*THE EFFECTS OF EXPLICIT TIMING AND FEEDBACK ON
COMPOSITIONAL RESPONSE RATE
IN ELEMENTARY SCHOOL CHILDREN¹*

RONALD VAN HOUTEN, ELIZABETH MORRISON,
ROWENA JARVIS, AND MARGO McDONALD

MOUNT SAINT VINCENT UNIVERSITY AND ALEXANDRA SCHOOL

The effects of several variables on compositional response rate were investigated in three classrooms. After establishing baseline composition rates in each of the three classrooms, an experimental phase was introduced that consisted of: explicit timing of the children's composition period with a stopwatch, immediate feedback on the number of words each child produced, public posting of the greatest number of words written by each child to date, and instructions to try to exceed their highest score. In the second-grade classroom, these conditions were introduced, removed, and re-introduced. In two fifth-grade classrooms, these conditions were introduced according to a multiple baseline, across classes. In all cases, introduction of the experimental conditions led to a doubling of rate of words written by students and an increase in subjective quality ratings of compositions made by independent judges.

The problem of increasing academic performance is perhaps one of the more important areas of classroom behavioral analysis. An indirect approach to this problem has concentrated on reinforcing children's attending to their work (Hall, Lund, and Jackson, 1968). Since this method does not directly involve the reinforcement of academic behavior, it has not always produced the desired increase in academic response rate (McLaughlin and Malaby, 1972). A more direct method involves reinforcing children's production of larger amounts of work (Ferritor, Buckholdt, Hamblin, and Smith, 1972; McLaughlin and Malaby, 1972) or high rates of work (Chadwick and Day, 1971; Hopkins, Schutte, and Garton, 1971; Lovitt and Curtiss, 1969; Lovitt and Esveldt, 1970).

Although the above approaches have effectively increased academic productivity, it should

be noted that in each of these experiments, points and backup reinforcers were employed. The logistics of administering token programs can consume a considerable portion of a teacher's time, because reinforcers must be provided and tokens earned and spent must be accounted for. In addition, not all teachers may feel comfortable introducing tokens or backup reinforcers into their classrooms. One alternative to a point system is the use of precise academic feedback.

The purpose of the present experiment was to determine whether explicit timing, immediate feedback, and public posting of highest scores without praise could increase the academic response rate of an entire class of children. Written composition was chosen as the target behavior for two reasons: it produced an objective, inexpensive, permanent record of behavior and both teachers had difficulty in getting their children to write compositions. Brigham, Graubard, and Stans (1972) found that points, which were exchangeable for classroom privileges, delivered for different aspects of compositional behavior, affected number of words written and number of different words written, as well as subjective judgements of story quality. Maloney and Hopkins (1973) showed that points, which were exchangeable for extra recess time and candy and

¹This research was supported by Grant A 8352 from the National Research Council of Canada to the senior author. We would like to give thanks to the Halifax School Board, and especially to Keith Sullivan, Principal of Alexandra School, for his support and friendly assistance. The authors would also like to thank Dr. John Wincze for his careful reading of the manuscript. Reprints may be obtained from Ronald Van Houten, Department of Psychology, Mount Saint Vincent University, Halifax, Nova Scotia, Canada.

which were delivered for different aspects of composition writing, increased story quality, according to subjective judgements.

METHOD

Subjects and Setting

The subjects were children in one second- and two fifth-grade classrooms located in a low socioeconomic multiracial area of Halifax, Nova Scotia. The class sizes were 21, 17, and 17 students respectively. The fifth-grade teacher taught language arts to both grade-five classes; the second-grade teacher taught all of the academic subjects to her class.

General Procedure

Before the experiment, both teachers were instructed to make a long list of composition topics which, in the teacher's judgement, were of approximately equal difficulty. Items in each list were then randomly ordered, and a topic was then pre-assigned for each day of the program. Following this, composition writing was introduced into the daily class schedule. Each child received his paper at the beginning of the time set aside for composition. The topic was then written on the board and the class was told to write as long a composition as they could. While the children were writing, the teacher sat at her desk and watched the children write. Discipline was maintained in both classes with the use of instructions ("sit in your seat and do your work please, Johnny") or on rare occasions with reprimands. At the end of 10 min, the teacher asked the students to stop writing, collected the papers, and moved on to the next topic. Papers were scored at the end of each day by the fourth author, who counted the number of words in each composition. Nonsensical or very repetitive sentences were not included in the word count. A nonsensical sentence was defined as a random list of words not comprising a sentence and a repetitious sentence was defined as a sentence that immediately followed another in which all words were the same except for a noun; *i.e.*,

"The rabbit sat on the paper. The rabbit sat on the desk. The rabbit sat on the book". The number of words excluded by these criteria was never more than 5% throughout the experiment. The experimenters usually accepted 100% of the words written by each child. At no point during the program were the children systematically praised for writing. An individual who was not aware of the purpose of the experiment counted words on two randomly sampled days during each experimental phase. Each scorer's counts were totalled and compared (low total score/high total score). Rescoring agreement was computed for the second-grade class on two or three days in each experimental condition. The mean rescoring reliability was 99%, with a range of 96% to 100%. Rescoring agreement for the fifth-grade classes was computed for three baseline and three experimental sessions for each class. The mean reliability for section A was 98%, with a range of 95% to 99%. The mean reliability for section B was 99.5%, with a range of 99% to 100%.

Second-Grade Class

Each day, the children were given a composition to write during the first afternoon period. The initial phase of the composition program was termed the baseline phase. During this phase, the children were not told that they were being timed and they received no feedback on their compositions. However, they were instructed to try to write as many words as they could. They were further instructed to try not to write repetitious sentences and were given several examples; *i.e.*, "Bill saw a man. Sue saw a man. John saw a man". At the end of the 10 min, the teacher said that it was time to move on to other activities and that papers should be passed in.

Following the baseline phase, the children were shifted to the timing and feedback condition. During the first day in this condition, the children were shown a chart on the board displaying the maximum number of words each student had written on any composition to date. They were then told that they would be timed

for 10 min with a stopwatch on all further compositions. At the end of 10 min, everyone would have to stop writing.

The object of this, it was explained, was to see if they could beat their own highest score. They were further instructed that words that did not conform to good sentences and the repetitious use of sentences would not be counted. Several examples were then provided for each type of sentence.

At the end of the 10-min period, each child had to put down his pencil and pick up a red pen. They then proceeded to count all words and enter their total on the top of the paper. These counts were not used for the purposes of the experiment but instead served to give the children immediate feedback. The children's count was compared for accuracy with the experimenter's count by dividing the lower of the two scores by the higher score. The mean reliability of the children's counts was 89% across both experimental phases with a range of 80% to 95%. On the following day and for each successive day during this phase of the experiment, the scores on the chart were changed every time a child exceeded his highest previous score. These changes were always based upon the experimenter's data and were always made during the lunch period when no students were in the classroom. All changes in score were circled and at the beginning of each writing period, the children were instructed to look at the chart on the board to see if they had beaten their score.

The next condition was a return to baseline. At the beginning of this phase, the teacher announced to the class that the compositions were no longer going to be timed but that they were to continue to try to write as much as they could; they were again reminded not to write repetitious or nonsensical sentences. The charts were removed from the board and the class was told that they would no longer count their words and that scores would no longer be charted. The teacher continued to call for the collection of compositions at the end of the 10 min, but this was done in such a way as to minimize the likeli-

hood of the children's knowing that they were being timed; *e.g.*, "We'd better stop, it's time for math" or "We have a lot of work to do this afternoon, I think we'd better stop".

The final phase of the experiment replicated the second phase. During this condition, timing and feedback were re-introduced and remained in effect until the end of the school year.

Fifth-Grade Classes

Because the end of the school year was near, a multiple baseline across classrooms design was chosen. This design required fewer sessions in order to provide an indication of scientific reliability.

The initial phase or baseline phase for both fifth-grade classes was identical in all important respects to the baseline phase for the second-grade class. The children were not told that they were being timed and they received no feedback on their compositions.

On the eighth day of the program, timing and feedback were introduced for one class (class B) and the other class remained in the baseline condition. On the eleventh day, timing and feedback were also introduced for the remaining class (class A). The timing and feedback phase was identical to the first timing and feedback phase for the second-grade class.

The student's count was compared for accuracy with the experimenter's count in the same manner as was done with the second-grade class. The mean reliability for section A was 94%, with a range of 91% to 97%, and 95% for section B, with a range of 93% to 98%.

In addition to the objective measures of response rate, two measures of subjective quality were employed. The first measure was the same as that employed by Brigham *et al.* (1972) to determine changes in the quality of their stories. Five dimensions were employed: (1) mechanical aspects, length, spelling, grammar, punctuation; (2) vocabulary, variety, and word usage; (3) number of ideas; (4) development of ideas; (5) internal consistency of the story. All dimensions were scored from zero to five in steps of 0.5.

No operational definitions were provided for any of the measures. The judge was a graduate student in the German department at Dalhousie University with a double undergraduate major in English and German. The judge had no knowledge of the design of the experiment or even that it was an experiment. She was simply told to rate the set of stories by the system described to her. The judge rated stories from each child present from at least two randomly selected days during each experimental phase in the second-grade class. In the two fifth-grade classes, the judge rated stories from each child present from four randomly selected days from each experimental phase.

Furthermore, two elementary school teachers rated randomly selected stories using a different method. These raters were not aware of the experimental procedures or results.

Each judge was given two pairs of compositions written by each of the students in all of the classes. One composition from each pair was written during the baseline period and the other was written during the timing and feedback phase. Each judge therefore compared a composition from the baseline phase to one from the experimental phase twice for each subject in each class. Their task was to rate which of the compositions in each pair was of better quality. They were further instructed to use the criteria that they would use in their own classroom. When these ratings were completed, each judge independently rated which of the two baseline compositions written by each child was best. Both independent raters were given the same sets of compositions at different times, so that occurrence reliability could be computed on their judgements. Percentage agreement between the two judges was calculated by dividing the number of ratings that they agreed on by their agreements plus disagreements. After the teachers had rated all of the compositions they were separately asked in what areas they found the compositions to be better. During this interview, each judge had free access to all of the compositions he had rated.

RESULTS

Changes in the mean number of words written per minute during each of the different experimental conditions are presented in Figure 1 for the entire second-grade class. During the first baseline phase, the class averaged approximately three words per minute.

Introduction of the timing and feedback condition produced a marked increase in response rate, the mean class response rate being more than double the baseline rate.

During the second baseline condition, the response rate decreased, although it did not drop to the level evidenced during the first baseline condition. One possible reason for this failure to recover completely the earlier baseline performance was that some of the children were observed to be still keeping track of the number of words produced, despite instructions to the contrary.

When timing and feedback were again introduced, rate again increased. By the end of the school year, the rate had increased to three times that during the first baseline phase.

Individual Subject Data

After timing and feedback were introduced, the rate of word production increased for all of the subjects. During the reversal phase, 13 of the subjects returned to their original baseline levels; eight subjects did not.

The data of four individual subjects are presented in Figure 2. The two graphs at the top (Subjects A.B. and D.S.) were typical of the results of subjects whose rates returned to their original level. The two graphs at the bottom (Subjects P.R. and M.S.) show results that were typical of subjects who did not show a return to baseline rates of responding during the reversal phase.

The results of the two fifth-grade classes are presented as mean response rates in Figure 3. These results also indicate that introduction of timing and feedback increased the response rate of composition writing.

GRADE 2

Fig. 1. The mean response rate in words per minute for the entire Grade-two class in each daily session across all conditions. The broken vertical lines represent the introduction of the experimental conditions.

The rates of most subjects in both sections showed a marked increase after the timing and feedback phase was introduced. However, some of the subjects in both sections showed less dramatic increases. The data of two subjects in section A (Subjects A.M. and S.G.) and two subjects in section B (Subjects M.A. and B.V.) are presented in Figure 4. The two graphs at the top show the results of two subjects who were typical of highly responsive subjects. The two graphs at the bottom show the results of two subjects who were typical of subjects less responsive to the introduction of the experimental phase.

Quality Ratings

The graduate student judge rated stories on five dimensions. The overall mean ratings of

Fig. 2. The response rate in words per minute of four individual children from the Grade-two class in each daily session across all conditions.

Fig. 3. The mean response rate in words per minute for both Grade-five classes in each daily session. The broken vertical lines represent the introduction of the experimental condition.

stories (25 points possible) for each phase in the second grade were: baseline 1, 6.4; timing and feedback 1, 12.8; baseline 2, 10.1; timing and feedback 2, 12.7. The failure to achieve com-

plete recovery of baseline ratings is in accord with the failure to recover baseline response rates.

The mean score for each of the five categories for each day that stories were rated is presented for the second-grade class in Table 1. This table shows that the increase in the total rating resulted from an increase in all aspects of the compositions evaluated.

The overall mean ratings of stories in the fifth-grade for each phase were: baseline, 6.9; timing and feedback, 13.3 for section A and baseline, 11.4; timing and feedback, 15.8 for section B.

The mean score for each of the five categories for both fifth-grade classes is presented in Table 2.

This table again shows the increase in ratings to result from an increase in all aspects of composition evaluated.

The results of the elementary teacher ratings supported the above finding. They found the compositions written during the timing and feedback phase to be superior to compositions written during the baseline phase the following percentage of times: Grade 2, 83%; Grade 5 section A, 88%; Grade 5 section B, 86%. The intersubject reliability of these judgements was 85%, 89%, 83% respectively. When rating the

Table 1

The mean rating for each of the five dimensions on each of the sampled days for the Grade-two class. An A indicates that the student was absent.

Condition: Day:	Baseline 1		Timing and Feedback 1		Baseline 2		Timing and Feedback 2			
	3	5	7	9	12	14	18	20	23	27
Mechanical aspects	1.3	1.2	2.4	2.5	2.1	2.0	2.3	2.8	2.1	2.8
Vocabulary	1.3	1.1	2.4	2.5	1.9	1.9	2.5	2.6	2.1	2.7
Number of ideas	1.7	1.3	2.8	2.6	2.1	2.2	2.8	3.0	2.2	3.2
Development of ideas	1.3	1.1	2.4	2.4	2.1	2.0	2.5	2.5	2.1	2.7
Internal consistency	1.5	1.3	2.6	2.6	2.0	2.1	2.5	2.6	2.2	2.8

two baseline compositions, the judges found the earlier of the two baseline compositions to be superior the following percentage of times: Grade 2, 50%; Grade 5 section A, 43%; Grade 5 section B, 50%. The intersubject reliability was 66%, 78%, and 50% respectively. Thus, although the raters saw the compositions written during the timing and feedback phase to be clearly superior to compositions written during the baseline phase, they did not rate the baseline compositions as different from each other. When asked in what way the higher rated compositions were better, both teachers said that ideas were better expressed and developed.

DISCUSSION

The results demonstrated that timing with simple and differential feedback increased the

rate of composition writing in elementary school children. Further increased writing rate was correlated with increased ratings of story quality by independent judges. The latter finding is similar to that reported by Brigham *et al.* (1972) with the exception that they reported only increases in mechanical aspects of writing and vocabulary when they increased the number of words written; in the present study, all five categories showed improvement.

In addition to the above findings, the classroom teacher made several interesting casual observations. Both teachers reported an improved "attitude" toward writing on the part of the children and a sharp decrease in social interactions while the children were writing during the experimental phase. It is unfortunate that objective data are not available to support these observations.

Fig. 4. The response rate in words per minute of four individual children from both Grade-five classes in each daily session. The two graphs at the top of the figure (Subjects A.M. and M.A.) show results of two subjects who were highly responsive and the graphs at the bottom (Subjects S.G. and B.V.) show the results of two subjects who were not very responsive.

Table 2

The mean rating for each of the five dimensions on each of the sampled days for both Grade-five classes.

Condition: Day:	Grade 5A							
	Baseline				Timing and Feedback			
	2	4	5	7	8	10	12	15
Mechanical aspects	1.3	1.3	1.3	1.8	2.6	2.8	2.6	3.3
Vocabulary	1.0	1.3	1.2	1.6	2.3	2.4	2.1	3.0
Number of ideas	1.1	1.4	1.3	2.1	2.5	2.5	2.1	3.0
Development of ideas	1.3	1.5	1.2	1.8	2.6	2.6	2.3	3.2
Internal consistency	1.4	1.5	1.2	1.9	2.6	2.7	2.4	3.3

Condition: Day:	Grade 5B							
	Baseline				Timing and Feedback			
	2	4	5	9	11	12	13	15
Mechanical aspects	2.2	2.0	2.5	1.8	3.4	3.2	3.4	3.3
Vocabulary	2.4	1.9	2.4	1.6	3.2	3.0	3.1	3.0
Number of ideas	2.8	2.3	2.8	1.9	3.1	2.6	3.5	2.8
Development of ideas	2.7	2.3	2.7	1.9	3.0	3.1	3.5	2.9
Internal consistency	2.8	2.3	2.8	1.9	3.1	3.1	3.7	3.1

Increased social interaction was observed by the experimenters just before and just after the writing. These interactions seemed to centre around the publicly posted scores. All of the authors noted that children made comparisons during the experimental phase of the experiment. In all classes, teachers noted comments such as: "Hey! I beat my score. How many words did you write? Hey! Look what — got, or — is the highest". There were also less-frequent statements of a negative variety, such as "I made more than you did". It is quite possible that these behaviors generated by the public posting of differential feedback facilitated the increased output in writing rate.

The gains in rate obtained in both the second- and fifth-grade classes in the present study were very similar to the gains reported by Brigham *et al.* (1972) with a fifth-grade adjustment class. In the present study, the gain in word productivity was achieved by the use of timing and several feedback procedures, whereas in the Brigham *et al.* study, points and backup reinforcers were employed.

In the Brigham *et al.* experiment, the number of points given was directly dependent on the number of words produced. Therefore, the number of points obtained provided the children with feedback as to how well they had done. Hence, it is possible that the major factor in increasing

rate may have been feedback, rather than the backup reinforcers provided. It is not possible to assess this possibility because the only control condition employed was points contingent upon attending to work. In order to assess properly the role of feedback, it would be necessary to provide a phase where backup reinforcers are not provided while continuing to provide points for improvement.

It should also be noted that Brigham *et al.* applied contingencies to other aspects of composition, such as the number of new words used in a given composition. Feedback could easily be made contingent upon these aspects of composition as well.

The use of timing and feedback procedures seems to offer several logistical advantages over a token or point system. Backup reinforcers need not be provided, which could save either money (if tangible reinforcers are employed) or teaching time (if free time or noneducational games are used). Also, there is no need to keep track of points earned or spent.

Another strength of the present method is that it offers an alternative technique for the modification of writing. Since one method might not appeal equally well to all teachers, the availability of alternative methods is highly desirable. It is possible that the idea of academic feedback might prove appealing to some teachers who would not be interested in using a token or point system or *vice versa*.

It should be emphasized that the present procedures involved timing, self scoring of responses, public display of scores, and instructions to students to beat their previous high scores. The contribution of each of these components to the effectiveness of the total package remains to be evaluated. Also, the procedures were used for

only 10 min each day. Whether similar effects would be obtained over longer time periods remains to be determined.

REFERENCES

- Brigham, T. A., Graubard, P. S., and Stans, A. Analysis of the effects of sequential reinforcement contingencies on aspects of composition. *Journal of Applied Behavior Analysis*, 1972, 5, 421-429.
- Chadwick, B. A. and Day, R. C. Systematic reinforcement: academic performance of underachieving students. *Journal of Applied Behavior Analysis*, 1971, 4, 311-319.
- Ferritor, D. E., Buckholdt, D., Hamblin, R. L., and Smith, L. The noneffects of contingent reinforcement for attending behavior on work accomplished. *Journal of Applied Behavior Analysis*, 1972, 5, 7-17.
- Hall, R. V., Lund, D., and Jackson, D. Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis*, 1968, 1, 1-12.
- Hopkins, B. L., Schutte, R. D., and Garton, K. L. The effects of access to a playroom on the rate and quality of printing and writing of first- and second-grade students. *Journal of Applied Behavior Analysis*, 1971, 4, 77-87.
- Lovitt, T. C. and Curtiss, K. A. Academic response rate as a function of teacher and self-imposed contingencies. *Journal of Applied Behavior Analysis*, 1969, 2, 49-53.
- Lovitt, T. C. and Esveldt, K. A. The relative effects on math performance on single- versus multiple-ratio schedules: a case study. *Journal of Applied Behavior Analysis*, 1970, 3, 261-270.
- Maloney, K. B. and Hopkins, B. L. The modification of sentence structure and its relationship to subjective judgements of creativity in writing. *Journal of Applied Behavior Analysis*, 1973, 6, 425-433.
- McLaughlin, T. F. and Malaby, J. Intrinsic reinforcers in a classroom token economy. *Journal of Applied Behavior Analysis*, 1972, 5, 263-270.

Received 13 August 1973.

(Revision requested 29 November 1973.)

(Revision requested 23 April 1974.)

(Revision requested 21 May 1974.)

(Final acceptance 22 July 1974.)