*WORD GAME BINGO: A BEHAVIORAL TREATMENT
PACKAGE FOR IMPROVING TEXTUAL
RESPONDING TO SIGHT WORDS*

KIMBERLY C. KIRBY, STEPHEN W. HOLBORN, AND
HARRY T. BUSHBY

UNIVERSITY OF MANITOBA AND MINNETONKA ELEMENTARY SCHOOL

Six third-grade students identified as deficient in reading skills tested the efficacy of word game bingo for acquisition and retention of sight word reading. The design was a modified multiple baseline in which treatment was implemented over 3 of 4 word sets and terminated on earlier sets when commencing treatment on later sets. Four sets of bingo cards were constructed on 7 × 9 cm paper divided into 25 equal-sized boxes. Sight words of each set were randomly placed into 24 of these boxes (the center box was marked "free"). Bingo winners were given tokens which were traded weekly for reinforcing activities. Noticeable improvements occurred for the word sets receiving the game treatment (sets A to C). Mean percentage points of improvement from baseline to treatment were approximately 30%. Terminal levels of correct responding exceeded 90%. Several variations of the game were suggested for future research and word game bingo was advocated as an effective behavioral technique for teachers to train sight word reading.

DESCRIPTORS: reading skills, sight words, multiple baseline, evaluation, education

Although behavioral principles have been applied and evaluated for the problems of teaching a variety of reading skills (e.g., Corey & Shamow, 1972; Staats & Butterfield, 1965; Van Houten, Hill, & Parsons, 1975; Whitlock & Bushell, 1967), very little has been done with specific reference to the acquisition of sight word vocabulary. Sight words are words that do not follow phonetic rules and must therefore be learned by memorization. Lahey and Drabman (1974) targeted this type of word in training second-grade students. They compared the relative effectiveness of verbal feedback alone with verbal feedback combined with token reinforcement. One-to-one training procedures were used for both treatment groups. Although rate of cor-

rect responding increased for both groups, those in the token group reached criterion in significantly fewer trials and had significantly better scores on retention tests as compared to the feedback-only group. Lahey and Drabman concluded that the token system was a superior training strategy, and they noted that for both strategies specific training for retention was unnecessary.

More recently Neef, Iwata, and Page (1977) examined acquisition and retention of sight words in a one-to-one training procedure with mentally retarded adolescents. They suggested that the procedures of shaping, differential reinforcement, and fading as traditionally used may not be the most efficient procedures for teaching some academic behaviors and proposed interspersing trials of previously trained items with trials of unknown items. During interspersal training sessions 10 known words from a pretest were alternately presented with 10 test words that were incorrect on the pretest. In baseline sessions, 10 words were presented without alternation of previously known words. The in-

---

The authors express their sincere appreciation to Heather Mullen and Rayleen De Luca for their dedication in implementing training procedures and to Joanne Swain for her cheerful assistance in collection of assessment data. Reprints of this article may be obtained from Stephen W. Holborn, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2.

terspersal procedure led to greater acquisition and retention of sight words. It was suggested that inclusion of known items helped maintain attending behavior to unknown items during reading instruction, presumably resulting from increased reinforcement due to increased correct responding during the early stages of skill acquisition.

Although both of the aforementioned techniques produced desirable results, they require the time and individual attention of a trainer. This is a serious problem for the classroom teacher who is not able to devote a great deal of time to preparing individual exercises and giving individual attention to each "slow reader." A training procedure that requires a minimum of the teacher's time, while providing maximum benefit to a number of students is needed.

Teachers have, for some time, been aware of the ability of games to fulfill this criterion with regard to reading instruction. Books have been dedicated to the topic of games for strengthening reading skills; many games have been suggested in teaching manuals; and various publishers have game kits available for purchase (Shankman, 1972). Thompson (1973) devoted a chapter of his book specifically to games for enlarging reading vocabularies. He suggested a diversity of card games and board games as well as two modifications on the traditional game of bingo. He reported that the games produced great learning enthusiasm, but no empirical data were evinced to verify the usefulness of games for increasing correctly read words. Indeed, there appear to have been no systematic evaluations indicating the serviceability of games for teaching reading skills.

Behavior modifiers, though familiar with the merit of games for producing desirable classroom behavior (Barrish, Saunders, & Wolf, 1969; Harris & Sherman, 1973; Medland & Stachnik, 1972), good oral hygiene habits (Swain, Allard, & Holborn, *in press*), creativity in writing behavior (Maloney & Hopkins, 1973) and increases in work output for the developmentally disabled (Lutzker & White-Blackburn,

1979), have apparently not researched the applicability of this kind of contingency control for reading behavior. The present investigation examined this issue using a modified bingo game for the training of appropriate textual responding to sight words.

METHOD

Subjects

Six students, four females and two males, ranging in age from 8 to 10 yrs were chosen from a third grade classroom. The resource teacher selected the participants, on the basis of his assessment of poor reading ability and difficulty pronouncing sight words.

Materials

Four sets of words, each containing 24 sight words, were randomly drawn from one of the resource teacher's manuals (Cordps, 1965). These words were identified by the manual as being third-grade reading material. Each of the sight words was printed with black felt pen on a blank 4×6 cm playing card. The cards were then grouped into their respective word sets with the 24 words in a standardized order within each set.

Each bingo card consisted of a 14×22 cm piece of laminated paper with lines dividing it into 25 equal-sized boxes. The 24 sight words of each set had been randomly typed in the boxes on each of six bingo cards. Thus, bingo cards for each 24-word set had the same words on them; however, the placement of the words differed for each card. The center box on each card was always marked "free." Clear plastic bingo chips were also used.

Procedure

Behavioral measures. Assessments of each child's reading were conducted every Monday, Wednesday, and Friday (except holidays) between the hours of 1:00-1:45 p.m. Children were taken one at a time to a small table in a secluded corner of the library and told that the experimenter would help them with their read-

ing, but first she wanted to know how well they could read already. The child was then presented with one set of the playing cards. In order to expedite data recording, the words were arranged in a standardized order. The child was asked to read through the set of words as quickly as possible and was told that if the word was not read quickly enough, she or he would be asked to go on to the next word. After the first set of words was read, there was about a 15-sec rest while the experimenter prepared the next data sheet. The second set of words was then presented and so on until all four sets had been tested. Word sets were presented in a counterbalanced fashion across days, although the daily order of presentation was consistent across children.

For sets A, B, and D the experimenter remained silent once the child started flipping through the deck of playing cards and reading the words (except to say "next word," if the child took more than approximately 3 sec to say the word). For set C this format was initially followed, but after treatment commenced on this word set, the experimenter began giving feedback to the children. Words that were read correctly were followed by a positive comment from the experimenter, such as "good" or "right." When a child read a word incorrectly or hesitated for more than 3 sec, the experimenter gave the correct response and had the child repeat it while looking at the word. Correctly repeating the word was also reinforced with verbal feedback. There were two reasons for this alteration in procedure. First, the children were regularly requesting feedback on their performance. During assessments they would frequently try to peer at the experimenter's recording sheet and, subsequent to completing the word set, would ask if they had answered all words correctly. Second, several children had a tendency to perseverate on incorrect answers to one or two of the sight words in a set. Feedback during assessments was proposed as a possible means of correcting this problem.

Approximately once every 2 wk a second ob-

server recorded during assessments to provide some measure of interobserver reliability. The observer sat across the table from the experimenter and the child. The experimenter and the observer avoided eye contact and interaction with each other during recording; thus simultaneous but independent records were obtained. Agreement was measured by comparing the two data sheets for agreement, response by response, calculating the ratio of agreements to total responses (agreements plus disagreements) and multiplying by 100.

Reinforcer survey. After 1 wk of assessments, reinforcer survey schedules (Cautela, 1977) were administered to the children individually after their assessment had been completed for that day. After collecting all six surveys, the responses were examined and a variety of activities were selected to serve as backup reinforcers. Activities were chosen in terms of high preference ratings by the children and of convenience in making the reinforcer available. Token costs were assigned according to these same factors.

Introduction of the game. The bingo game was introduced after the first week of assessments and was played every school day thereafter between the hours of 2:45 and 3:30 p.m. All games were played at the same small table in the library where the assessments took place. Upon arriving at the school, an assistant, who was blind to assessment results, arranged the bingo cards and chips on the table and escorted the six children from the classroom to the library. Once seated at the table, the assistant and the children jointly decided what the criterion for a "bingo" would be: one line any direction, one line horizontal, vertical, or diagonal in direction, an "X," or two lines any way.

Words were called from the same playing cards that were used during testing. Before beginning to call the words, the assistant shuffled the set of cards so that presentation occurred randomly. The assistant then drew the card from the top of the deck, announced the word to the children, then showed the card for approximately 15 sec. Duration of card presentation

was quickly faded so that by about the fourth day after treatment implementation, children were responding only to the announcement of the word. The children were told to find the word on their bingo card and to cover it with a bingo chip. They were encouraged to seek their neighbor's help if they had difficulty locating the word. Part of the actual training process, therefore, was executed by the child's neighbors.

The assistant waited until each child had located the word on the card before proceeding to the next word. When one of the children achieved the previously agreed upon criterion, she or he would call "bingo!" and the assistant would check the card to ensure that the correct words had been covered in the line or lines required. If an error had been made, the incorrectly covered word was uncovered and the game continued. If the child had correctly covered the words, a star was rubber-stamped by her or his name on a poster which was displayed nearby. Cards were then cleared for the next game. Games continued sequentially until about 45 min of bingo playing had occurred. Approximately six games were completed daily.

On Fridays, fewer games were played because bingo terminated at 3:00 p.m. leaving one-half hour for reinforcement. At this time children were able to trade in stars for a choice of activities such as coloring, painting, reading, playing with model cars and trucks, or playing with plasticene. Stars were simply crossed off the chart as the children traded them. On two Fridays, when the children had saved up enough stars to go on a special outing for ice cream or beverage, bingo was not played at all. To allow enough time to participate in this activity, the children were gathered at 2:30 p.m., directly after recess.

Experimental design. The bingo game was played with only one word set at a time, starting with set A and proceeding to set C, with 2½ wk of bingo being played on each set. Treatment periods of 2½ wk were used primarily due to time constraints; however, the children's performance rates were monitored to ensure that of the chil-

dren who attended regularly, a minimum criterion of 80% had been achieved during the last three consecutive assessments. The fourth word set, D, was never used in the bingo game. This continuous baseline assessed any effects of repeated exposure to the words or of the learning occurring in the regular classroom situation.

The sequential administration of treatment on sets A to C was used in constructing the experimental design which was a variation of the multiple baseline design. The specific variations from the standard multiple baseline design were twofold. First, the target for sequential implementation of treatment was atypical. The multiple baseline is usually introduced across subjects, situations, or behaviors (Hersen & Barlow, 1976). Here multiple baselines across the content of the material were used. All children played the same game at the same time, the situation remained constant, and the behavior treated was always reading sight words. It was only the actual set of sight words that changed from treatment implementation to treatment implementation. Thus, a multiple baseline across materials was used.

A second, minor alteration was made in the standard procedure. When treatment was implemented on a new set of words it was terminated on the previous set of words, contrary to the usual situation where treatment continues in the previous conditions. This approach was taken because it was thought, as suggested by Lahey and Drabman (1974), that after reaching criterion, performance was likely to remain fairly stable without programming for generalization over time. An additional consideration was convenience of treatment procedures.

RESULTS

Observer Reliability

Reliability of assessment procedures was analyzed on four occasions. Agreement on correct responses was 96.2, 95.6, 98.4, and 99.7% from reliability checks 1 to 4, respectively. This represents a mean rate of agreement of 97.7%. Dis-

agreement, although infrequent, occurred in the rare instances when a child responded to a stimulus word with slight changes in pronunciation or when subtle discriminations were required in determining the correctness of an answer. (It should be noted that the vast majority of incorrect responses were simply failures to respond or totally incorrect words.) To illustrate, a situation leading to discrepancy in scoring arose when a child would respond "wonder" to the word stimulus "wander." In this case, slight changes in the pronunciation of these two words created difficulty in reliability agreements.

Reading Performance

The measure of reading performance was stated in terms of the percentage of words read correctly during each assessment of each word set. As the group data in Figure 1 show, initial baseline levels of correct responding are in the area of 45-60% on all word sets. The initial level of correct responding appears to be slightly higher on set D as compared to other word sets. Here the baseline rate begins above 55%, quickly rises, and initially stabilizes around the 60% level. Sets A through C show an initial rate of approximately 50% and increases to the 60% level are more gradual. The trend for a gradual increase in baseline responding is present in all word sets. Increases of approximately 10-15% occur across all baselines with the higher percent increases occurring in the more extended baselines.

Slight upward drifts in baseline, though typically problematic, were in this case to be expected due to the effects of repeated exposure to the words and the learning occurring in the regular classroom situation. The upward trends did not prove troublesome as for all three word sets that received treatment, increases in level of correct responding were immediate. In all three cases the first assessment after treatment show a sharp jump such that no overlapping data points can be detected between baseline and treatment. During treatment, performance consistently increased and eventually stabilized

above 90% correct responding. For set C, where corrective feedback was given during assessments, responding stabilizes above 95% correct by the fourth assessment after treatment. For sets A and B, slight increases in correct responding are noted after treatment terminated. During this posttreatment retention period, responding stabilizes around 95% for these word sets.

These overall trends are generally supported by examination of the individual data. Representative data for two of the six children are presented in Figures 2 and 3. Wendy's data (Figure 2) were selected, as they are representative of clear treatment effects. For Wendy, baselines are quite stable and responding increases rapidly after treatment implementation leaving only one overlapping data point between baseline and treatment phases (set C). For all three treated word sets, responding reaches 90% correct by the second assessment after treatment implementation. Responding on set D, which received no treatment, increases only slightly and never exceeds 80% correct.

Cathy's data are shown in Figure 3, as they represent the least potent (although still clear) treatment effects. Baselines are less stable; gradual increases of up to 30% are present during baseline periods. Increases after treatment implementation are less dramatic than in Wendy's case; an overlapping data point is present for both word sets B and C. For word set A, responding stabilizes around 95% by the fifth assessment after treatment implementation; however, this level is never exceeded. On set B, responding stabilizes above 95% by the sixth assessment after the treatment phase commences. On set C increases in responding are more dramatic with stabilization above 95% occurring by the fourth assessment after treatment. Cathy's responding on set D increased 30%, but as is the case for Wendy, responding never exceeds 80% correct for the untreated set of words.

Improvements in the children's responding from baseline to treatment phases ranged from approximately 20 to 40%. Mean percentages of improvement from sets A to C were 30.8, 31.0,

Fig. 1. Mean percentage of correct responses for the group on word sets A through D.

WENDY

Fig. 2. Percentage of correct responses for Wendy on word sets A through D.

CATHY

Fig. 3. Percentage of correct responses for Cathy on word sets A through D.

Table 1

Improvement of mean percentage points of correct responding from baseline to treatment phases for individual children.

<i>Child</i>	<i>Word Set</i>		
	<i>A</i>	<i>B</i>	<i>C</i>
Cathy	29.1	35.5	38.0
Carman	35.8	24.9*	37.8
Chris	29.2*	24.2*	33.8*
Robert	41.4	35.2	20.8*
Shelly	21.2	26.5	25.2
Wendy	28.0	39.8	36.8

*Treatment days missed.

and 32.1, respectively. Improvements below 25% typically occurred for children who were absent from some of the treatment sessions or had baseline levels of responding that were higher than those of the other children. This can be seen in Table 1 which represents individuals' mean improvements. Scores with asterisks (indicating absenteeism during treatment) are typically lower, with the exception of Chris's score on set C. For Chris, the first three treatment days were missed, yet performance increased sharply after attendance became regular. Lower changes in mean percentage of improvement were achieved by Shelly despite regular attendance. For Shelly, baseline rate of responding was typically somewhat higher than for the other children, leaving less room for improvement.

DISCUSSION

The present study illustrates the effectiveness of word game bingo as a treatment package for training appropriate textual responding to sight words. Introduction of the game to a new set of words consistently led to increases in the percentage of words read correctly. In most instances the changes were immediate and of a magnitude that is clinically significant (i.e., 30%).

The design did not permit clear assessment of feedback for reducing perseveration of incorrect responses on set C words, but casual observation during assessment suggested that this pro-

cedure did alleviate the problem. Such perseveration occurred when a child responded to a word stimulus card with another word quite similar to the one presented. Often the incorrect response was correct for one of the other 96 words used in the study. Although feedback did not eliminate perseveration (a child occasionally slipped back to the incorrect response), the frequency of such errors did appear to decrease. An experimental analysis of the contribution of feedback would be a suitable focus for a future study.

The stability of high levels of responding during the posttreatment retention phases of the study not only has positive implications with regard to the applied value of this training procedure, but also lends support to the suggestion of Lahey and Drabman (1974) regarding the phenomenon of stable performance during posttreatment without training for generalization over time. Lahey and Drabman's results suggested that the same procedures that facilitate acquisition also facilitate retention. The retention data shown on sets A and B in this study suggest that this may be so; however, the influence of frequent assessments cannot be overlooked. It may be that the repeated practice of assessments aided in maintaining high, stable retention levels. Periodic follow-up probes are needed to eliminate this factor. Such probes were not conducted because the school year ended and the children left for summer vacation.

Word game bingo has several appealing features for the applied researcher. It does not require extra time in preparing lessons nor a high degree of individual attention. The group training situation is likely to appeal to teachers, as it allows them to give attention to a number of students at a time. Because there is no restriction as to the number of participants in bingo, the game could conceivably be extended to include the entire class. Thus, a number of students can benefit from the teacher's time. Moreover, because much instruction occurs via peer modeling and feedback, the degree of teacher input necessary is lessened.

The training procedure does not require an

extensive knowledge of behavioral principles, so it can be learned with relative ease in a short period of time. The procedure is simple and could be efficiently managed by peers. The teacher could assign an advanced student to implement the training procedure and save even more time. Because an advanced student would have the majority of the sight words in her or his repertoire, little or no tutor training would be required.

The implication of the present investigation is that word game bingo provides an efficient training procedure for the acquisition and retention of sight words. The procedure is effective, easily implemented, and relatively inexpensive. Furthermore, the game proved very enjoyable for both the children and the trainer. It is possible that the provision of a positive environment for the introduction of reading skills could facilitate later development and maintenance of reading skills.

Comparison of the game to other training procedures along the dimensions of effectiveness, ease of implementation, expense, and long-term benefits would be interesting areas for future investigation. Extension of the game to other reading skills is a viable possibility and as earlier mentioned, was suggested some time ago (Thompson, 1973). Though the variations of the game remain untested, the implications of the present results are that such evaluation would prove fruitful.

REFERENCES

- Barrish, H. H., Saunders, M., & Wolf, M. M. Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 1969, 2, 119-124.
- Cautela, J. R. *Behavior analysis forms for clinical intervention*. Champaign, Ill.: Research Press, 1977.
- Cordps, A. D. *Phonics for the reading teacher*. New York: Holt, Rinehart and Winston, 1965.
- Corey, J. R., & Shamow, J. The effects of fading on the acquisition and retention of oral reading. *Journal of Applied Behavior Analysis*, 1972, 5, 311-315.
- Harris, V. W., & Sherman, J. A. Use and analysis of the "Good Behavior Game" to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis*, 1973, 6, 1-14.
- Hersen, M., & Barlow, D. H. *Single-case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press, 1976.
- Lahey, B. B., & Drabman, R. S. Facilitation of the acquisition and retention of sight-word vocabulary through token reinforcement. *Journal of Applied Behavior Analysis*, 1974, 7, 307-312.
- Lutzker, J. R., & White-Blackburn, G. The good productivity game: Increasing work performance in a rehabilitation setting. *Journal of Applied Behavior Analysis*, 1979, 12, 488.
- Maloney, K. B., & Hopkins, B. L. The modification of sentence structure and its relationship to subjective judgments of creativity in writing. *Journal of Applied Behavior Analysis*, 1973, 6, 425-433.
- Medland, M. B., & Stachnik, T. J. Good behavior game: A replication and systematic analysis. *Journal of Applied Behavior Analysis*, 1972, 5, 45-51.
- Neef, N. A., Iwata, B.A., & Page, T. J. The effects of known item interspersal on acquisition and retention of spelling and sightreading words. *Journal of Applied Behavior Analysis*, 1977, 10, 738.
- Shankman, F. *Games and activities to reinforce reading skills*. New York: MSS Information Corporation, 1972.
- Staats, A. W., & Butterfield, W. H. Treatment of non-reading in a culturally deprived juvenile delinquent: An application of reinforcement principles. *Child Development*, 1965, 3, 925-942.
- Swain, J. J., Allard, G. B., & Holborn, S. W. The good toothbrushing game: A school-based dental health program for increasing the toothbrushing effectiveness of children. *Journal of Applied Behavior Analysis*, in press.
- Thompson, R. A. *Energizers for reading instruction*. West Nyack, N.Y.: Parker Publishing Company, 1973.
- Van Houten, R., Hill, S., & Parsons, M. An analysis of a performance feedback system: The effects of timing and feedback, public posting, and praise upon academic performance and peer interaction. *Journal of Applied Behavior Analysis*, 1975, 8, 449-457.
- Whitlock, C., & Bushell, D. Some effects of "back up" reinforcers on reading behavior. *Journal of Experimental Child Psychology*, 1967, 5, 50-57.

Received August 26, 1980

Final acceptance February 6, 1981