenever they were on-task and whenever their score on the chart changed. Some examples of praise statements the teacher used were: "Good _____, you beat your score.", "Good _____, your score came up.", "If you keep working like that you will likely beat your score _____.". This phase lasted seven days.

Timing + feedback + public posting 3. During this final phase, which lasted three sessions, praise was removed.

RESULTS

Changes in the mean number of words written per minute, the mean percentage of on-task behavior, and the total number of performance comments made during each of the experimental conditions are presented for both Classes A and B in Figures 1 and 2 respectively.

Writing rate. During the baseline phase, Class A averaged approximately 4.5 words per minute and Class B 1.8. Introduction of the timing and feedback condition increased the mean response rate of Class A to 8.3 words per

Fig. 1. The mean response rate in words per minute, the mean percentage of intervals on-task, and the number of performance comments in each daily session across all conditions for Class A. The broken vertical lines represent the introduction of the experimental conditions. The broken data line represents an unrecorded session.

Fig. 2. The mean response rate in words per minute, the mean percentage of intervals on-task, and the number of performance comments in each daily session across all conditions for Class B. The broken vertical lines represent the introduction of the experimental conditions. The broken data line represents an unrecorded session.

minute and Class B to 3.5, almost double the baseline rates.

Introduction of public posting of scores further increased mean response rates to 10.8 and 5.4 words per minute for Class A and B respectively. After removal of public posting, mean rates decreased to 7.9 and 3.4 words per minute for Class A and B. These rates are very close to the mean rates during the previous timing and feedback phase. Reintroduction of public posting increased the mean rate of Class A to 11.0 and the mean rate of Class B to 5.0.

Introduction of praise increased the rate of Class A to 13.0 but had little effect on Class B (mean rate of 5.3 words per minute). Removal of praise decreased rate for Class A to 10.0 words per minute and produced no change for Class B (6.1 words per minute).

On-task. The daily mean percentage of on-task behavior for each class shows changes that parallel changes in writing rate. This correspondence is more striking for Class B than for Class A.

Performance comments. During the baseline phase, no performance comments were observed

in Class B and only one comment was observed in Class A. After timing and feedback were introduced, the number of performance comments made in both classes increased markedly. The mean number of comments per day during this phase was 13.3 for Class A and 8.5 for Class B. A further increase to a mean of 19.8 per day for Class A and a mean of 23.0 per day for Class B followed introduction of public posting of scores and a decrease to a mean of 11.6 and 5.2 respectively followed its removal. Reintroduction of public posting increased the number of comments in Class A and B to 17.6 and 12.0 respectively. The introduction of praise resulted in a very large increase in comments for Class A and a smaller increase for Class B (Class A, 47.1; Class B, 19.0). Removal of praise resulted in a large decline in comments for Class A and a smaller drop for Class B (Class A, 15.3; Class B, 12.0). Throughout all phases, almost all comments were made either before or just after the timed writing period. Comments were rarely made during the timed writing period itself.

DISCUSSION

The results demonstrated that timing + feedback, public posting of scores, and praise each contributed somewhat to the effectiveness of the total package. Further, increased writing rate was correlated with increased on-task behavior and an increased rate of performance comments made by the children. Although each of the three experimental phases improved the performance of children in Class A, the praise condition did not seem to be effective with Class B. Why teacher praise was effective with one class and not with the other remains an unanswered question.

Although the highest rate of performance comments followed introduction of the public posting and praise condition, comments did increase with introduction of the timing and feedback phase. This increase may have resulted from the emphasis that feedback places on performance and the emphasis timing places on competition.

Story quality was also rated by two independent observers who were unaware of the purpose of the experiment. All stories were rated on five aspects of story quality as described by Van Houten *et al.* (1974). The results of the primary rater were in accord with the changes in writing rate, but the reliability rater tended to rate the stories written in all phases to be of similar quality. This failure to find reliable differences in story quality may be related to the children's poor reading and spelling skills.

One question raised by the results of this experiment concerns the effect of performance comments on writing rate. The correlation between these two variables suggests that a functional relationship may exist between comments and writing performance. Further research is needed to establish the existence of such a relationship.

EXPERIMENT II

METHOD

Subjects and Setting

Nineteen children from a fifth-grade class of 31 children taught by the third author, and located in a low socioeconomic area of Halifax, Nova Scotia, served as subjects.

General Procedure

Comprehension exercises. Before beginning the experiment, the teacher prepared all of the comprehension questions for each of the stories left in the Grade 4 reader, *Adventure Awaits*, Ginn and Company. A set of 10 questions was prepared for approximately every eight pages of the reader. These assignments were then randomly ordered, and preassigned to different sessions. Items consisted of how, why, when, where, what, and who questions about the assigned readings. At the beginning of each reading period, students were given a stencilled sheet with the 10 questions about the story and a space to write in the answer. The children were instructed to answer each question in sentence form on their stencilled sheets. Any new words

in the readings were put on the board and their meanings explained. If a student encountered additional unfamiliar words, the teacher helped him with them. Students were not permitted to talk or to help one another. At the end of 25 min, all papers were collected and the students moved on to reading workbook exercises. The assignment was scored at the end of each day by the teacher and returned on the following day. Students then went over the questions and corrected their work before beginning a new exercise. An individual who was not aware of the purpose of the experiment scored the assignments on three randomly sampled days during each phase of the experiment. Each scorer's count was then totalled and compared (low total score/high total score). Mean rescoring reliability was 95.3%, with a range of 90.5% to 98.7%.

Multiple word-meaning exercises. Before beginning the experiment, the teacher prepared all of the multiple word-meaning exercises for the duration of the experiment. Multiple word-meaning questions consisted of words taken from the Grade 4 vocabulary with a list of three different meanings. Below these meanings were three specific sentences, each representing a different use of the word. A space was provided below each sentence for the student to write in the correct definition, depending upon how the word was contextually used. Each definition was used once in each exercise.

At the beginning of the language period, each student was given a stencilled sheet with 10 words requiring three responses each and instructed to begin their work. Students were also instructed to work independently and not to talk. At the end of 15 min, the work was collected by the teacher and the children were instructed to begin their next assignment. The marking procedure for multiple word-meaning exercises was the same as for comprehension exercises. An individual who was not aware of the purpose of the experiment rescored the assignments during the three randomly chosen sessions during each phase. Each scorer's counts were

then totalled and compared (low total score/high total score). Mean rescoring reliability was 99.2%, with a range of 98.7% to 99.6%.

Experimental Design

A multiple baseline across behaviors, with reversal design was employed. After stable baselines were recorded in both comprehension and multiple word-meaning exercises, timing, feedback, praise, and public posting of scores were introduced, first during the comprehension exercises and later with the multiple-meaning exercises. A return to baseline conditions for both types of exercises followed.

Baseline 1. During this phase, the students were not told that they were being timed and received no immediate feedback on either comprehension or multiple word-meaning exercises.

Performance feedback system. On the eighth day of the program, explicit timing, self-scoring, public posting of scores, and praise for improvement was introduced during the comprehension exercises. On the first day of this condition, students were given the following instructions: "Starting today, we are going to do something different. While you are doing your reading, I am going to time you with this stopwatch and you are going to have exactly 25 minutes to do your work. At the end of this time, you'll have a chance to score your work and to see how many problems you have done correctly. You will then put your pencils in your desks and pick up your red pens and draw a line under the last problem you have completed. You will then grade it as we go along and count the number of problems worked correctly and put this score at the top of your page."

The chart was then introduced and its purpose explained. The chart was a 71 by 56 cm sheet of white bristleboard with the following information printed on it: each student's name; the days of the week; a space for each of the following, each student's daily score, the total daily group performance scores, the total weekly group performance scores, the total weekly performance scores for each student, and the cumu-

lative number of perfect papers for each student. This sheet was covered in plastic and all data were recorded with water-based pens. It was placed above the blackboard at the front of the class in clear view.

It was explained that the purpose of the chart was to see if the group as a whole could beat their highest total daily and weekly score, and whether individual students could beat their own total weekly score. The number of perfect papers was also cumulated, since the fixed number of questions given each day placed a ceiling on how many problems a child could work correctly each day. The scores on the chart were changed by the teacher at the beginning of each new session. Students were praised if they exceeded their highest weekly total and the group was praised whenever they exceeded their highest daily or weekly group score. Perfect scores were also pointed out and praise was delivered by the teacher.

The same basic procedure was followed for multiple word-meaning exercises starting on Day 13 of the program, except that the children were timed for only 15 min.

Baseline 2. At the start of this phase, both charts were removed and the teacher announced that she was no longer going to put their scores on the chart. In all other respects this phase, which lasted seven sessions, was identical to the first baseline phase.

RESULTS

The mean number of comprehension and multiple word-meaning problems worked correctly and incorrectly during all experimental phases are presented in Figure 3. The number of comprehension and multiple word-meaning problems worked correctly increased substantially after the performance feedback system was introduced; the number of comprehension problems worked correctly nearly doubled and the number of multiple word-meaning problems worked correctly increased by approximately 50%. Comprehension errors showed a slight decrease after the experimental package was in-

troduced, while multiple word-meaning errors showed no initial change. However, multiple word-meaning errors increased slightly during the last 10 days of the experimental phase. During the second baseline phase, the comprehension correct rate decreased to baseline levels while errors seemed to increase above baseline levels. Word-meaning correct rate decreased during the reversal phase while the error rate remained about the same as during the previous phase.

DISCUSSION

The results of this experiment demonstrated that the experimental package of timing, feedback, public posting of scores, and praise can effectively improve performances on reading and language exercises. This result increases the generality of the findings obtained by Van Houten *et al.* (1974), as well as the results of Experiment 1.

It is interesting to note that although the experimental contingencies were placed only upon the number of problems worked correctly, comprehension errors decreased during the experimental phase. Why a similar decrease was not observed for word-meaning errors remains an unanswered question. One way in which errors might be further suppressed would be to post them publicly along with the correct scores.

Although it was not possible for trained observers to record performance comments made by students during this experiment because of logistical problems, the experimenters did make some interesting observations. First, students frequently praised and applauded students who exceeded their highest score and second, students often praised poorer students who showed improvement. These comments tended to be similar to those observed in Experiment I.

GENERAL DISCUSSION

The introduction of timing, feedback, public posting of scores, and praise increased the performance of elementary school children in story

Fig. 3. The mean number of reading comprehension and word-meaning problems worked correctly and incorrectly in each daily session across all conditions for Classes A and B. The broken vertical lines represent the introduction of the experimental conditions.

writing, reading comprehension, and word-meaning exercises with the effects of those variables showing no sign of weakening over time. These contingencies were effective spanning work times from 10 to 25 min. Further analysis revealed that timing and feedback and public posting of scores each individually contributed to the effectiveness of the experimental package upon story writing rate while praise was effective with only one of the two classes. These results increase the generality of the previous finding that the experimental package of timing, feedback, and public posting of scores improved story writing performance (Van Houten *et al.*, 1974). Although these results further demonstrate the utility of this approach, it still remains to be established whether or not an entire academic program could be based on use of a performance feedback system. Only further research will determine the limitations of this approach.

In both experiments, all children used the same textual materials. However, there is no reason why this approach could not be employed in an "open concept" classroom, where children are working on individualized materials. This might be accomplished by dividing each individual assignment into roughly equivalent units or lessons that each child could self-score. The

number of lessons completed during timed activity periods could then be posted for each individual and for the class as a team.

Another result of interest was the marked increase in the rate of performance comments following introduction of the experimental variables. In light of Soloman and Wahler's (1973) finding that peer comments can affect classroom behavior, it is possible that these comments may have mediated many of the changes in performance observed. Further research should focus upon the direct manipulation of performance comments in order to assess their role in the classroom.

REFERENCES

- Broden, M., Hall, R. V., and Mitts, B. The effect of self-recording on the classroom behavior of two eighth-grade students. *Journal of Applied Behavior Analysis*, 1971, 4, 191-199.
- Cooper, M. L., Thomson, C. L., and Baer, D. M. The experimental modification of teacher attending behavior. *Journal of Applied Behavior Analysis*, 1970, 3, 153-157.
- Hall, R. V., Cristler, C., Cranston, S. S., and Tucker, B. Teachers and parents as researchers using multiple baseline designs. *Journal of Applied Behavior Analysis*, 1970, 3, 247-255.
- Herbert, E. W. and Baer, D. M. Training parents as behavior modifiers: self-recording of contingent attention. *Journal of Applied Behavior Analysis*, 1972, 5, 139-149.
- Leitenberg, H., Agras, W. S., Thompson, L. E., and Wright, D. E. Feedback in Behavior Modification: an experimental analysis in two phobic cases. *Journal of Applied Behavior Analysis*, 1968, 1, 131-137.
- McKenzie, T. L. and Rushall, B. S. Effects of self-recording on attendance and performance in a competitive swimming training environment. *Journal of Applied Behavior Analysis*, 1974, 7, 199-206.
- Panyon, M., Boozer, H., and Morris, N. Feedback to attendants as a reinforcer for applying operant techniques. *Journal of Applied Behavior Analysis*, 1970, 3, 1-4.
- Soloman, R. W. and Wahler, R. G. Peer reinforcement control of classroom problem behavior. *Journal of Applied Behavior Analysis*, 1973, 6, 49-56.
- Van Houten, R., Morrison, E., Jarvis, R., and McDonald, M. The effects of explicit timing and feedback on compositional response rate in elementary school children. *Journal of Applied Behavior Analysis*, 1974, 7, 547-555.

Received 18 November 1974.

(Final acceptance 21 May 1975.)