JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1996, **29,** 451–469 NUMBER 4 (WINTER 1996)

TEACHING READING AND SPELLING: EXCLUSION AND STIMULUS EQUIVALENCE

JULIO C. DE ROSE AND DEISY G. DE SOUZA

UNIVERSIDADE FEDERAL DE SÃO CARLOS

AND

ELENICE S. HANNA

UNIVERSIDADE DE BRASÍLIA

In Experiment 1, 7 nonreading children were exposed to a program designed to teach reading of 51 training words. The program featured an exclusion-based procedure in which the children (a) matched printed to dictated words and (b) constructed (copied) printed words with movable letters and named them. All children learned to read the training words. Five children also read generalization words and showed progress in spelling. Experiment 2 applied the program to 4 different children, omitting the word-construction task. They also learned to read the training words, but only 1 participant read generalization words. The data support a stimulus equivalence account of reading acquisition and suggest that reading generalization may be obtained, especially when the teaching program includes word construction.

DESCRIPTORS: reading, spelling, stimulus equivalence, recombination of textual units, exclusion, matching to sample, children

Equivalence relations, defined by the properties of symmetry, transitivity, and reflexivity, provide behavioral criteria for the definition and assessment of meaning and comprehension (Sidman, 1986, 1994; Sidman & Tailby, 1982). For instance, if a child learns to select a pictorial representation of one fifth (as a comparison stimulus) in the presence of the printed ratio 1/5 (as a sample), this conditional relation is symmetrical if the child also selects 1/5 in the presence of the pictorial representation, without direct training. If the child also learns to select the printed decimal 0.2 (as a comparison stimulus) in the presence of the pictorial representation, transitivity of the trained conditional relations is demonstrated by the child selecting the decimal 0.2 in the presence of the ratio 1/5 without direct training. Reflexivity is demonstrated if the child selects an identical comparison stimulus in the presence of either 1/5, 0.2, or the pictorial representation, without direct training. Lynch and Cuvo (1995) taught children to select pictorial representations of fractions in the presence of fraction ratios and to select

Dedicated, in memoriam, to Fred S. Keller, who taught all of us. This research was supported by Grants 500619/90-5 and 500679/90-8 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brazilian Research Council) and Grant 90/3435-4 from Fundação de Amparo à Pesquisa do Estado de São Paulo. Portions of this research were presented at the annual convention of Sociedade Brasileira de Psicologia, Ribeirão Preto, Brazil, October 1992, and at the annual convention of the Association for Behavior Analysis, Chicago, May 1993. We thank Mariza Borges, Monica Fonseca, Laura Gomes, Lucia Sallorenzo, Gladys Duarte, Ana Lucia Rossito Aiello, Adriana Pereira, Naldeli Fontes, Lidia Balduino, Gustavo Carvalho, Rosangela Cesaretti, Adriana C. Bernardes, and M. Angelica Zanotto for help in conducting the research. We also thank Maria Amelia Matos, Gina Green, and William Dube for their many valuable comments on earlier versions of this paper, and José Gonçalves Medeiros for suggestions about data presentation.

Correspondence concerning this article should be addressed to Julio C. de Rose or Deisy G. de Souza, Departamento de Psicologia, Universidade Federal de São Carlos, Caixa Postal 676, 13565-905, São Carlos, S.P., Brazil (E-mail: djcc@power.ufscar.br to J. C. de Rose or ddgs@power.ufscar.br to D. G. de Souza).

printed decimals in the presence of pictorial representations, and showed that these relations were symmetrical and transitive, indicating that each fraction ratio, its pictorial representation, and the corresponding decimal formed an equivalence class. For these children, therefore, the pictorial representation became the meaning of the symbols 1/5 and 0.2.

The stimulus equivalence paradigm provides economical and effective methods to teach complex repertoires. For instance, Sidman (1971) worked with a youth with severe mental retardation who began the experiment able to (a) match 20 pictures presented as comparison stimuli to their corresponding dictated words presented as sample stimuli and (b) name the 20 pictures. The participant then was taught to match 20 printed-word comparisons to the same set of dictated samples. Afterwards, the participant could match picture comparisons to printed-word samples and printed-word comparisons to picture samples, without direct training. The participant could also name (read) the printed words. The emergent matching performances demonstrated in this study verified that the dictated words, printed words, and pictures were related by equivalence (Sidman, 1994). Sidman concluded that the formation of equivalence relations enabled the youth to read the 20 words with comprehension.

Mackay and Sidman (1984) and Mackay (1985) found that equivalence classes involving printed words could be established by teaching subjects to construct the printed words with movable letters. For instance, in one of their studies, Mackay and Sidman (1984) taught students who were mentally retarded to name color patches and construct the color names using movable letters (constructed-response spelling). Students subsequently matched printed-word comparisons to color patches and color comparisons to printed-word samples, and also named printed words. Teaching students to construct printed color names, therefore, established equivalence classes that consisted of dictated color names, printed color names, and colors. These findings suggest that equivalence relations may integrate reading and spelling performances. A recent study by Stromer and Mackay (1992) supports this contention. They taught students with mental retardation to construct three different printed names in the presence of a picture, and showed that the picture and three printed words formed an equivalence class.

Although the applied potential of stimulus equivalence research has been noted, the actual use of stimulus equivalence in applied settings has been very limited (see Mace, 1994; Sidman, 1994). This is also true for reading and spelling instruction, in spite of major advances in research on basic processes of equivalence class formation and methods to teach prerequisite performances involving printed and dictated words (e.g., Dube, McDonald, McIlvane, & Mackay, 1991; Matos & D'Oliveira, 1992; Stromer & Mackay, 1992).

An applied study by de Rose, de Souza, Rossito, and de Rose (1992) investigated the acquisition of reading performance closer to that expected in the classroom, with elementary school children who had a history of failure in learning to read. Exclusion procedures (Dixon, 1977; McIlvane & Stoddard, 1981) were used to increase the children's repertoires of conditional relations between printed-word comparisons and the corresponding dictated words. Exclusion procedures minimize errors and are more effective than trial-and-error procedures for generating auditory-visual matching as well as naming of visual stimuli (Ferrari, de Rose, & McIlvane, 1993). On exclusion trials, an undefined sample (not yet related to a particular comparison) is presented with an undefined comparison and other comparisons

already defined. The participant may then "exclude" the defined comparisons, selecting the undefined one and often learning the relation between this stimulus and the sample. De Rose et al. (1992) presented novel (untrained) words as samples on exclusion trials, together with two comparison stimuli: the corresponding untrained printed word and a baseline word that had been previously taught. For instance, the printed words *vaca* (cow) and *bolo* (cake) were presented as comparison stimuli, and the untrained word "vaca" was dictated. The children could then exclude the printed word *bolo,* which they had already learned to relate to the corresponding dictated word, and correctly select the printed word *vaca.* A constructed-response task required children to copy these new words with movable letters. Probes inserted periodically during training assessed whether children named (read) the printed words to which they had previously responded in the exclusion matching procedures. Generalization probes tested naming of novel words, formed by recombination of syllables of the training words (such as the word *boca,* formed by a syllable of *bolo* and a syllable of *vaca*). De Rose et al. (1992) replicated the results of Sidman (1971) and Sidman and Cresson (1973), showing that subjects learned to name the printed words used in the training. Further, as the repertoire of trained and emergent relations increased, most subjects began to read novel words, suggesting acquisition of generalized reading performances.

The present study replicates and extends our previous findings. It was conducted with Brazilian children from elementary school who had not made progress in reading and spelling. They had, however, acquired an extensive copying repertoire. This repertoire was much less than expected from first graders in Brazil, who are required to read, spell, take dictation, and write meaningful sentences or phrases. A large number of students, especially those of low socioeconomic backgrounds, fail at this level and are required to repeat the initial grades several times. This massive failure is generally attributed to inadequate teaching conditions (e.g., Brandão, Baeta, & Rocha, 1983; Fletcher & Costa-Ribeiro, 1987).

In Experiment 1, we applied the same teaching program used in our earlier study (de Rose et al., 1992), periodically assessing the reading and spelling of trained and untrained words. Experiment 2 evaluated the extent to which the constructed-response task may have contributed to reading and spelling generalization.

EXPERIMENT 1

Method

Participants

Participants were 7 first-grade students (3 girls and 4 boys, mean age = 8.3 years; range, 7.5 to 10.8 years) of low socioeconomic status, who had been in school for at least 8 months. Students were referred by their teachers because they had not learned to read and spell simple words. Paulo, Elena, Elia, and Ari lived in an orphanage and spent half a day in an elementary public school. Dora, Rai, and Diego lived with their families and attended a public school. IQs were assessed before the beginning of the study, using the Wechsler Intelligence Scale for Children (Wechsler, 1949). Students' mean scores were 82 (range, 75 to 87).

Before the beginning of the study, the children were asked to read and write simple words. No feedback was given for correct or incorrect responses. Words used in the reading assessment were different from those used in the teaching program, but, like those words, were formed by simple syllables (a consonant followed by a vowel, with each consonant corresponding to only one speech

sound) and were chosen from words presumed to be highly familiar to the children. The reading assessment was discontinued after a child made five consecutive errors. Elia read one word correctly and the other students read none of the words. For the writing assessment, students received pencil and paper and were asked to write a series of five words and one short sentence. The words varied between one and four syllables: *pão* (bread), *mesa* (table), *banana* (banana), and *telefone* (telephone). The experimenter showed the students a picture of each object, and asked them to name it and then write the corresponding word. The final picture showed a boy reading a book. Children were asked to tell what the picture showed. All children described it with a short sentence like "O menino está lendo" (The boy is reading). The child then wrote the sentence just produced. If the students stated that they did not know how to write a word or sentence, they were gently encouraged to write "in the way they knew" (Ferreiro & Teberosky, 1982). All students presented a writing pattern described as "presyllabic" (Ferreiro & Teberosky, 1982); they wrote sequences of characters (letters or scribbles that resembled letters) with approximately the same number of characters regardless of the length of the dictated word or phrase. When the characters written resembled letters, they did not correspond to the actual sounds in the words or the sentence. In a final assessment, a series of printed and handwritten words were presented, and participants were asked to copy each word. All students copied the printed and handwritten words.

Setting and Materials

While the study was in progress, students were attending public elementary school, where they were exposed to regular reading and spelling instruction. A teaching program was organized in a series of units. The program used common words in Portuguese (the students' native language) that could be easily represented by pictures. In general, the words had two or three syllables, of the consonant-vowel type. The few exceptions were some words that contained a one-vowel syllable, for example, *diabo* (devil) and *navio* (ship). The selection of the whole set of words also ensured that each consonant corresponded to just one phoneme.

The visual stimuli were printed words and pictures. Words were printed in lower case 65-point Arial type. A drawing in black ink, approximately 5 cm by 5 cm, was made to represent each picture. Printed words and pictures were photocopied and glued on sheets of letter paper. Each sheet displayed stimuli for one trial, was enclosed in a transparent plastic envelope, and was placed in a binder. The binder was placed on a desk so that the sheets were in front of the participant, and the experimenter could turn the sheets to present successive trials. The child sat facing the binder, and the experimenter sat beside the child. One or more observers sat behind the child and the experimenter and recorded the child's responses as well as other relevant events in the session.

Paulo, Elena, Elia, and Ari had daily sessions in a laboratory at the university, 5 days per week, for approximately 6 months (range, 4 to 7 months). Dora, Rai, and Diego had sessions 3 days per week in a small office at their school for an average of 8 months (range, 6 to 11 months). The actual interval between sessions was sometimes longer than scheduled because students missed some school days. Average session duration was approximately 20 min, and no session was longer than 40 min.

Procedures

Overview of the teaching program. The program was designed to teach students to read a set of 51 training words, using two teaching activities: matching printed-word

comparisons to dictated-word samples, and a constructed-response task in which students used movable letters to copy a printed word. After the establishment of an initial baseline of matching and reading three words, an exclusion procedure was used to establish new matching relations. Tests conducted periodically verified whether students could read the words presented in exclusion trials (training words) and other words produced by recombination of the phonemes of the training words (generalization words). Spelling of training and generalization words was also probed.

The program was comprised of 25 units, each applied in an individual session and repeated in successive sessions until a criterion was achieved. Unit 1 established the initial baseline of matching and reading with three words. Unit 2 verified the formation of equivalence classes consisting of dictated word, printed word, and picture for the three training words of the first unit. Subsequent units were exclusion units designed to teach new matching relations between dictated and printed words and to obtain emergent reading of these new sets of words, and equivalence units to verify the formation of equivalence classes involving the new words. After every two exclusion units, an equivalence unit was conducted to verify class formation. Two extensive reading tests were interspersed in the middle and at the end of the program (Units 12 and 25).

Correct responses of matching, constructing words with movable letters, and reading were always followed by social reinforcement (confirmation and praise) and, for Paulo, Elena, Elia, and Ari, also by tokens that could be exchanged for items available on a shelf at the end of the session. Consequences for incorrect responses differed depending on the nature of the response. No feedback was given for spelling responses. Before spelling assessments, students were told that they no longer would be informed whether responses were correct.

Table 1
Function of Each Unit in the Teaching Program, Number of Training and Generalization Words Presented in Each Unit, and Number of Trials with Each Word

Unit	Function	Words		Number of trials[a]
		Train	Gen	
1	Teaching: baseline	3		10
2	Equivalence test	3		4W-4P
3	Teaching: exclusion	2	2	5E-5C
4	Teaching: exclusion	2	2	5E-5C
5	Equivalence test	7	4	2W-2P
6	Teaching: exclusion	2	2	5E-5C
7	Teaching: exclusion	3	2	5E-5C
8	Equivalence test	12	8	1W-1P
9	Teaching: exclusion	3	2	5E-5C
10	Teaching: exclusion	4	2	5E-3C
11	Equivalence test	19	14	1W-1P
12	Extensive test	19	14	1
13	Teaching: exclusion	4	2	4E-4C
14	Teaching: exclusion	4	2	4E-4C
15	Equivalence test	27	18	1W-1P
16	Teaching: exclusion	4	2	4E-4C
17	Teaching: exclusion	4	2	4E-4C
18	Equivalence test	35	22	1W-1P
19	Teaching: exclusion	4	2	4E-4C
20	Teaching: exclusion	4	2	4E-4C
21	Equivalence test	43	26	1W-1P
22	Teaching: exclusion	4	2	4E-4C
23	Teaching: exclusion	4	2	4E-4C
24	Equivalence test	51	30	1W-1P
25	Extensive test	51	45	1

Note: Equivalence units presented all training and generalization words from previous teaching units. Extensive tests in Units 12 and 25 presented all training words from previous units, plus a set of generalization words not included in any other teaching unit (the 45 generalization words in Unit 25 include the 14 already presented in Unit 12).

[a] For exclusion units, the numbers are for exclusion trials (E) and control trials (C) with each training word. For equivalence units, the numbers are for trials with each printed word as sample (W) and with each picture as sample (P). These numbers do not include pretest and posttest trials, maintenance trials, and matching trials to teach and test picture naming.

Table 1 summarizes the teaching program. The table shows the function of each unit (exclusion teaching, equivalence test, etc.), the number of training and generalization words introduced in each unit, and the number of specific trial types in each unit.

Teaching children to name pictures. Children named common pictures; however,

some training words were difficult to represent unambiguously by pictures. For example, to represent a beak it was necessary to draw a bird's face. Thus, students could name this picture "bird" as well as "beak." In other cases, students could produce common synonyms. It was necessary, therefore, to teach participants to name pictures with the specific words intended in the study.

Training was conducted first by a match-to-sample format, followed by a naming assessment. For a given set of words, a block of matching trials was conducted, with each word as a sample in one trial. Three pictures were presented as comparison stimuli in each trial, and selections of the picture corresponding to the dictated word were reinforced. If the student made an incorrect selection, the experimenter pointed to the correct comparison and said, "This is the picture of —. Point to the picture of —." Correct responses following this prompt were reinforced. The block was repeated until the student made correct unprompted selections in all trials. Then the pictures were presented one at a time in successive trials, and the participant was asked to name each picture. If the student produced any name different from the name assigned to the particular picture in the experiment, the block of matching trials was repeated to criterion and was followed again by the block of naming trials. This cycle of matching trials followed by naming trials was repeated until the student named all pictures with the names assigned to them in the experiment.

For naming training, pictures were distributed in sets. Each set contained the training and generalization words of two consecutive exclusion units (which were always followed by an equivalence unit). For each pair of exclusion units, naming training was conducted at the beginning of the initial unit. There was also another set of three pictures, corresponding to the training words of Unit 1. Naming training for these pictures was conducted at the beginning of this unit.

Teaching the initial baseline. The first unit taught an initial baseline of matching and reading three words. After participants named all three pictures, training of matching printed words to dictated words began. On the six initial trials (two with each training word), only one printed word was presented. The experimenter said, "This word is —. Point to the word —." The child pointed to the word and this response was reinforced. These trials were followed by 30 matching trials in which the sample was a dictated word (each training word was presented 10 times as the sample), and comparison stimuli were two printed training words. Matching trials began with the experimenter turning the page and presenting the two comparison stimuli. The experimenter waited 2 to 3 s after presentation of the comparison stimuli and then dictated the sample word. In the beginning of the program, the sample word was often presented in a phrase (e.g., "Which of these words is —" or "Point to the word —."). Correct responses were followed by programmed consequences. Incorrect responses were followed by the verbal prompt, "Are you sure?" The student could then point to the other word. After this correction procedure, the trial was repeated. The unit ended with a reading (oral naming of printed words) posttest. Each posttest trial presented only one printed word, centered in the lower half of the page, and the experimenter asked, "Which word is this?" There were six posttest trials, two with each training word. Correct responses were reinforced. If the student made an incorrect response or stated that he or she did not know how to read the word, the page was turned and the next trial was initiated. If one or more incorrect responses occurred in the posttest, in the next session the block of matching trials was repeated and was followed again by the posttest. This

cycle of matching trials followed by reading trials was repeated until the student read the three words on the posttest.

Exclusion units. Each exclusion unit was designed to enable the student to read two to four printed words (the training words for the particular unit). A pretest and a posttest verified reading training and generalization words, the latter produced by recombining syllables and phonemes of training words. Each unit was repeated until the student read all training words in the posttest as well as in a maintenance assessment conducted at the beginning of the next session. These words were then added to a cumulative baseline comprising all words that the student had already learned to match and read (henceforth referred to as baseline words).

The pretest was conducted in a matching-to-sample context. The sample was a dictated baseline word, and the comparison stimuli were the corresponding printed word and either a training or generalization word. After the student selected the baseline printed word and reinforcement was provided, he or she was asked to read the other word. Correct responses were reinforced. If the student made an incorrect response or stated that he or she did not know how to read the word, the page was turned to initiate the next trial. This format for the pretest was adopted because it was expected that students would make reading errors with most training and generalization words. Because participants virtually always selected baseline printed words correctly in the matching-to-sample task, a high rate of reinforcement would be assured.

Training by exclusion began after the pretest, with three types of matching-to-sample trials. The comparison stimuli on exclusion trials were two printed words, one of them a baseline word and the other a training word. The training word was dictated as the sample, and the student could respond correctly by excluding the baseline word. Control trials, mixed with exclusion trials, also presented a training printed word together with a baseline word as comparison stimuli. The sample in control trials was the dictated baseline word. Mixed with these trials in the first three exclusion units were 10 baseline trials on which the two comparison stimuli were words that had been previously taught. Baseline trials were included to ensure high rates of reinforcement for correct responding within a session and maintenance of the baseline repertoire. All trials were conducted as programmed beforehand, even if the student already read one or more of the training or generalization words in the pretest.

The first two exclusion trials with a new word also required a constructed-response matching to sample. The student was asked to copy the word by sequencing a set of randomly presented letter tiles. After the printed word in the matching task was selected correctly, the set of letters that formed the word to be constructed was displayed. The printed word remained present while the child selected and sequenced the letters. A correctly constructed response was followed by confirmation and praise. The participant was asked to name the word just constructed and was reinforced for correct reading. If the constructed response was incorrect, the tiles were mixed again, and the participant was asked to repeat the task. After correct constructed response and naming, the next matching-to-sample trial was presented.

A posttest verified reading of training and generalization words for that session, plus a sample of training words of preceding units. Each trial presented only one printed word, centered at the lower half of the page, for the participant to read. Words were presented in a randomized order, and correct reading responses were reinforced. If the student did not know how to read the word or made an error, the page was turned to present the next trial.

Criterion for advancement to the next

unit was reading all training words correctly on the posttest and on a maintenance test conducted at the beginning of the next session. Otherwise, the training unit was repeated, with the pretest omitted.

Equivalence units. Equivalence units assessed matching of pictures to printed words and vice versa. Each trial presented a picture or printed word as a sample, centered in the upper half of the page, and three pictures or printed words as comparison stimuli, in the lower half of the page. The child was asked to name the sample and then select the corresponding comparison stimulus. Correct naming and selections were reinforced. If the student made a naming error, the experimenter turned the page before the student made any selection and initiated the next trial. Responding on a trial was scored as incorrect if an error occurred either on the naming or on the selection response.

The stimuli used in equivalence units were drawn from the set of all training and generalization words from preceding exclusion units. This set became larger at each new equivalence unit. For each word there was at least one trial in which the printed word appeared as the sample (with three pictures as comparison stimuli) and one trial in which the picture appeared as the sample (with three printed words as comparison stimuli). Table 1 shows the number of words tested on equivalence units and the number of trials on which each word and each picture appeared as the sample. For example, Unit 5 assessed equivalence by asking whether the student would match the seven pictures and seven printed words involved in direct training up to that point. In addition, equivalence was assessed with four generalization words (two tested for reading on Unit 3 and two on Unit 4). The total number of trials in Unit 5 was then 44 (2W means that each printed word appeared as a sample on two trials; 2P means that each picture also appeared as a sample on two trials). Trials were arranged in a randomized order.

Spelling tests. Spelling probes were conducted at the end of equivalence units, from Unit 5 onward. Participants were given paper and a pencil and asked to spell each word to dictation. The experimenter dictated a sample of training and generalization words from the exclusion units that the student had already completed. The number of dictated words varied among subjects and usually increased gradually along successive tests, ranging from 5 to 28. Training and generalization words were presented in a mixed order in each test. The words changed along successive probes. The experimenter did not provide differential feedback for spelling responses. Correctness of spelling responses was evaluated by at least two judges, by inspection of the child's written product. In cases in which it was difficult to decide on the basis of the written material, judges watched the videotaped records that focused on the movements of the child's hand and showed the handwriting while the word was being written.

Extensive tests for reading generalization. The reading performance was also assessed in two extensive tests conducted at the middle and the end of the program (Units 12 and 25), with all baseline words and a set of generalization words that did not appear in other units of the program. These reading probes were conducted in the matching-to-sample format used in reading pretests described before. The number of words tested was 19 baseline and 14 generalization words in Unit 12 and 51 training and 45 generalization words in Unit 25.

Remedial procedures. Remedial procedures were used when a child repeatedly failed to achieve criterion in a particular unit and when a child made repeated mistakes reading a particular baseline word. If a student repeatedly failed to achieve criterion in an exclusion unit, the initial procedure was to

reduce the number of training words in the unit and to reinstate the original training words progressively. The number of training words was reduced by withdrawal of all except one of the training words not read in the last posttest. Hence, the modified unit presented only the training words read successfully in the last posttest plus one training word not read in this posttest. After criterion was achieved with this modified unit, each succeeding session reintroduced one of the removed training words.

A different remedial procedure was used when a student made repeated errors reading a baseline word (a training word of an earlier unit). At the beginning of the next session, the word on which the student made repeated mistakes was presented on a flash card, side by side with another baseline word. These cards were displayed as comparison stimuli in six matching-to-sample trials. Each word was dictated as a sample in three trials (in a randomized order), and then the scheduled unit was conducted.

Interobserver agreement. A sample of 136 sessions, of the total of 327 sessions conducted, was used to assess interobserver agreement. Two independent observers recorded selections on matching trials, naming (pictures or printed words), constructed responses, and handwriting upon dictation. Each of these responses was considered as a unit and scored as an agreement or disagreement. Interobserver agreement was obtained using a point-by-point formula (Kazdin, 1982): The total number of agreements was divided by agreements plus disagreements and multiplied by 100%. The mean agreement score was 96.8% (range, 75.0% to 100%). Disagreements usually consisted of omissions in recording subjects' selections on matching trials, or different interpretations of subjects' handwriting and their utterances on naming trials. In most cases, watching the videotape of the session made clear what response to consider in data analysis.

Table 2
Accuracy Scores on Matching Printed Words to Pictures and Pictures to Printed Words Averaged Across Eight Equivalence Units of Experiment 1

Participants	Printed words to pictures		Pictures to printed words	
	%	Range	%	Range
Training words				
Paulo	97.8	89.5–100	99.3	94.7–100
Elena	99.5	96.3–100	96.4	83.3–100
Dora	99.1	91.6–100	95.6	83.3–100
Elia	99.0	91.7–100	100	
Rai	95.4	90.0–100	95.9	86.0–100
Diego	89.4	76.4–100	89.0	74.0–100
Ari	98.1	92.9–100	98.8	92.9–100
Generalization words				
Paulo	100		94.0	75.0–100
Elena	90.8	75.0–100	99.4	96.0–100
Dora	80.3	66.6–96.0	70.3	25.0–92.8
Elia	96.3	87.5–100	98.2	87.5–100
Rai	71.4	50.0–92.8	70.1	50.0–94.1
Diego	66.0	50.0–92.8	64.8	37.5–82.1
Ari	80.2	50.0–93.8	77.4	62.5–100

Results

Accuracy on training tasks was very high for all students, and the number of errors in all training trials was negligible. Table 2 shows accuracy for each participant on trials that assessed formation of equivalence classes (matching printed-word comparisons to picture samples and picture comparisons to printed-word samples). The scores are averages for each participant across the eight equivalence units for trials with training and generalization words. All children showed high accuracy scores with training words. Paulo, Elena, and Elia showed similar scores with generalization words. The other subjects showed lower and more variable scores for generalization words, but most of them achieved scores close to 100% in the final units.

Figure 1 shows scores on reading pretests and reading posttests that were conducted in exclusion units. Scores for Unit 1, in which the first three words were taught, are also included in Figure 1. This unit is indicated

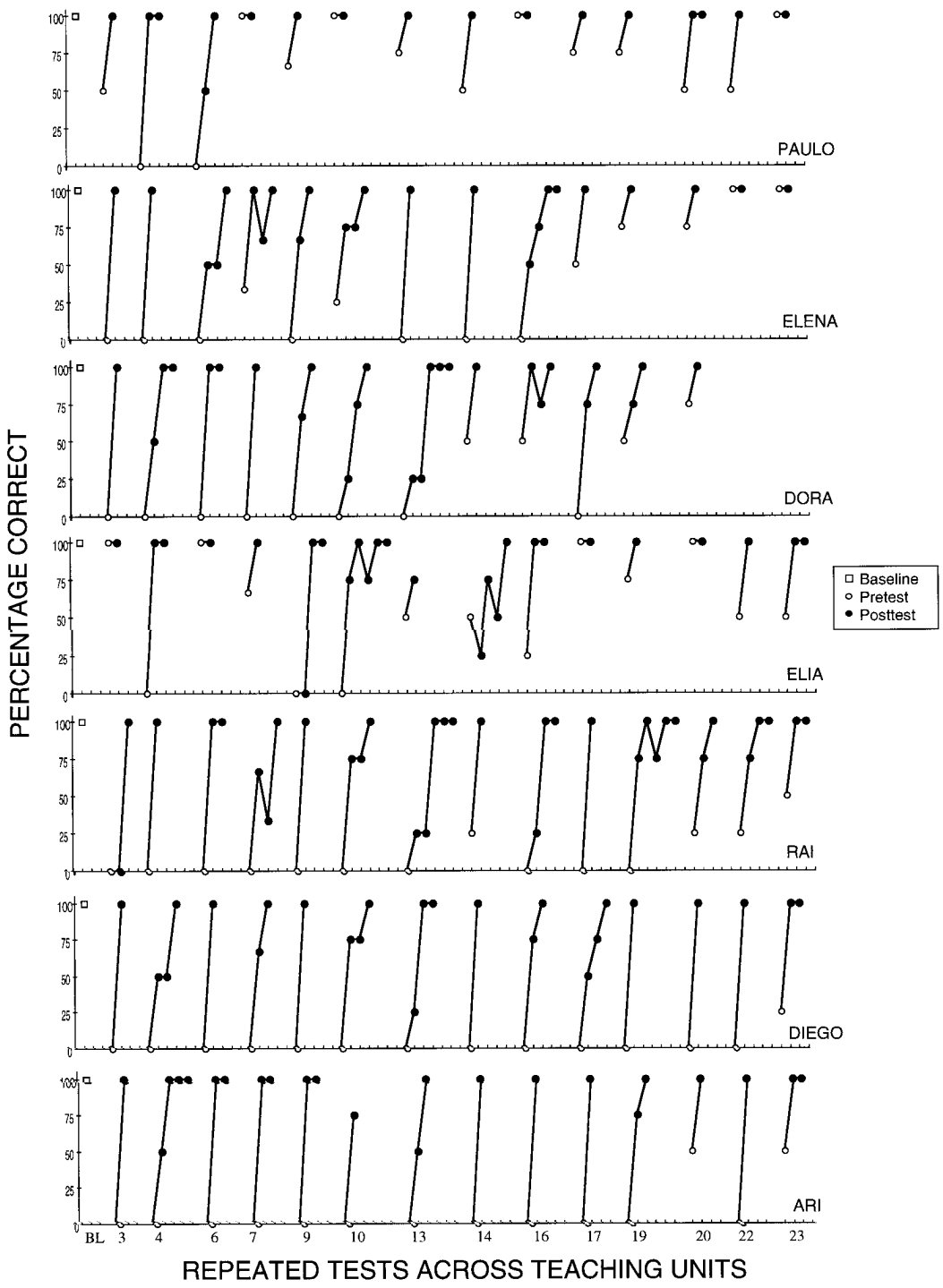

Figure 1. Reading scores (percentage correct) for training words on the posttest for Unit 1 (baseline training) and pretests and posttests of subsequent exclusion units of Experiment 1. The lines connect data points within each unit. The numbers on the abscissa indicate the corresponding exclusion units (see Table 1).

Table 3
Reading Scores for Training and Generalization Words on Units 12 and 25 (Extensive Tests) of Experiment 1

	Training words				Generalization words			
	Unit 12	Unit 25			Unit 12	Unit 25		
Students	$n = 19$	Early $n = 19$	Late $n = 32$	Overall $n = 51$	$n = 14$	Early $n = 14$	Late $n = 31$	Overall $n = 45$
Paulo	100	100	100	100	71.4	85.7	80.6	82.2
Elena	89.4	100	100	100	14.3	92.9	83.9	86.7
Dora	84.2	100	93.8	96.0	0.0	42.9	74.2	64.4
Elia	78.9	100	87.5	92.2	35.7	42.9	22.6	28.9
Rai	89.4	94.7	87.5	90.2	14.3	14.3	22.6	20.0
Diego	89.4	94.7	71.9	80.4	0.0	7.1	0.0	2.2
Ari	100	94.7	93.8	94.1	0.0	0.0	0.0	0.0

Note. Scores for Unit 25 are presented separately for training words taught before Unit 12 (early), after (late), and overall; for generalization words, scores on Unit 25 are presented separately for words already tested on Unit 12 (early), words presented only on Unit 25 (late), and overall.

by BL (baseline), and the open squares represent scores on posttests conducted at the end of that unit. All participants required only one session to reach 100% correct in reading posttests for Unit 1. In subsequent units, all children showed low scores on pretests followed by marked increases on posttests, reaching 100% correct after a few repetitions of each unit. There was individual variability in the number of repetitions of each unit and in the number of units that required repetition. Dora and Rai also required the use of remedial procedures. This happened on Unit 10 for Dora and on Units 7, 13, and 19 for Rai. Under some conditions, students reached the criterion on the posttest, but the unit was repeated because the retention test conducted at the beginning of the following session was not passed.

Although initial pretest scores were low, all children eventually showed an increase in these scores. Paulo showed this pattern of increasing pretest scores relatively early, whereas for the remaining students, pretest scores began to increase during the second half of the program. Elia showed a slightly different pattern, oscillating between high and low scores on pretests for the initial units, falling back to low pretest scores midway through the program, and showing a gradual increase toward the end. All 7 students finished the program reading all words on the last posttest of each unit. (Elia and Ari were mistakenly moved to the next unit with posttest scores of 75% on Units 10 and 13, respectively, but they both passed the retention test for those units.)

Reading performances were maintained during the two extensive tests conducted midway through (Unit 12) and at the end of the program (Unit 25). Table 3 presents reading scores for training and generalization words on both tests. For training words, each test included all words trained until that point. For Unit 25, the table presents a separate score for words trained before Unit 12 (labeled "early"). These were the same words included in the reading assessment at Unit 12. Another score, labeled "late," corresponds to reading scores with words trained after Unit 12. An overall score for Unit 25 is also presented. Generalization tests on Units 12 and 25 were conducted with a set of words that had not been included in any other training unit. Scores for Unit 25 also distinguish early and late generalization words (early words were presented in the test at Unit 12). All children achieved high accuracy scores for training words, and their scores tended to increase

Table 4
Spelling Scores for Training and Generalization Words on Units 11 and 24 of Experiment 1

	Training words		Generalization words	
Participants	Unit 11	Unit 24	Unit 11	Unit 24
Paulo	4/7 (57.1)	4/8 (50.0)	4/16 (25.0)	2/4 (50.0)
Elena	1/7 (14.3)	5/8 (62.5)	1/15 (6.6)	3/4 (75.0)
Dora	0/14 (0.0)	6/11 (54.5)	0/8 (0.0)	6/10 (60.0)
Elia	0/7 (0.0)	3/8 (37.5)	1/16 (6.2)	1/4 (25.0)
Rai	1/5 (20.0)	3/10 (30.0)	0/1 (0.0)	2/7 (28.5)
Diego	0/5 (0.0)	0/14 (0.0)	0/3 (0.0)	0/8 (0.0)
Ari	0/7 (0.0)	0/8 (0.0)	0/15 (0.0)	0/4 (0.0)

Note. The number of dictated words varied across children. Scores present the number of words spelled correctly on each test and the number of dictated words (correct/total), followed by percentage correct in parentheses.

from Unit 12 to Unit 25 (except Diego and Ari). Students with scores lower than 100% on Unit 12 increased their scores for early words on Unit 25. The median number of training words read on Unit 12 was 17 (range, 15 to 19). On Unit 25, the median number of training words read was 48 (range, 41 to 51). Accuracy scores remained high at the end of the program, even though the number of training words had increased. Maintenance of reading was similar for words learned in the initial section of the program (early) and words learned in the final section (late).

In general, participants also demonstrated some degree of generalized reading (they named new printed words involving new combinations of phonemes used in the training words), but the unit of the program in which generalization began to occur and the final score for reading generalization words varied across students. These tests provided the most stringent measure of generalization, because generalization words used therein did not appear in any other units of the program. On Unit 12, the median number of generalization words read by the students was 2 (range, 0 to 10), with 3 students showing no generalization at all. On Unit 25, the median number of generalization words read was 13 (range, 0 to 39), and 5 of the 7 participants increased their reading scores for early generalization words. There were differences in scores for early and late words, but there was no systematic trend in the percentage of correct responses among students. Paulo, Elena, and Dora showed marked increases in generalization during the final part of the program. Paulo and Elena achieved accuracy levels comparable to those with the training words. Elia and Rai read approximately 20% to 30% of generalization words. Diego and Ari had the poorest performance with generalization words throughout the program. Ari did not read any generalization word correctly on these tests, and Diego read only one generalization word on Unit 25. These scores for generalization words on Units 12 and 25 corresponded closely to scores on generalization posttests conducted during exclusion units (not presented).

Five of 7 students showed improvement in spelling, but the improvement was variable across students. Although their spelling performances on the training and generalization words were never actually pretested, their spelling scores on Unit 11 surpassed those obtained during the initial assessment using a different set of words of comparable difficulty. Table 4 presents spelling scores measured at the end of Units 11 and 24 for training and generalization words. In general, there was an increase from Unit 11 to

Unit 24 in the percentage of words that were completely and correctly spelled, except for Paulo, whose scores with training words were higher in Unit 11. The results were similar for both training and generalization words, showing a systematic increase in spelling accuracy. Although the scores for spelling were somewhat lower than the scores for reading (see Table 3), students who showed higher scores in reading generalization words tended to show higher scores in spelling.

EXPERIMENT 2

In Experiment 1, most of the students learned to read not only the training words but the generalization words as well. In addition, these students scored highest on the spelling tests. Experiment 2 addressed the possibility that the constructed-response component of the procedure may have contributed to these generalized reading and spelling performances. Four new students were given the same reading program, but they were not required to construct the words used during exclusion training.

Method

Participants

Four first-grade students of low socioeconomic status (mean age = 9.1 years; range, 8.9 to 10.1 years) participated. Toni, Ilto, and Gil were boys and Elis was a girl. They all lived in an orphanage and spent half a day in an elementary public school. They were referred with students of Experiment 1 because they had not learned to read and spell simple words and had had their reading and writing repertoires and IQs assessed at the beginning of the school year (in March).

In the initial reading assessment, participants did not read any words. In the writing assessment, all showed the writing pattern described as presyllabic (see Experiment 1). Mean IQ was 70 (range, 64 to 74). Experiment 2 began 6 months after the beginning of Experiment 1 (in September). Between March and September, participants assigned to Experiment 2 periodically came to the university for reassessment of their writing repertoires and continued to attend regular school. During this time, they showed no apparent progress in writing and continued to show the presyllabic writing pattern. Reading pretests conducted during Experiment 2 also revealed no noticeable progress in reading during this time. Students continued to attend regular school while the experiment was in progress.

Setting, Materials, and Procedure

All aspects of the procedure were the same as in Experiment 1, except that the constructed-response task was omitted on exclusion units. The mean interobserver agreement score for a sample of 192 sessions, of the total of 248 sessions conducted, was 97.4 % (range, 70.3% to 100%).

Results

As in Experiment 1, the children's performances on training tasks were nearly errorless, and all of them showed very high accuracy on trials that assessed the formation of equivalence classes with training words. Table 5 shows individual scores for training and generalization words, averaged across the eight equivalence units. For training words, the average was close to 100%, both for matching printed words to pictures and vice versa. Except for Elis, accuracy was lower and more variable for generalization words. The lower scores usually occurred on the first two equivalence units.

Figure 2 shows reading scores for training words on exclusion units. Low pretest scores were followed by marked increases in posttest scores. Students often reached 100% correct on the first posttest or after a few repetitions of each unit. In most cases, repetition was due to failure on the retention

Table 5
Accuracy Scores on Matching Printed Words to Pictures and Pictures to Printed Words Averaged Across Eight Equivalence Units of Experiment 2

Participants	Printed words to pictures		Pictures to printed words	
	%	Range	%	Range
Training words				
Elis	100		99.1	92.9–100
Toni	97.5	89.5–100	96.8	91.1–100
Ilto	95.1	83.3–100	92.2	72.7–100
Gil	97.8	88.2–100	95.4	85.0–100
Generalization words				
Elis	97.7	93.8–100	96.8	87.5–100
Toni	80.4	60.0–100	75.7	37.5–100
Ilto	85.7	61.5–100	78.5	54.5–100
Gil	88.4	75.0–100	85.6	70.8–100

test (maintenance data are not shown, but failures are indicated by successive posttests with 100% correct). In a few cases (e.g., Toni, Unit 13, and Ilto, Unit 16) students showed a gradual increase on successive posttests, requiring several sessions to achieve the criterion to pass to the next unit. Elis showed a pattern different from the other 3 students that was very similar to the one showed by Paulo in Experiment 1. Despite her low reading scores in the assessments conducted before the experiment, Elis read one word on the pretest of Unit 3 and scored zero only on the pretest of Unit 4; after that, she always read some of the words on the pretests and, except on Unit 9, required only one session to achieve the criterion. Her pretest scores increased markedly

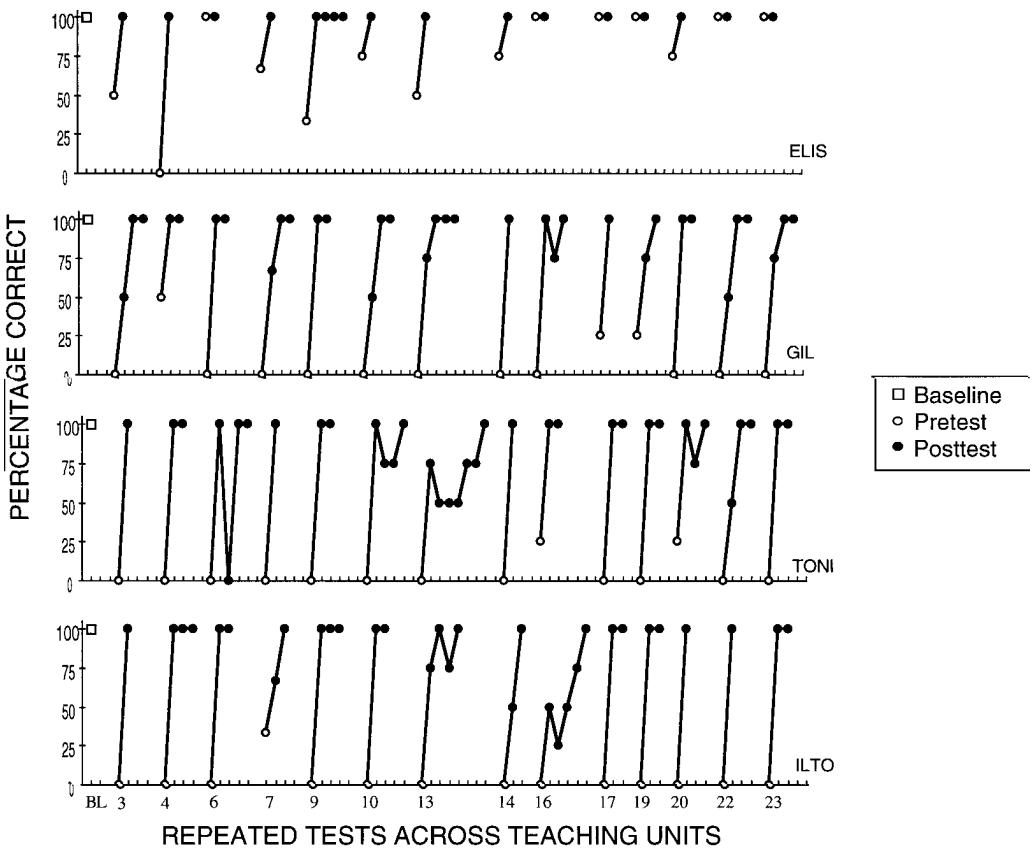

Figure 2. Reading scores (percentage correct) for training words on the posttest for Unit 1 (baseline training) and pretests and posttests of subsequent exclusion units of Experiment 2.

Table 6
Reading and Spelling Scores for Training and Generalization Words on Tests in the Middle and at the End of the Program for Participants of Experiment 2

Participants	Training words		Generalization words	
	Middle	End	Middle	End
Reading[a]				
Elis	100	100	35.7	91.1
Toni	73.7	100	0.0	0.0
Ilto	100	86.3	0.0	0.0
Gil	100	82.3	0.0	0.0
Spelling[b]				
Elis	85.7	62.5	50.0	50.0
Toni	14.7	12.5	0.0	0.0
Ilto	0.0	0.0	0.0	0.0
Gil	0.0	0.0	0.0	0.0

[a] Reading tests were conducted on Units 12 (middle) and 25 (end); *n* was 19 and 51, on Units 12 and 25, respectively, for training words, and 14 and 45 for generalization words.

[b] Spelling tests were conducted on Units 11 (middle) and 24 (end); for training words, *n* was 7 and 8 on Units 11 and 24, respectively; for generalization words, *n* was 4 on both units.

during the program, reaching 100% in the final units. The other 3 subjects showed low pretests on most teaching units, replicating the pattern observed for Rai, Diego, and Ari in Experiment 1.

Except for Elis, students in this experiment performed differently from those in Experiment 1 on probes for reading generalization and spelling. Table 6 shows number of training and generalization words read on Units 12 and 25, with the corresponding percentage scores. Table 6 also shows the number of training and generalization words spelled correctly on spelling probes conducted on Units 11 and 24. All children read training words with high accuracy in Units 12 and 25, but only Elis read generalization words, and her scores increased from Unit 12 to Unit 25. This child was the only one who presented high accuracy in spelling scores for training words on Unit 11 (85% correct), and she also correctly spelled two of four generalization words on Unit 11. On Unit 24, her spelling score for training words decreased (62.5% correct), and the score for generalization words remained stable. None of the other children correctly spelled a generalization word. Toni spelled one training word correctly on Unit 11 and again on Unit 25. The other students did not spell any training word correctly.

DISCUSSION

This study replicated earlier findings (Sidman, 1971; Sidman & Cresson, 1973), showing that teaching children to match picture comparisons to corresponding dictated-word samples and printed-word comparisons to the same dictated-word samples resulted in the formation of equivalence classes comprised of dictated and printed words, together with corresponding pictures. Students also learned to name the printed words. Because the printed words were equivalent to their pictured referents, it may be concluded that children learned to read with comprehension the set of training words.

Experiment 1 also replicated earlier results (de Rose, de Souza, Rossito, & de Rose, 1989; de Rose et al., 1992; Melchiori, de Souza, & de Rose, 1992) in showing that 5 of 7 participants also developed generalized reading of words formed by recombination of training word syllables. Also, these past findings were extended in Experiment 1 by showing that children made progress in spelling training and generalization words. Participants of Experiment 2 also learned to read the set of training words, but only 1 of the 4 children showed reading and spelling generalization. The achievements of these children support the contention that stimulus control methods, especially those based on stimulus equivalence, may provide the foundation for effective methods to teach reading and spelling (e.g., Sidman, 1994; Stromer, Mackay, & Stoddard, 1992). However, the average IQs and ranges between the groups were different, with students in Experiment 2 having lower measured IQs. This

difference, although minor, may have contributed to the differences in generalization.

Each teaching unit in the present study was designed as an AB experiment, with an initial pretest and final posttest. Pretest scores were generally low and posttest scores increased markedly, documenting the effect of the teaching procedure. Pretest scores themselves often increased during the program, accompanying increases in scores for reading generalization. This trend may reflect the control by minimal textual units (Skinner, 1957); children read new words based on speech sounds that corresponded to letters or groups of letters. It is arguable, however, that students could have learned these textual units via classroom instruction. Typical reading instruction in Brazil focuses on learning to read syllables, which are then recombined and read in many different words. Studies on the Brazilian education system have repeatedly shown, however, that students with poor repertoires do not receive remedial instruction and cannot keep up with progress made by their classmates (Brandão et al., 1983; Fletcher & Costa-Ribeiro, 1987). This happened with participants of Experiment 2, who received initial reading and spelling assessment concurrently with participants of Experiment 1 but were exposed to the teaching procedure only 6 months later. Their performance at the beginning of Experiment 2 suggests that these students gained little in specific reading and spelling skills from 6 months of instruction in the classroom. Because students in Experiment 1 participated in the same teaching conditions, we may infer that they would also make little progress in reading and spelling if they had not been exposed to the teaching program. However, the mean differences in measured IQ may have also influenced these results.

Several features of the teaching procedures may have contributed to their effectiveness. When individuals learn to match visual stimuli to dictated names, they often produce these names when presented with the visual stimuli (Ferrari et al., 1993; Lipkens, Hayes, & Hayes, 1993; Sidman, 1971; Sidman & Cresson, 1973). The exclusion procedure may have contributed to this outcome because it virtually eliminated errors in matching to sample, thus preventing errors from interfering with learning (cf. Stoddard & Sidman, 1967). Reinforcing correct reading responses during the posttests also could have contributed to children learning to read training words and also may have established the absence of reinforcement for a particular reading response as discriminative for changing the response in subsequent tests. It is unlikely, however, that this kind of trial-and-error learning could have produced the highly accurate training word performance demonstrated by all participants; the outcomes of trial-and-error learning are usually not so consistent (e.g., Ferrari et al., 1993; Stoddard & Sidman, 1967).

In our study, children learned to match whole printed words to the corresponding dictated words, but no activities in the teaching program required correspondence between textual elements and sounds. Skinner (1957) suggested that control by smaller textual units may gradually develop, even in the absence of special training, as the repertoire of textual behavior controlled by whole words increases. Control by "minimal" textual units would require that students recognize the correspondence between sounds and minimal textual units, and a "whole-word" approach leaves this to chance (Sidman, 1994). This is probably the reason for the high intersubject variability in reading generalization in the present study. Reading generalization could possibly occur more promptly if procedures to teach explicitly the correspondence between textual units and sounds were added to the program. Two activities in our teaching program may have provided implicit opportu-

nities for students to learn the correspondence between textual elements and sounds. The constructed-response procedure in Experiment 1 may have facilitated acquisition of minimal textual control and, therefore, generalized reading, because it required explicit manipulation of textual units to produce printed words. This suggestion is supported by the absence of reading generalization for most children in Experiment 2, when the constructed-response procedure was omitted. Also, matching printed words to pictures and vice versa in equivalence units possibly provided clues to strengthen emergent control by textual units smaller than words. Recognizing part of a printed word could provide the basis for correct selection, and the ensuing reinforcement would strengthen control by the textual elements that were correctly recognized. This assumption is supported by the fact that reading generalization was shown much earlier when children selected the picture corresponding to a printed generalization word than when they were asked to name printed words. Even students who named few or no generalization words achieved high scores in matching these printed words to pictures. Sidman (1994) suggested that when children acquire control by minimal textual units, the equivalence phenomenon makes it possible for them to advance from sound recognition to reading comprehension. Reading comprehension emerges because the new sound combinations produced by the students have meaning, in the sense that they can be matched to pictures. In addition, as discussed above, the matching task used to assess equivalence may also provide opportunities to strengthen sound recognition.

During the initial assessment, none of the participants were able to spell. They produced sequences of letters or pseudo-letters in which the number of characters was unrelated to the length of the word, and the characters themselves were unrelated to the sounds of the word. Students who showed higher reading generalization scores tended to show more progress in spelling, measured by percentage of words spelled correctly on periodic spelling assessments. If spelling progress were also assessed by approximations to conventional spelling (Lee & Pegler, 1982; Lee & Sanderson, 1987), then all participants improved their spelling. Even children who spelled few or no words entirely correctly showed a gradual increase in the production of sequences of letters that corresponded to the sounds of the words. Because spelling was not directly taught in this study, the reason for this progress is not clear. One possibility is that children had already learned to write (copy) the words in the presence of printed words. Then, as the printed words became equivalent to the dictated words and pictures, the printed words and pictures themselves acquired some degree of control over writing. The constructed-response procedure used in Experiment 1 apparently facilitated this transfer of stimulus control, because in Experiment 2, in which the constructed-response procedure was absent, most children did not show progress in spelling. Again, however, the reason for this is not entirely clear.

The present study extends the literature on stimulus equivalence showing the acquisition of reading and spelling skills of the kind and extent required in a regular classroom. The methods used in this study can be adapted to the needs of the classroom. They can be used with individual students, either in face-to-face or computer-assisted teaching. It is also possible to adapt these methods to group instruction (see Stromer et al., 1992). The results of this study suggest several ways in which stimulus equivalence contributed to the acquisition of reading and spelling skills. The teaching procedures that formed equivalence classes also generated naming of the printed words. Equivalence assured that these words had

meaning, so that students acquired rudimentary reading comprehension. When students learned that words have meaning, they could possibly search for the meaning of those words that they could read only partially, thus strengthening emergent control by smaller textual units. After equivalence classes formed, control exerted by printed words over writing (copying) may have transferred to the other class members (pictures and dictated words), so that students could write these words upon dictation. Future research should address the extent to which equivalence classes contributed to these gains and how equivalence interacted with other aspects of the teaching procedures, such as the constructed-response training.

REFERENCES

Brandão, Z., Baeta, A. M. B., & Rocha, A. D. C. (1983). O estado da arte da pesquisa sobre evasão e repetência no ensino de 1o.grau no Brasil (1971–1981) [The state of the art of research on dropping out and school failure in Brazilian elementary school]. *Revista Brasileira de Estudos Pedagógicos, 64,* 38–69.

de Rose, J. C., de Souza, D. G., Rossito, A. L., & de Rose, T. M. S. (1989). Aquisição de leitura após história de fracasso escolar: Equivalência de estímulos e generalização [Reading acquisition after a history of school failure: Stimulus equivalence and generalization]. *Psicologia: Teoria e Pesquisa, 5,* 325–346.

de Rose, J. C., de Souza, D. G., Rossito, A. L., & de Rose, T. M. S. (1992). Stimulus equivalence and generalization in reading after matching-to-sample by exclusion. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 69–82). Reno, NV: Context Press.

Dixon, L. S. (1977). The nature of control by spoken words over visual stimulus selection. *Journal of the Experimental Analysis of Behavior, 27,* 433–442.

Dube, W. V., McDonald, S. J., McIlvane, W. J., & Mackay, H. A. (1991). Constructed-response matching-to-sample and spelling instruction. *Journal of Applied Behavior Analysis, 24,* 305–317.

Ferrari, C., de Rose, J. C., & McIlvane, W. J. (1993). Exclusion vs. selection training of auditory-visual conditional relations. *Journal of Experimental Child Psychology, 56,* 49–63.

Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. Exeter, NH: Heinemann.

Fletcher, P. R., & Costa-Ribeiro, S. (1987). O ensino de primeiro grau no Brasil de hoje [First grade teaching in contemporary Brazil]. *Em Aberto, 33,* 1–10.

Kazdin, A. E. (1982). *Single-case research designs*. New York: Oxford University Press.

Lee, V. L., & Pegler, A. M. (1982). Effects on spelling of training children to read. *Journal of the Experimental Analysis of Behavior, 37,* 311–322.

Lee, V. L., & Sanderson, G. M. (1987). Some contingencies on spelling. *The Analysis of Verbal Behavior, 5,* 1–13.

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology, 56,* 201–239.

Lynch, D. C., & Cuvo, A. (1995). Stimulus equivalence instruction of fraction-decimal relations. *Journal of Applied Behavior Analysis, 28,* 115–126.

Mace, F. C. (1994). Basic research needed for stimulating the development of behavioral technologies. *Journal of the Experimental Analysis of Behavior, 61,* 529–550.

Mackay, H. A. (1985). Stimulus equivalence in rudimentary reading and spelling. *Analysis and Intervention in Developmental Disabilities, 5,* 373–387.

Mackay, H. A., & Sidman, M. (1984). Teaching new behavior via equivalence relations. In P. H. Brooks, R. Sperber, & C. MacCauley (Eds.), *Learning and cognition in the mentally retarded* (pp. 493–513). Hillsdale, NJ: Erlbaum.

Matos, M. A., & D'Oliveira, M. M. H. (1992). Equivalence relations and reading. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 83–94). Reno, NV: Context Press.

McIlvane, W. J., & Stoddard, L. T. (1981). Acquisition of matching-to-sample performances in severe mental retardation: Learning by exclusion. *Journal of Mental Deficiency Research, 25,* 33–48.

Melchiori, L. E., de Souza, D. G., & de Rose, J. C. (1992). Aprendizagem de leitura através de um procedimento de discriminação sem erros (exclusão): Uma replicação com pré-escolares [Reading acquisition through an errorless discrimination procedure (exclusion): A replication with preschoolers]. *Psicologia: Teoria e Pesquisa, 8,* 101–111.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14,* 5–13.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

Sidman, M. (1994). *Equivalence relations: A research story*. Boston, MA: Authors Cooperative.

Sidman, M., & Cresson, O., Jr. (1973). Reading and crossmodal transfer of stimulus equivalences in se-

vere retardation. *American Journal of Mental Retardation, 77,* 515–523.
Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.
Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice Hall.
Stoddard, L. T., & Sidman, M. (1967). The effects of errors on children's performance on a circle-ellipse discrimination. *Journal of the Experimental Analysis of Behavior, 10,* 261–270.
Stromer, R., & Mackay, H. A. (1992). Delayed constructed-response identity matching improves the spelling performance of students with mental retardation. *Journal of Behavioral Education, 2,* 139–156.
Stromer, R., Mackay, H. A., & Stoddard, L. T. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education, 2,* 225–256.
Wechsler, D. (1949). *Wechsler intelligence scale for children (WISC).* New York: Psychological Corporation.

Received April 28, 1995
Initial editorial decision August 1, 1995
Revisions received January 8, 1996; April 10, 1996
Final acceptance June 17, 1996
Action Editor, Anthony J. Cuvo

STUDY QUESTIONS

1. In their introduction, the authors noted that three properties define an equivalence relation among stimuli. Define these properties and, using a ball, a picture of a ball, and the printed word *ball* as stimuli, describe the matching relationships that exemplify each property.

2. What is meant by the term *exclusion training* during match-to-sample instruction? How is it similar to and different from typical match-to-sample training and what are its potential advantages?

3. What types of stimuli were used to form equivalence classes?

4. Briefly describe the instructional procedures that were used to establish the different properties of the equivalence relationship.

5. The exclusion units included three types of matching trials. Describe both the composition and the function of each type of trial.

6. What type of experimental design was used to evaluate the effects of the teaching program? What features of the design increased the likelihood that results were not merely a function of other processes such as maturation or training done outside of the experimental setting?

7. How did Experiments 1 and 2 differ in terms of procedures and results, how were the results interpreted by the authors, and how might their interpretation have been strengthened?

8. The authors did not include an evaluation of the exclusion training component as part of either experiment. Briefly describe how such an evaluation could have been conducted.

Questions prepared by Iser DeLeon and SungWoo Kahng, University of Florida