JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1992, 25, 893–904 NUMBER 4 (WINTER 1992)

SPELLING AND EMERGENT PICTURE-PRINTED WORD RELATIONS ESTABLISHED WITH DELAYED IDENTITY MATCHING TO COMPLEX SAMPLES

ROBERT STROMER AND HARRY A. MACKAY

EUNICE KENNEDY SHRIVER CENTER AND NORTHEASTERN UNIVERSITY

Students with academic deficits learned delayed matching-to-sample tasks that used complex sample stimuli, each consisting of a picture and a printed word. A touch to the sample complex removed it from the computer display and produced either picture comparisons or a choice pool of letters. If the comparisons were pictures, selecting the picture identical to the preceding sample was reinforced. If the letters appeared, letter-by-letter construction of the preceding printed word sample was reinforced. The procedure engendered new constructed-response spelling performances and arbitrary relations among pictures and printed words in matching to sample. The emergence of relations among different sets of printed words (paired with the same pictures) suggested classes of equivalent stimuli. Outcome tests involving spoken words as sample stimuli suggested expansion of subjects' spelling repertoires and stimulus classes.

DESCRIPTORS: language arts instruction, spelling, stimulus equivalence, matching to sample, special education

Studies of stimulus equivalence contribute to the analysis of the relationships among language skills and illuminate how methodological factors may influence potential training outcomes (e.g., Mackay, 1991; Mackay & Sidman, 1984; Singh & Singh, 1986; Stromer, 1991; Stromer, Mackay, & Stoddard, 1992). Sidman's (1971) classic study of equivalence is illustrative: A youth with mental retardation was able to perform the receptive task of matching picture comparison stimuli to their spoken names presented as sample stimuli and the expressive task of naming the pictures aloud. However, the subject was unable (a) to match printed words to the same spoken words, (b) to match the printed words to the pictures, (c) to match the pictures to the printed words, or (d) to name the printed words. The subject then learned to match printed words to spoken words, a form of auditory receptive reading. Remarkably, the three remaining performances emerged without direct teaching. The emergence of matching printed words to pictures, and vice versa, qualify as rudimentary forms of reading comprehension; naming the printed words qualifies as oral reading.

The matching-to-sample performances that emerged in Sidman's (1971) study suggest the formation of classes of equivalent stimuli. Equivalence classes are verified when two or more independently established stimulus relations have a stimulus term in common and when the properties of *reflexivity, symmetry,* and *transitivity* are demonstrated behaviorally (Sidman & Tailby, 1982). The property of reflexivity is verified by generalized identity matching with the experimental stimuli. Symmetry is verified if the sample–comparison relations trained directly are reversible: If picture samples control selections of printed-word comparisons, the printed words should control selections of the pictures. Transitivity is verified by the emergence of new stimulus relations involving those trained separately but linked to one another by a common stimulus. For example, if spoken-word samples control selections of picture comparisons, and the same pictures control selections of printed words, then the spoken words should control selections of the printed words.

This research was supported by the National Institute of Child Health and Human Development (Grants HD04147 and HD25995) and the Massachusetts Department of Mental Retardation (Contract 3404-8403-306). We are grateful to Carol Cole, Tucker Walton, and their colleagues at the Dearborn Academy for their assistance and cooperation throughout the investigation. We also thank Bill Dube for writing the computer software and Bill McIlvane for helpful comments on a draft of the paper.

Address reprint requests to Robert Stromer, Behavioral Sciences Division, Eunice Kennedy Shriver Center, 200 Trapelo Road, Waltham, Massachusetts 02254.

In Sidman's (1971) study, the relations established between the spoken-word samples and pictures (before the experiment) and between the same spoken words and printed words (within the experiment) provided the prerequisites for assessing equivalence. The performances that emerged among printed words and pictures permit an inference of both symmetry and transitivity that, in turn, supports an inference of equivalence class formation, each class consisting of a spoken word, a picture, and a printed word (for details of such "combined tests," see Sidman, 1986, 1990; Sidman & Tailby, 1982).

Mackay (1985) extended Sidman (1971) by analyzing spelling performances in the context of stimulus equivalence (see also Mackay & Sidman, 1984). Subjects with mental retardation displayed the same initial performances as Sidman's subject, but the stimuli were colored patches and their corresponding spoken and printed words. Next, they learned an anagram spelling performance that involved constructing words with individual letter tiles. For example, given a patch of yellow as the sample stimulus, the subjects learned to rearrange the letters *l, w, y, l, e,* and *o* to "spell" the word *yellow.* Subjects then showed the same emergent performances observed in Sidman's study, suggesting the formation of equivalence classes.

Dube, McDonald, McIlvane, and Mackay (1991) recently described a computerized version of Mackay's constructed-response spelling procedure. Their procedure involved displaying a printed word with its picture as a complex sample and gradually fading out the printed word. Here, we describe a computerized method for establishing word constructions to pictures without the need to research and develop such fading programs, and extend Dube et al.'s work further by asking whether the procedure provides the basis for equivalence class formation.

Our procedure involved either delayed indentity matching of the pictures or delayed construction of the words presented in complex sample stimuli, each composed of both a picture and a printed word. Put simply, the procedure encouraged looking at, and remembering, both the picture and the printed word. For example, in an initial training phase, the sample complex involved a picture of a dog and the printed word *dog.* A touch to the sample removed it from the computer display and immediately produced the comparison stimuli. On some trials the comparison stimuli were pictures, and selecting the picture of the dog was reinforced. On other trials, a pool of letters appeared, and selecting the letters *d, o,* and *g* in that order was reinforced. As in spelling more generally, construction of the printed word required discrimination of each of the letters in the word and their left-to-right order.

The critical training next involved pairing different printed words with the same pictures used initially. For example, one new complex sample involved the picture of the dog and the printed word *canine.* On some trials, selecting the identical picture was reinforced as before; but on other trials, selecting the letters *c, a, n, i, n,* and *e* was reinforced. Performances assessed without further teaching included (a) constructing the word *canine* to the picture of dog, (b) matching the word *canine* and the picture of dog to one another, and (c) matching the words *dog* and *canine* to one another. The emergence of relations between the new and the initial printed words would satisfy the requirements of symmetry and transitivity and would suggest the formation of equivalences (cf. Sidman & Tailby, 1982). Such results would also replicate previous findings with college students using complex-sample procedures (Stromer & Stromer, 1990a, 1990b). In the final phase of the study, we assessed constructed-response spelling with spoken words as sample stimuli; we also examined the expansion of equivalence classes by assessing the emergence of relations between printed and spoken words (e.g., matching the printed word *canine* to the spoken word "dog").

METHOD

Subjects and Setting

Three boys, 9 to 13 years of age, participated. Each boy was enrolled in a private, nonprofit, therapeutic day school providing services for students

Table 1
Subject Characteristics

	Curt	Andy	Bill
Chronological age (years-months)	9-7	11-3	13-11
Intelligence testing[a]			
Verbal IQ	95	80	69
Performance IQ	74	93	68
Full-scale IQ	84	85	67
Achievement testing (grade equivalent scores)			
Reading[b]	2.7	1.8	1.3
Spelling[c]	2.5	2.5	2.2
Arithmetic[d]	3.2	3.5	5.3
General information[e]	5.6	3.4	3.9

[a] Wechsler Intelligence Scale for Children–Revised.
[b] Woodcock Reading Mastery.
[c] Wide Range Achievement Test.
[d] Keymath Diagnostic Arithmetic Test.
[e] Peabody Individual Achievement Test.

with emotional, behavioral, and learning disorders. Classrooms at the school were ungraded but were administratively organized into elementary-, middle-, and secondary-school programs. Curt and Andy participated in the elementary program, and Bill was in the middle-school program. Table 1 summarizes data from standardized tests. Curt participated in a prior study of delayed matching to sample; Andy and Bill were experimentally naive. Except for unavoidable schedule changes, three sessions were conducted each week (Monday, Wednesday, and Friday) in a quiet area at school. A trainer, seated behind and to the subject's right, monitored all sessions.

Apparatus

A Macintosh® computer with a touch-sensitive screen presented stimuli, recorded responses, and stored session data for analysis (Dube & McIlvane, 1989). Visual sample stimuli (pictures and printed words) appeared at top-center in the computer display (Figure 1); auditory samples (digitized spoken words) were presented by a speaker attached to the computer. The sample area was also used to display the words constructed during spelling trials. Comparison stimuli were presented in the lower portion of the display and were either trios of pictures or printed words or displays of 10 letters. The locations at which visual samples and comparisons appeared functioned as response "keys." Other aspects of the computerized procedure are described below.

Phase 1 Procedure

Overview and design. Table 2 outlines the procedures. A pretraining phase acquainted subjects with the operation of the computer and some of the basic tasks. Subsequent conditions established new constructed-response spelling performances and three four-member stimulus classes. The "dog class" consisted of a picture of a dog and the printed words *dog, canine,* and *pisces*; the "cat class" consisted of a picture of a cat and the words *cat, feline,* and *virgo*; the "owl class" consisted of a picture of an owl and the words *owl, avian,* and *taurus*. In addition to the pictures there were three sets of printed-word training stimuli: *dog, cat,* and *owl* (Set 1); *canine, feline,* and *avian* (Set 2); and *pisces, virgo,* and *taurus* (Set 3). The influence of extraexperimental variables was minimized because the matching relations studied involving Set 3 words had no "real-world" basis but were experimentally contrived (e.g., matching the printed words *pisces* and *canine* to one another).

A multiple baseline design was used to assess outcome performances derivable from the three

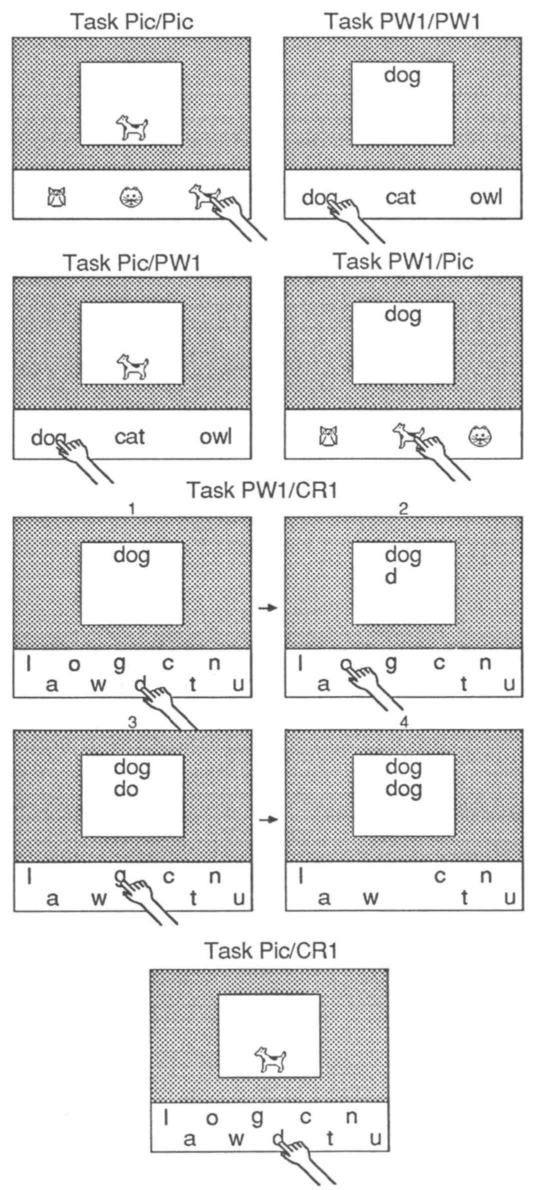

Figure 1. Panels depict computer displays of six trial types used during pretraining with Set 1 stimuli in Phase 1 of the experiment. Top four panels illustrate matching picture comparisons to picture samples (Pic/Pic), matching Set 1 printed words to one another (PW1/PW1), matching printed words to pictures (Pic/PW1), and vice versa (PW1/Pic). Panels 1 through 4 depict constructed-response identity matching (PW1/CR1); bottom panel depicts constructed-response matching (spelling) to pictures.

training conditions (cf. Horner & Baer, 1978). Initially, all relevant relations involving the pictures and printed words from Set 1 were trained; various outcome performances with the other stimuli were then pretested (Tests 1 through 7). Accuracy scores on these tests were expected to be below or near chance levels. Next, training directly related the same pictures with words from Set 2, and the outcome tests were repeated. At this point we anticipated increases only in performances involving the pictures and printed words from Sets 1 and 2 (Tests 1 through 3); all other performances should remain at pretest levels. Finally, training was given using the same pictures and words from Set 3, and the outcome tests were repeated again. These tests should now reflect increases in all remaining performances (Tests 4 through 7).

Pretraining with Set 1. Subjects were first given six simultaneous matching tasks involving pictures of a dog, cat, and owl and their corresponding printed words (Set 1). The top left panel in Figure 1 shows a trial requiring selection of a picture comparison given a picture sample. This task is denoted Pic/Pic to express the sample–comparison relation. Three other tasks are PW1/PW1 (matching Set 1 printed words to one another), Pic/PW1 (matching Set 1 printed words to pictures), and PW1/Pic (matching pictures to Set 1 printed words). Panels at the bottom depict a constructed-response trial in which the sample stimulus is a printed word (PW1/CR1). The subject's task was to touch letters from the choice pool that corresponded to the printed-word sample. For example, with the printed word *dog* as the sample, a touch to the letter *d* moved it from the choice pool to the construction area; subsequent touches to letters *o* and *g* moved the letters to the construction area and completed the trial. A similar task with picture samples (Pic/CR1) was also pretrained. Blocks of six trials were given on each pretraining task in the order listed in Table 2. The criteria were one session with (a) an accuracy score of 5 of 6 or better on each of the six tasks and (b) at least 92% accuracy across all 36 trials.

All trials began with a sample stimulus displayed in the upper window of the computer display. During simultaneous matching trials, touches to sample stimuli added comparison stimuli that remained on the display until completion of the subject's comparison selection or word construction. Selections of a correct comparison or correct word construc-

Table 2
Outline of Procedures and Trial Types Used in Phase 1

Pretraining with Set 1 (dog, cat, owl)
Pic/Pic: Matching pictures to pictures
PW1/PW1: Matching Set 1 printed words to one another
Pic/PW1: Matching Set 1 printed words to pictures
PW1/Pic: Matching pictures to Set 1 printed words
PW1/CR1: Constructing Set 1 words to printed words
Pic/CR1: Constructing Set 1 words to pictures
Further training with Set 1
Pic/CR1: Constructing Set 1 words to pictures
PW1/Pic: Matching pictures to Set 1 printed words
Pic+PW1/Pic: Matching pictures to complex samples (pic + Set 1 words)
Pic+PW1/CR1: Constructing Set 1 words to complex samples (pic + Set 1 words)
Testing
Test 1: Pic/CR2: Constructing Set 2 words to pictures
Test 2: Pic/PW2: Matching Set 2 printed words to pictures
PW2/Pic: Matching pictures to Set 2 printed words
Test 3: PW1/PW2: Matching Set 2 printed words to Set 1 printed words
PW2/PW1: Matching Set 1 printed words to Set 2 printed words
Test 4: Pic/CR3: Constructing Set 3 words to pictures
Test 5: Pic/PW3: Matching Set 3 printed words to pictures
PW3/Pic: Matching pictures to Set 3 printed words
Test 6: PW1/PW3: Matching Set 3 printed words to Set 1 printed words
PW3/PW1: Matching Set 1 printed words to Set 3 printed words
Test 7: PW2/PW3: Matching Set 3 printed words to Set 2 printed words
PW3/PW2: Matching Set 2 printed words to Set 3 printed words
Training with Set 2 (canine, feline, avian); repeat tests
Pic+PW2/Pic: Matching pictures to complex samples (pic + Set 2 words)
Pic+PW2/CR2: Constructing Set 2 words to complex samples (pic + Set 2 words)
Pic/CR1: Constructing Set 1 words to pictures
PW1/Pic: Matching pictures to Set 1 printed words
Pic+PW1/Pic: Matching pictures to complex samples (pic + Set 1 words)
Pic+PW1/CR1: Constructing Set 1 words to complex samples (pic + Set 1 words)
Training with Set 3 (pisces, virgo, taurus); repeat tests
Pic+PW3/Pic: Matching pictures to complex samples (pic + Set 3 words)
Pic+PW3/CR3: Constructing Set 3 words to complex samples (pic + Set 3 words)
Pic+PW1/Pic: Matching pictures to complex samples (pic + Set 1 words)
Pic+PW1/CR1: Constructing Set 1 words to complex samples (pic + Set 1 words)
Pic/CR1: Constructing Set 1 words to pictures
PW1/Pic: Matching pictures to Set 1 printed words
Pic+PW2/Pic: Matching pictures to complex samples (pic + Set 2 words)
Pic+PW2/CR2: Constructing Set 2 words to complex samples (pic + Set 2 words)

Note. Figures 1 through 4 illustrate the trial types outlined in these training and testing conditions.

tions (performances consistent with English-speaking conventions) produced a melodic tune and a flashing computer screen. In addition, at the end of each session, subjects received 1 cent for each correct trial (amount earned displayed on the computer screen at session's end). Trials were separated by an intertrial interval in which all stimuli were absent from the display for 1.5 s. Touches to an incorrect comparison (or letters out of sequence) produced the intertrial interval followed by the next trial. The 10 letters in the choice pool always included the eight needed to spell Set 1 words (Figure 1). (On other constructed-response trials, the letters in the choice pool were those that made up the particular word appearing in the sample and selected letters from the other two words in the set.) The particular sample stimulus and the positions of comparison stimuli and letters in the choice pool varied unsystematically from trial to trial.

Further training with Set 1. Training consisted of 36 trials with Set 1 stimuli; nine with each of the four trial types shown in Table 2. Tasks Pic/CR1 and PW1/Pic involved constructed-response matching to picture samples and the matching of picture comparisons to printed-word samples, respectively. Task Pic+PW1/Pic involved matching pictures to sample stimuli that combined pictures and printed words; Task Pic+PW1/CR1 involved constructed-response matching to the same complex samples. These tasks provided a relevant baseline to which the critical training and test trials were added in this and subsequent conditions.

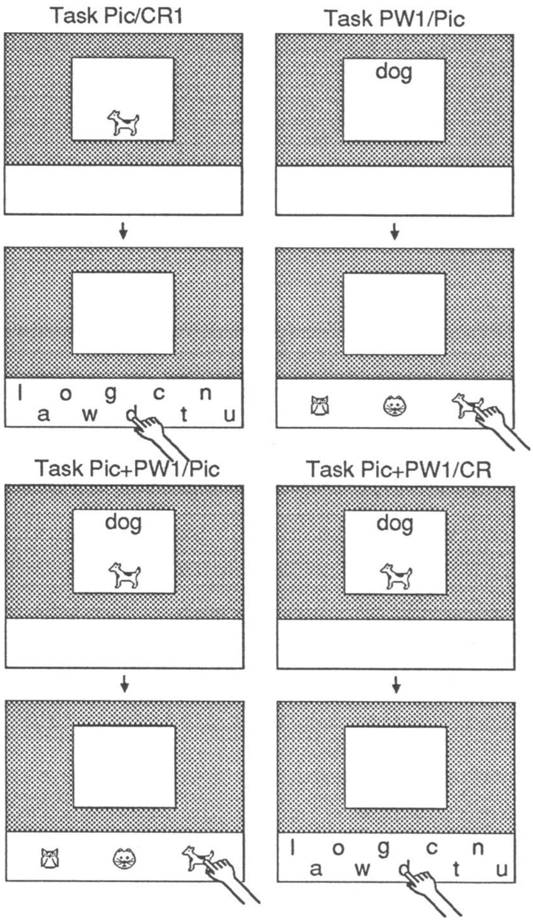

Figure 2. Examples of delayed matching trials used in Phase 1 training with Set 1 stimuli.

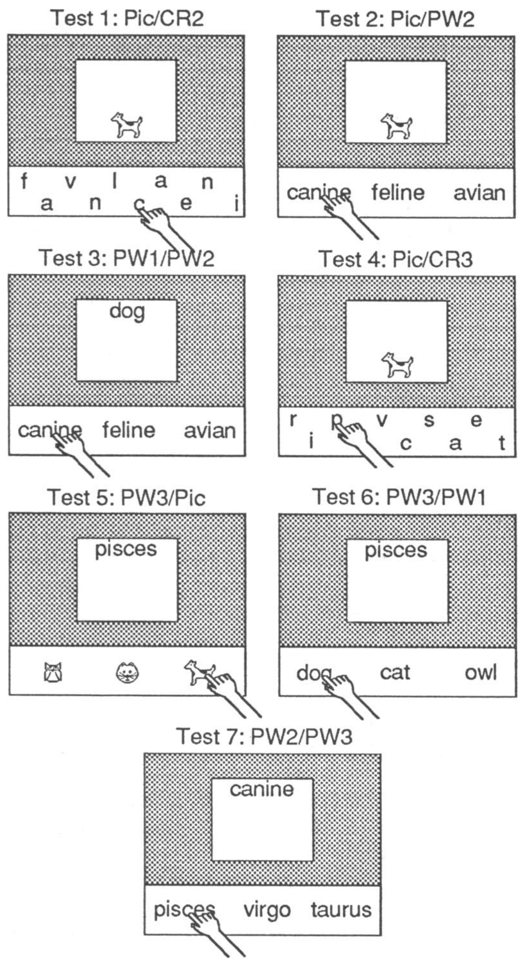

Figure 3. Examples of delayed matching trials used in Phase 1 testing (delay periods are not illustrated, as in Figure 2, to save space).

For Andy and Bill, these four trial types occurred first as simultaneous matching tasks, and training continued until performance in three consecutive sessions was at least 92% correct under the contingencies used in pretraining; the same criterion was then applied to zero-delay sessions. Curt received only zero-delay sessions. Figure 2 illustrates the zero-delay matching procedure in which a touch to the sample stimulus removed it from the display at the same time the comparisons appeared. The bottom panels illustrate the delayed matching arrangement with the complex-sample trials. After the sample stimulus was removed from the display, the comparisons were either three pictures or the choice pool of letters.

Testing. Testing consisted of a block of three consecutive 64-trial sessions (zero delay) without immediate consequences. However, 64 cents was given at the end of each session regardless of performance. Subjects were told of these contingencies: "For the next few days the computer won't jingle or flash but it will still tell you how much money you get at the end." During testing, 28 test trials were mixed with 36 baseline trials. Of the 28 test trials, 20 involved two each of the 10 matching-to-sample trial types in Tests 2, 3, 5, 6, and 7 (see Table 2). The trial types differed across sessions so that each of the six possible matching relations in each test was assessed twice (12 total trials for each test). The eight other test trials in each session were

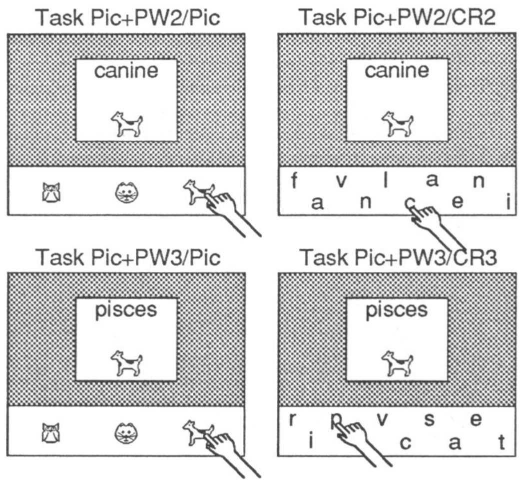

Figure 4. Examples of delayed matching trials used in Phase 1 training with Set 2 and Set 3 stimuli.

word constructions, four each of Tests 1 and 4 (12 total for each).

Figure 3 shows representative trial types from Tests 1 through 7. For example, at top left is a Pic/CR2 trial (Test 1): Given the picture of a dog, and the choice pool of letters, arranging the letters to spell *canine* was correct. Task Pic/CR3 (Test 4) involved the same sample, but construction of the words in Set 3 was assessed. The letters in the choice pool were those that made up the particular word being tested and selected letters from the other two words. The remaining tasks in Figure 3 (and outlined in Table 2) involved matching pictures and printed words to one another (e.g., Pic/PW2 and PW3/Pic) and matching printed words from the three sets to one another (e.g., PW1/PW2, PW2/PW1, and PW3/PW1).

Training with Set 2; repeat tests. Figure 4 shows the complex-sample trials added to the baseline. Trial types Pic+PW2/Pic and Pic+PW2/CR were trained initially by establishing constructed-response matching with only printed words as sample stimuli; the criterion was six consecutive correct trials. Next, a 36-trial mixture of both complex-sample types (18 of each) was given until one session of at least 92% accuracy occurred. These two trial types (six of each) were then combined with 24 of the Set 1 baseline trials (six of each). The block of three test sessions was given after three consecutive baseline sessions were at least 92% correct.

Training with Set 3; repeat tests. The trial types Pic+PW3/Pic and Pic+PW3/CR3 (Figure 4) were used in training as described above, and then were mixed with trials of previously trained performances. The testing baseline in this phase consisted of (a) three of each of the four Set 1 baseline trials (Table 2), (b) six of each of the two complex-sample trials used during the Set 2 training, and (c) six of each of the trials trained during this phase. Only Curt and Andy participated in this phase; Bill was not enrolled in the school at the time of these conditions.

Results

Training. Pretraining took two sessions for Curt and Andy and four sessions for Bill. During training with Set 1, Curt was perfect in three baseline sessions given under the zero-delay procedure. Andy and Bill met the training criteria in this condition in six sessions and eight sessions, respectively. With Set 2, Curt's initial training involved 89 trials, and the mixed baseline criterion was met in three sessions. For Andy, these measures were 85 initial training trials and six baseline sessions, and for Bill 139 trials and seven baseline sessions. During training with Set 3, Curt's initial training took 89 trials and baseline lasted three sessions. Andy's initial training took 155 trials and baseline lasted three sessions.

Testing. Figure 5 shows each subject's probe data plotted as the number of correct responses (out of 12) during Tests 1 through 7. The top panel shows data gathered after the initial baseline training involving the Set 1 printed words; the middle and lower panels show data gathered after trained relations involving Sets 2 and 3, respectively, were added to the testing baseline. The first three sets of bars (reading left to right) represent test performances derivable from training with Sets 1 and 2; the remaining four sets of bars reflect outcomes linked to training Sets 1, 2, and 3.

The top panel of Figure 5 shows that test scores for all 3 subjects initially ranged from 0 to 6 (e.g., Curt and Bill on Tasks PW1/PW2 and PW2/

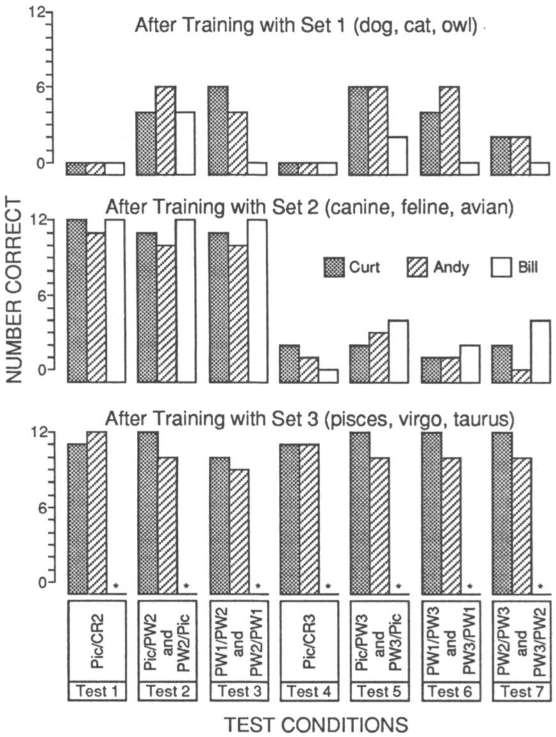

Figure 5. Subjects' performance on Phase 1 tests as a function of training applied to three sets of stimuli. Bars represent number of correct trials out of 12 for a given trial type tested. Trial types used in each test are listed along the abscissa. (Bill was unavailable for tests denoted by an asterisk.)

PW1). The middle panel shows that after training with Set 2, scores for all subjects improved to 10 correct trials or better on Tests 1 through 3, the tasks involving Sets 1 and 2. In contrast, performance on the remaining tasks, which involved Set 3, was low for all 3 subjects. However, the performance of Curt and Andy on these tasks improved markedly after they were trained on relations involving Set 3 printed words (bottom panel).

Phase 2 Procedure

The purpose of Phase 2 was (a) to assess constructed-response spelling of words in Sets 2 and 3 given spoken words as sample stimuli, and (b) to assess relations among spoken samples and either pictures or printed words as comparisons. Positive results would demonstrate an additional spelling performance that was not taught directly and would suggest that spoken words had been added to the existing equivalence classes. Table 3 outlines the procedures.

Training. The delayed matching baseline included the six trial types (three trials of each) displaying picture or printed-word samples as shown in Figure 2 (bottom panels) and Figure 4. The baseline also included the four auditory-visual trial types depicted in Figure 6 (six of each). Auditory samples were repeated every 2 s until the blank (black) sample area was touched. The touch to the sample area terminated the auditory stimulus and turned the sample area to white. This 42-trial baseline was presented until 93% accuracy occurred for three consecutive sessions under continuous reinforcement.

Testing. Figure 7 illustrates the test displays, and Table 3 outlines the 10 trial types that were arranged into the five tests given. Each of two consecutive test sessions involved 30 test trials (three of each of the 10 trial types) and 42 baseline trials.

Results

The baseline criteria were reached in four sessions by Curt and Andy. Figure 8 shows that the performance of both subjects was nearly perfect across all test conditions.

DISCUSSION

The present results recommend delayed identity matching to complex samples as a method to teach spelling and relations among pictures and printed words. The procedure offers an alternative to prompt-fading methods for teaching constructed-response spelling (Dube et al., 1991). The findings also extend prior studies of stimulus equivalence between pictures and their printed and spoken names (Sidman, 1971; Sidman & Cresson, 1973) and the analysis of spelling performances (Mackay, 1985; Mackay & Sidman, 1984; Stromer, 1991). Finally, the results contribute to the development of procedures for establishing an integrated set of rudimentary language performances (Stromer et al., 1992).

In Phase 1, we established delayed matching

Table 3
Outline of Procedures and Trial Types Used in Phase 2

Training

"Wrd1"/Pic: Matching pictures to Set 1 spoken words
"Wrd1"/PW1: Matching Set 1 printed words to Set 1 spoken words
"Wrd2"/PW2: Matching Set 2 printed words to Set 2 spoken words
"Wrd3"/PW3: Matching Set 3 printed words to set 3 spoken words
Pic+PW1/Pic: Matching pictures to complex samples (pic + Set 1 words)
Pic+PW1/CR1: Constructing Set 1 words to complex samples (pic + Set 1 words)
Pic+PW2/Pic: Matching pictures to complex samples (pic + Set 2 words)
Pic+PW2/CR2: Constructing Set 2 words to complex samples (pic + Set 2 words)
Pic+PW3/Pic: Matching pictures to complex samples (pic + Set 3 words)
Pic+PW3/CR3: Constructing Set 3 words to complex samples (pic + Set 3 words)

Testing

Test 1: "Wrd 2"/CR2: Constructing Set 2 words to spoken words
"Wrd3"/CR3: Constructing Set 3 words to spoken words
Test 2: "Wrd2"/Pic: Matching pictures to Set 2 spoken words
"Wrd3"/Pic: Matching pictures to Set 3 spoken words
Test 3: "Wrd1"/PW2: Matching Set 2 printed words to Set 1 spoken words
"Wrd1"/PW3: Matching Set 3 printed words to Set 1 spoken words
Test 4: "Wrd2"/PW1: Matching Set 1 printed words to Set 2 spoken words
"Wrd3"/PW1: Matching Set 1 printed words to Set 3 spoken words
Test 5: "Wrd2"/PW3: Matching Set 3 printed words to Set 2 spoken words
"Wrd3"/PW2: Matching Set 2 printed words to Set 3 spoken words

Note. Figures 6 and 7 illustrate the trial types outlined in these training and testing conditions.

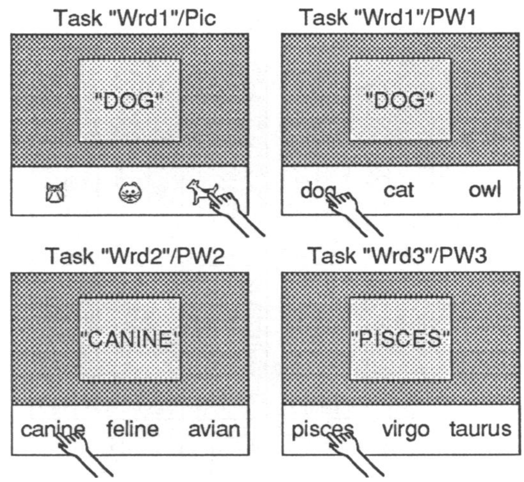

Figure 6. Examples of delayed matching trials in Phase 2 training.

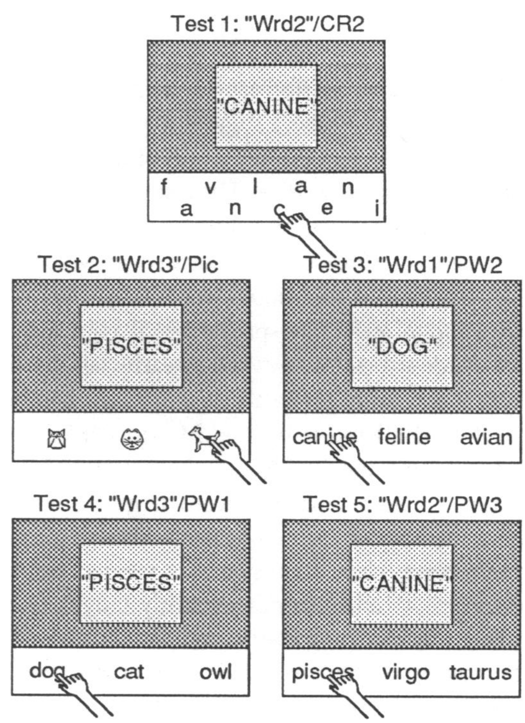

Figure 7. Examples of delayed matching trials in Phase 2 testing.

performances that included trials in which a picture and a printed word served simultaneously as sample stimuli. The potential for equivalence classes existed because these sample elements may become related to one another and each of the different complex stimuli contained a common picture element (cf. Stromer & Stromer, 1990a, 1990b). Indeed, the test data suggest that a history with the complex-sample trials engendered new relations among pictures and printed words: (a) Word construction was accomplished in response to the pictures alone (Tests 1 and 4), and (b) printed words and pictures were matched to one another (Tests 2 and 6).

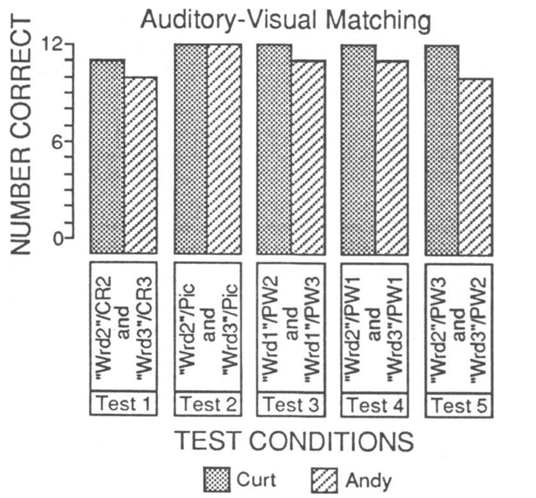

Figure 8. Subjects' performance on Phase 2 tests following all teaching conditions. Bars represent number of correct trials out of 12 for a given trial type tested. Trial types used in each test are listed along the abscissa. (Bill was unavailable for these tests.)

Previous research suggests that either constructing or matching words to pictures suffices to establish equivalence relations (Mackay, 1985; Mackay & Sidman, 1984; Sidman, 1971; Sidman & Cresson, 1973).

The emergent performances demonstrated in Tests 3 and 7 of Phase 1 suggest that the relations between the different printed words of each set were symmetrical and transitive. The test outcomes thus support the inference of equivalence class formation: Each of three classes consisted of a picture (e.g., dog) and three printed words (e.g., *dog, canine,* and *pisces*). The properties of equivalence then made it possible to expand each class by relating a new stimulus to an existing class member (e.g., Sidman, Kirk, & Willson-Morris, 1985). This was accomplished in Phase 2 by first training a relation between a spoken word (e.g., "canine") and a member of each equivalence class (e.g., *canine*). Next, class expansion was verified by the emergence of relations not explicitly taught (e.g., in Test 3 by matching *canine* and *pisces* to "dog"). The test results suggest that each of the classes established in Phase 1 now included three spoken words (e.g., "dog," "canine," and "pisces").

The internal validity of the study could be questioned because in Phase 1 the design was only a two-tier multiple baseline and in Phase 2 the emergent relations were only posttested. However, the potential influence of extraexperimental variables was minimized by studying some matching relations that had no basis in everyday experience (those involving zodiac names). By doing this, of course, we raise concerns about external validity. In retrospect, we probably would have achieved the same results by using a name for each of the animals (e.g., Rover, Goldie, and Snowy), as in reading primers, rather than arbitrarily assigned zodiac names. Alternatively, the classes studied here could include printed and spoken words from languages other than English, classes of numerical stimuli, or naturally occurring language categories (e.g., fruits and vegetables). Our future research will conduct more formal analyses of the procedures with educationally relevant stimuli such as these.

The present delayed matching procedures may have ensured discrimination of all sample elements and discouraged the selective stimulus control that often appears in the performance of children with specific or general developmental delays (e.g., Bailey, 1981; Lovaas, Koegel, & Schreibman, 1979). In particular, our procedures may have limited the potential for pictures to "block" control by printed words (Singh & Solman, 1990) because we (a) established identity word construction performance before training with complex samples and (b) required both picture matching and word construction to the complex samples. These methods also have been effective for students with moderate mental retardation (Stromer & Mackay, in press). We acknowledge, however, possible limitations with students functioning at lower levels. Data from our laboratory suggest that many students with mental retardation lack the prerequisites for acquiring the baselines used in the present study. A research priority is to discover those prerequisites.

Equivalence classes might well emerge even if the present general procedure required only identity matching of pictures and printed words. However, it is unlikely that such a procedure would produce word constructions. For example, subjects in the study by Stromer and Mackay (1992) could match pictures and printed words to their spoken names, and could match the pictures and printed words to one another. They could not, however, construct the words in response to the pictures or the spoken words. Matching-to-sample trials could be arranged to require discrimination of individual letters and their order just as in the construction procedure. For example, with the word *dog* as the sample, the comparisons might be *dog, dig, dot,* and so on. If a student learns to make such discriminations, word constructions and other forms of spelling (e.g., written and oral) may occur. In contrast, word-construction methods have the advantage of teaching directly the left–right sequencing required in spelling, can accomplish the same outcomes, and are implemented relatively easily.

"True" spelling requires that the letters of a word function as a unitary or compound stimulus. The unique arrangement of letters makes a word discriminable from all other words. In contrast, the construction procedure could establish performances involving only rote sequences of letter selection under the control of the particular letters available in the choice pool on a trial. Some word-construction tasks clearly required conditional discrimination because the letters for spelling all words in a set were available on every trial (e.g., Set 1). However, the letters in the choice pools were limited on other trials, making it possible to spell only one training word (e.g., Set 3). The matching-to-sample data do not clarify this aspect of performance because the matching procedures did not require discrimination of each letter of the printed words and their order. The use of the computer keyboard as the response device may circumvent the foregoing constraints because all letters of the alphabet are available on every trial (e.g., Stevens, Blackhurst, & Slaton, 1991). However, the demands of operating the keyboard will exclude some students from its use.

Word-construction methods may lend themselves to teaching many behaviorally limited students, and the inclusion of written and oral tests could verify true spelling as an outcome performance. Initial assessments of this possibility have been successful (Stromer & Mackay, 1992, in press). The results of one of these studies (Stromer & Mackay, in press) also support prior observations suggesting that equivalence-based teaching procedures may yield emergent naming performances (Mackay, 1985; Mackay & Sidman, 1984; Sidman, 1971; Sidman & Cresson, 1973). These and the present results suggest that stimulus equivalence concepts and methods provide a valuable framework for analyzing language performances involving pictures, printed words, and dictated names (cf. Stromer et al., 1992).

A major benefit of this approach is its economy of teaching; a tremendous amount of new behavior occurs after only a few skills are trained directly. Such findings contribute to the development of a technology for teaching generative academic performances (cf. Alessi, 1987; Mackay, 1991; Stokes & Baer, 1977; Stromer, 1991). The analysis also has implications for *how* language arts are taught. For example, establishing equivalences between spoken words, pictures, and printed words alone may not ensure reading and spelling (Stromer, 1991). Further, the addition of reading to such equivalences may or may not yield spelling (Lee & Pegler, 1982; Stromer, 1991; Stromer & Mackay, 1992). In contrast, an explicit focus on spelling from the outset can yield both equivalences and reading (Mackay, 1985; Mackay & Sidman, 1984; Stromer & Mackay, in press). Such a focus goes against traditional practice in the early grades but may be crucial if all children are to benefit fully from instruction that seeks to establish integrated repertoires of language arts skills (cf. Tierney & Shanahan, 1991).

REFERENCES

Alessi, G. (1987). Generative strategies and teaching generalization. *The Analysis of Verbal Behavior,* **5,** 15–27.

Bailey, S. L. (1981). Stimulus overselectivity in learning disabled children. *Journal of Applied Behavior Analysis,* **14,** 239–248.

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.

Dube, W. V., & McIlvane, W. J. (1989). Adapting a microcomputer for behavioral evaluation of mentally retarded individuals. In J. A. Mulick & R. F. Antonak (Eds.), *Transitions in mental retardation* (Vol. 4, pp. 104–127). Norwood, NJ: Ablex.

Horner, R. D., & Baer, D. M. (1978). Multiple-probe technique: A variation of the multiple baseline. *Journal of Applied Behavior Analysis,* **11,** 189–196.

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.

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin,* **86,** 1236–1254.

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

Mackay, H. A. (1991). Stimulus equivalence: Implications for the development of adaptive behavior. In R. Remington (Ed.), *The challenge of severe mental handicap: An applied behaviour analytic approach* (pp. 235–259). London: Wiley.

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.

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. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. Lejeune (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 93–114). Hillsdale, NJ: Erlbaum.

Sidman, M., & Cresson, O., Jr. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency,* **77,** 515–523.

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior,* **43,** 21–42.

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.

Singh, N. N., & Singh, J. (1986). Reading acquisition and remediation in the mentally retarded. In N. R. Ellis & N. W. Bray (Eds.), *International review of research in mental retardation* (Vol. 14, pp. 165–199). New York: Academic Press.

Singh, N. N., & Solman, R. T. (1990). A stimulus control analysis of the picture-word problem in children who are mentally retarded: The blocking effect. *Journal of Applied Behavior Analysis,* **23,** 525–532.

Stevens, K. B., Blackhurst, A. E., & Slaton, D. B. (1991). Teaching memorized spelling with a microcomputer: Time delay and computer-assisted instruction. *Journal of Applied Behavior Analysis,* **24,** 153–160.

Stokes, T., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis,* **10,** 349–367.

Stromer, R. (1991). Stimulus equivalence: Implications for teaching. In W. Ishaq (Ed.), *Human behavior in today's world* (pp. 109–122). New York: Praeger.

Stromer, R., & Mackay, H. A. (1992). Delayed constructed-response identity matching improves the spelling performances of students with mental retardation. *Journal of Behavioral Education,* **2,** 139–156.

Stromer, R., & Mackay, H. A. (in press). Delayed identity matching to complex samples: Teaching students with mental retardation spelling and the prerequisites for equivalence classes. *Research in Developmental Disabilities.*

Stromer, R., Mackay, H. A., & Stoddard, L. T. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education,* **2,** 225–256.

Stromer, R., & Stromer, J. B. (1990a). The formation of arbitrary stimulus classes in matching to complex samples. *The Psychological Record,* **40,** 51–66.

Stromer, R., & Stromer, J. B. (1990b). Matching to complex samples: Further study of arbitrary stimulus classes. *The Psychological Record,* **40,** 505–516.

Tierney, R. J., & Shanahan, T. (1991). Research on the reading-writing relationship: Interactions, transactions, and outcomes. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 246–280). White Plains, NY: Longman.

Received September 23, 1991
Initial editorial decision January 23, 1992
Revision received July 1, 1992
Final acceptance July 16, 1992
Action Editor, John Parrish