JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2011, **44,** 781–792 NUMBER 4 (WINTER 2011)

A COMPARISON OF TWO FLASH-CARD METHODS FOR IMPROVING SIGHT-WORD READING

SARA KUPZYK, EDWARD J. DALY, III, AND MELISSA N. ANDERSEN

UNIVERSITY OF NEBRASKA–LINCOLN

Flash cards have been shown to be useful for teaching sight-word reading. To date, the most effective flash-card instruction method is incremental rehearsal (IR). This method involves the instructor interspersing unknown stimulus items into the presentation of known stimulus items. In this study, we compared IR to a modified IR procedure—strategic incremental rehearsal (SIR)—to determine whether the effects of IR might be improved by incorporating variables likely to increase word acquisition. These included increased opportunities to respond to unknown stimuli, using learner responding as a basis for changing instructional items, and systematic prompting methods. An A-B-A-B design was used to compare the effects of IR and SIR for increasing sight-word reading with 4 elementary school students. Results indicated that students read more words correctly with SIR than with IR. In addition, similar patterns of responding were seen at a 2-week follow-up.

Key words: discrimination training, incremental rehearsal, modeling, prompt delay, sight words, stimulus control

Goldiamond and Dyrud (1966) proposed that reading could be productively investigated as a form of operant behavior. For example, with the use of differential reinforcement, word reading should come under stimulus control of the configuration of letters and spaces in a text. In fact, one of the most successful reading programs involves direct instruction (DI), which is based on principles of instructional design that are derived from a stimulus control paradigm (Adams & Carnine, 2003; Gersten, Carnine, & White, 1984). Although phonics is a vital part of any reading instruction program (National Reading Panel, 2000), some words (e.g., *through*) do not contain predictable grapheme–phoneme correspondence, making them difficult to decode based on phonics rules learned through DI programs. Each item of the response class that makes up what educators refer to as *sight words* contains a unique configuration of letters whose phonemic properties do not conform to conventional pronunciation rules (e.g., *the*), distinguishing it from all other reading words. Thus, word reading must come under the stimulus control of the entire word.

Flash cards are a convenient, simple, and popular format for presenting discrete stimulus items (e.g., sight words) during discrimination training. Sight words can be presented singly while the instructor delivers prompts, reinforcement, and corrective feedback. Each repetition of the three-term contingency (presentation of the flash card, response of the learner, and consequence delivered by the teacher) further strengthens future responding in the presence of the relevant antecedent (Catania, 2007). Use of the three-term contingency (otherwise referred to as the learning trial) to conceptualize the components of effective instruction has been helpful in explaining the differential effectiveness of various instructional procedures (Heward, 1994), with conditions that deliver more learning trials consistently leading to greater learning (Belfiore, Skinner, & Ferkis, 1995; Skinner, Fletcher, & Henington, 1996). Traditionally, this type of flash-card instruction draws items only from material that has not yet been learned (i.e., unknown stimulus items; MacQuarrie,

We thank Breanna Gal and April Turner for their assistance with data collection.

Correspondence concerning this article should be addressed to Edward J. Daly, III, University of Nebraska–Lincoln, 33 Teachers College, Lincoln, Nebraska 68588 (e-mail: edaly2@unl.edu).

doi: 10.1901/jaba.2011.44-781

Tucker, Burns, & Hartman, 2002; Nist & Joseph, 2008). However, some experimenters began to intersperse previously mastered items (i.e., known stimulus items) with unknown stimulus items (Neef, Iwata, & Page, 1977, 1980; Roberts, Turco, & Shapiro, 1991) and found improved responding relative to conditions that contained only unknown stimulus items. The investigators speculated that improvements were due to better attending (as a function of a richer schedule of reinforcement in the interspersal condition; Neef et al., 1977, 1980) or more appropriate instructional difficulty level (Roberts et al., 1991).

Recommendations for controlling the ratio of known to unknown material during interspersal teaching trials began to appear in the professional literature out of a concern for ensuring appropriate difficulty level (Gickling & Havertape, 1981; Gickling & Rosenfield, 1995; Shapiro, 2004), which has been shown to have positive effects on on-task behavior, task completion, and comprehension (Gickling & Armstrong, 1978). Studies have examined various ratios of known to unknown items for sight words and math computation problems to identify the optimal ratio for improving learning during flash-card instruction (Burns, 2007; Cates et al., 2003; Cooke, Guzaukas, Pressley, & Kerr, 1993; Cooke & Reichard, 1996; MacQuarrie et al., 2002; Nist & Joseph, 2008; Roberts & Shapiro, 1996; Roberts et al., 1991). Although the results appeared initially to be somewhat contradictory (e.g., compare Roberts et al., 1991, to Roberts & Shapiro, 1996), one finding that consistently emerged was that the condition associated with the largest number of response opportunities led to the greatest increases in newly acquired responses (Burns, 2004, 2007; Cates et al., 2003; MacQuarrie et al., 2002; Szadokierski & Burns, 2008).

Nist and Joseph (2008), for example, compared three flash-card methods: interspersal, incremental rehearsal (IR), and drill and practice. All conditions contained six unknown words, but the interspersal and IR conditions also contained three known words that were either interspersed with unknown words throughout the session (interspersal) or interspersed incrementally over the course of the session (IR). Results suggested that participants learned more total words in the IR condition than in the other two conditions. A critical difference among the conditions was the number of unknown-word presentations associated with each condition. Unknown words were presented nine times per session during both the interspersal and drill-and-practice conditions. However, unknown words were presented 9 to 45 times per session in the IR condition. Therefore, five of the six unknown words for each session were presented two to five times more often in the IR condition than in the other conditions. The frequency of stimulus presentations is obviously a critical property of any discrimination training program. However, the advantage of IR is diminished when learning per unit of time is considered. When Nist and Joseph divided the number of words learned by the session length, they found that the drill-and-practice condition (containing only known words) outperformed both conditions that contained interspersal procedures (interspersal and IR). The incremental rehearsal and interspersal conditions take longer to administer because they include both known and unknown items. It appears, however, that the response opportunities for unknown stimulus items per se account for more learning (Burns, 2007; MacQuarrie et al., 2002), making conditions that increase response opportunities (when all other things are equal) more efficient.

The method by which unknown stimulus items are interspersed with known stimulus items sets IR apart from other methods (including other interspersal techniques). An unknown item is presented first, followed by a known item. The unknown item is then presented again, followed by the previous known item and a new known item. This

sequence is repeated, with the instructor adding an additional known item each time (i.e., incrementally), until all known items are exhausted. Next, the instructor removes the last known item and moves the unknown item into the first known item position while he or she introduces a new unknown item. The same sequence is used for the second unknown item. This whole process is followed until new unknown items are presented with nine known items in a single session (MacQuarrie et al., 2002). Because unknown items become designated as known items within and across sessions, the learner benefits from a lot of practice with previously unknown items and new unknown items in each instructional session, which is perhaps the reason for its superiority to other popular flash-card methods (Daly, Hintze, & Hamler, 2000; MacQuarrie et al., 2002).

Despite this potential benefit of IR, it has some drawbacks. The approach sets an artificial constraint on the number of unknown words that can be presented in an instructional session if the session is to be kept to a reasonable length. Most often, only three to five unknown items are introduced in a session while the learner repeatedly responds to the presentation of up to three times more known items (Burns, 2007; MacQuarrie et al., 2002; Nist & Joseph, 2008). Cates et al. (2003) and Nist and Joseph (2008) showed that instructional time was longer with IR than with other procedures, which further compounds the problem of an artificial constraint on the number of stimulus items. Furthermore, although IR creates frequent response opportunities for unknown items, decisions about when a word becomes a known item have been based on arbitrary procedures rather than on the learner's responding to the stimulus item. That is, an unknown item was treated as unknown until it was presented with all of the known items, at which time it replaced one of the known items; thus, it became a known item regardless of whether the response was acquired. Discrimination training should proceed according to learner responding. Finally, although IR incorporates the three-term contingency, it lacks prompting strategies that have proven to be very effective at accelerating learning (Coleman-Martin & Wolff-Heller, 2004; Wolery, Bailey, & Sugai, 1988).

Enhancing IR with additional features might produce a number of benefits. For example, keeping the number of stimulus items presented to a reasonable number within a session by removing known items may produce more and faster learning as well as make it more feasible for classroom use. Also, the effectiveness and efficiency of IR might be increased if prompting strategies (e.g., modeling, delayed prompting) were presented with unknown stimulus items to increase the pace of instruction, to reduce the number of stimulus presentations necessary for a given stimulus item, and to increase the rate of reinforcement for responding (because a response is more likely in the presence of the antecedent). In addition, IR might be more effective and efficient if items were added based on learner responding rather than on a prescribed number of stimulus presentations.

The purpose of the current study was to compare the effects of IR and a revised version of IR. Because IR has been consistently demonstrated to be superior to other flash-card methods, we specifically wanted to determine whether modest modifications of IR would further improve its effects. We refer to the revised version of IR as strategic incremental rehearsal (SIR) because it involved (a) the addition of an antecedent prompt (initially) with prompt fading, (b) use of unknown items only, and (c) the change of stimulus items based on learner proficiency rather than on the order of presentation (as in IR).

METHOD

Participants and Setting

Participants were four first-grade students (ages 6 to 7 years). All participants attended the same public elementary school and were eligible

for the federal free or reduced-price lunch program. Participants did not receive special education services or supplementary reading instruction while they participated in the study. Sessions were conducted at a table in the school psychologist's office or in an empty classroom.

Materials

Sight words were drawn from the first- and second-grade level Dolche (http://www.k12reader.com/dolch-word-list/) and Fry (http://www.usu.edu/teachall/text/reading/Frylist.pdf) word lists. Words used during the screening (described below) were printed in black ink and arranged in four columns on a single sheet of paper (8.5 in. by 11 in.). Each page contained 64 to 68 words. Words used during sessions (identified for inclusion through prior screening) were printed individually in black ink on index cards (3 in. by 5 in.) in landscape orientation (one word per index card).

Dependent Variable

The dependent variable was the total number of correctly read words (CRW) presented during an assessment session. Words were scored as correct if the participant correctly read the word within 3 s of presentation. If the participant corrected an incorrect response within 3 s, it was scored as correct. Words were scored as incorrect if the participant did not pronounce the word correctly or did not respond within 3 s.

Interobserver Agreement and Treatment Integrity

All sessions were audiotaped to assess interobserver agreement and treatment integrity. A second independent observer listened to audiotapes of the assessment sessions and scored responses for the purposes of determining interobserver agreement. The observer was given the list of the words in the order presented during the assessment and marked the word as correctly or incorrectly read by the student. Agreements and disagreements between the independent observer and the experimenter were determined for the sessions that were sampled. An agreement was defined as both the experimenter and the observer marking a word as correctly or incorrectly read by the participant. Interobserver agreement was calculated by dividing the total number of agreements by the total number of agreements plus disagreements and multiplying by 100%. Interobserver agreement was assessed during 33% of sessions across all phases. The overall mean agreement was 98% (range, 67% to 100% across all sessions sampled) across participants (i.e., 95%, 97%, 100%, and 99% for Andre, Jake, Julia, and Victor, respectively). Interobserver agreement was low (67%) for one session because the effect of one discrepancy between scorers was magnified by the small number of items for that session (i.e., three). Otherwise, the next lowest percentage agreement for any single session was 92%.

During all conditions, the experimenter followed a condition-specific protocol that was organized as a checklist (available from the second author). An independent observer listened to 30% of the audiotaped sessions and scored treatment integrity using the checklist that corresponded to the condition in effect and the list of words presented in that session (the order of words was specified for IR). The observer recorded whether or not the experimenter correctly implemented each step (i.e., order of word presentation, modeling, delayed prompt, error correction). Many of the steps occurred multiple times; however, for it to be recorded as correct, it had to be correctly implemented by the experimenter every time. The percentage of steps correctly implemented was calculated by dividing the number of steps correctly implemented by the total number of steps for the condition. The mean treatment integrity was 99.4% (range, 90% to 100%) for SIR and 100% for all IR sessions.

Design

An A-B-A-B design (with counterbalancing to control for sequence effects) was used to compare the effectiveness of the flash-card methods. For two participants (Andre and Julia),

the instructional sequence was IR, SIR, IR, SIR. For the other two participants (Jake and Victor), the instructional sequence was SIR, IR, SIR, IR. Each phase included five instructional sessions that were conducted approximately 4 days per week.

Procedure

Screening. An initial screening was conducted to identify an initial set of stimulus items for inclusion in the study. Unknown words were needed for both flash-card conditions (IR and SIR). Known words were needed for the IR condition. Lists of sight words from first- and second-grade Dolche and Fry word lists were generated to create potential stimulus items. Sessions were conducted individually with each participant. An experimenter presented the word lists to the participant and instructed him or her to try to read each word on the list. Words read correctly within 3 s were identified as known items for possible inclusion in the IR condition. Words read incorrectly during the first session were re-presented. Words read incorrectly during the second session were selected for inclusion in the unknown word pool and printed on individual index cards. Intermittent praise was delivered throughout screening sessions for on-task behavior (e.g., following directions, effort), but no programmed consequences were delivered for correct or incorrect responding.

Ongoing screening also was conducted throughout the study to determine specific stimulus items to be used. Immediately prior to each instructional session, the experimenter randomly presented words from the pool of unknown words (as determined in the first and second screening sessions) to identify stimuli to be used in the instructional session (third screening session). A word was designated as unknown only after the participant failed to read the word correctly in all three screening sessions. Unknown words were then randomly assigned to flash-card conditions as the need for new unknown words arose. Three unknown words were identified for each instructional session in the IR condition, and 10 unknown words were identified for sessions in the SIR condition. A larger number of unknown words was needed for the SIR condition because words were added contingent on participant responding during instructional sessions (although all 10 words were not necessarily presented in SIR sessions).

Assessments. An assessment session was conducted immediately prior to each instructional session. All the words trained on the previous day were administered during the assessment. For example, all the words presented on Monday were assessed immediately prior to the instructional session on Tuesday. Two weeks after each participant's final instructional session, assessments were conducted to assess maintenance. All words trained in a particular instructional phase were administered in a single session. Two conditions and two phases per condition meant that there were four separate assessments during maintenance, one for all the words in each phase. Prior to assessments, the experimenter shuffled all the flash cards. During the session, the experimenter presented each word individually and prompted the participant to say the word aloud. Following the response, the experimenter immediately presented the next word. No reinforcement or feedback was provided during assessment.

Instruction. During IR sessions, all words were presented according to the procedure prescribed by MacQuarrie et al. (2002), which included three unknown words and nine known words (see condition description below). During SIR sessions, words were presented in a random order and were incrementally added contingent on correct responding, which resulted in more words presented during sessions in which participants displayed more correct responses. Pilot testing with the IR procedure revealed that sessions lasted approximately 8 min. Therefore, SIR sessions were fixed at 8 min and as many of the 10 available unknown

words were presented as possible, depending on responding. Sessions sometimes took slightly longer in both conditions if 8 min elapsed before the experimenter finished presenting all known words for the final unknown word of the session. An instructional session was not conducted on a particular day (e.g., a Friday) if no assessment could be conducted on the following day (e.g., Saturday). On those days, only an assessment was conducted.

Incremental rehearsal. Prior to the initial IR instructional session, three unknown words and nine known words were selected based on screening results. During each session, the experimenter began by presenting the first unknown word (U1), which was followed by presentation of the first known word (K1). Next, U1 was presented again, followed by K1 and then the second known word (K2). Next, U1 was presented again, followed by K1, K2, and the third known word (K3). The experimenter added an additional known word (K4, K5, etc.) while repeating the sequence until U1 had been presented with all the known words. Therefore, U1 was presented nine times before the other unknown words were presented at all.

After U1 had been presented with all nine known words, it was moved to the K1 position and K9 was removed from the sequence before repeating the same sequence with U2 (i.e., U2-K1, U2-K1-K2, etc.). After U2 had been presented with all known words (i.e., K1 to K9), it was moved to the K1 position, the previous U1 became K2, and K9 was removed from the sequence. The same procedure was followed for U3. Thus, there were always 10 words in the sequence. In subsequent sessions, previously taught unknown words remained in the sequence until they were removed as the ninth known word (i.e., K9).

During all instructional sessions and for all words (known and unknown), the experimenter presented the stimulus item and prompted the participant to say the word aloud if the participant did not read the word immediately. If the participant correctly read the word within 3 s, the experimenter provided praise (e.g., ''good!''). If the participant incorrectly read the word or did not respond within 3 s, the experimenter provided corrective feedback and error correction. That is, the experimenter said, ''No, the word is —.'' The participant was then prompted to read the word correctly by having him or her say the word. Participants always read the word correctly following this prompt.

Strategic incremental rehearsal. Prior to the initial SIR instructional session, 10 unknown words were selected from the pool of available words based on screening results. During the first instructional session, the experimenter simultaneously presented and modeled the correct pronunciation of U1 (i.e., ''The word is —'') and asked the participant to repeat the word. If the student did not respond correctly, the experimenter gave corrective feedback, repeated the word, and prompted the student to say it aloud by having him or her say the word (''No, the word is —, say —''). Next, U2 was presented using the same procedure. This procedure was repeated for U1 and U2 before a prompt-delay procedure was initiated. At the next presentation of U1, the experimenter modeled correct reading of the word if the student did not respond correctly within 2 s (prompt delay). If the student read the word correctly before 2 s elapsed, the experimenter moved on to the next stimulus item. Next, U2 was presented using the same prompt-delay procedure. When participants responded correctly to the presentation of both words without a prompt, the experimenter introduced a new unknown word (U3). Modeling, corrective feedback, and error correction were used when U3 was introduced for the first time. Then, U1 and U2 were presented in random order with the prompt delay, corrective feedback, and error correction. U1, U2, and U3 were then randomized, and the words were presented with the same procedure. The experimenter followed this procedure (prompt delay, corrective feedback, and error correction) until

the participant read all three words correctly without the prompt or error correction and then introduced a new word (U4). This sequence was followed until the session time expired.

Flash cards were shuffled each time the experimenter finished presenting all the words and before starting the sequence with another unknown word. Each time a new word was added, a modeling prompt was used the first time the word was presented. Prompt delay was used during all subsequent presentations of the word. This sequence was repeated with the experimenter adding words (from the available pool of 10 unknown words) according to student responding. Sessions lasted for 8 min or until the last word in the series was presented for the most recent unknown word.

RESULTS

Results for all four participants are presented in Figure 1. Both flash-card strategies resulted in steep increases in CRW for all participants. In all cases, however, participants read more words correctly in the SIR condition than in the IR condition. Andre read more words correctly in the SIR condition than in the IR condition across an equal number of sessions. In the final session of both phases of SIR, Andre read 20 and 22 words correctly, respectively. In the final session of both phases of IR, he read 14 and 15 words correctly. Jake also read more words correctly in the SIR condition than in the IR condition. In the final session of both phases of SIR, he read 16 and 21 words correctly. In the final session of both phases of the IR condition, he read 15 and 13 words correctly. The same findings were obtained for Julia, whose performance exceeded the others in the SIR condition. In the final session of both phases of SIR, Julia read 21 and 24 words correctly. Interestingly, her performance dropped slightly between the fourth (22 CRW) and fifth (21 CRW) assessment sessions for the first SIR phase. This finding was not replicated in the second SIR phase (i.e., her performance increased following every instructional session). In the final session of both phases of IR, she read 13 and 14 words correctly. Victor read 19 and 17 words correctly in the final session of both phases of SIR. Similar to Julia, Victor's performance dropped off in the final session of one of the SIR phases. In Victor's case, it was during the second phase. In the final session of both phases of IR, he read 15 words correctly.

The maintenance results indicate that all students retained words in both conditions; however, they read more words correctly in the SIR condition than in the IR condition. This result is shown in Figure 1. Jake had fewer maintenance sessions due to an unexpected departure from the school. The difference between conditions was only one CRW for Jake. All other participants displayed an average difference of five CRW or more. Furthermore, every SIR data point during maintenance exceeded the highest IR data point. These results are based on the total number of words learned over the course of the study. Participants learned more words in SIR and therefore retained more words 2 weeks later. However, large differences were not found for accuracy of responding when the percentage of correct responses after a word presentation was calculated for maintenance sessions. For Andre, average correct responding was 90% for IR and 85% for SIR. For Jake, correct responding was 93% for IR and 88% for SIR. For Julia, average correct responding was 80% for IR and 80.5% for SIR. For Victor, average correct responding was 87% for IR and 100% for SIR.

DISCUSSION

We compared the effects of two flash-card methods on the acquisition and maintenance of sight words in a sample of first-grade students. Holding instructional length constant across both conditions made it possible to evaluate outcomes based on a standard unit of instructional time while it allowed each procedure to determine the number of response opportunities presented to participants. Although participants learned to read previously unknown

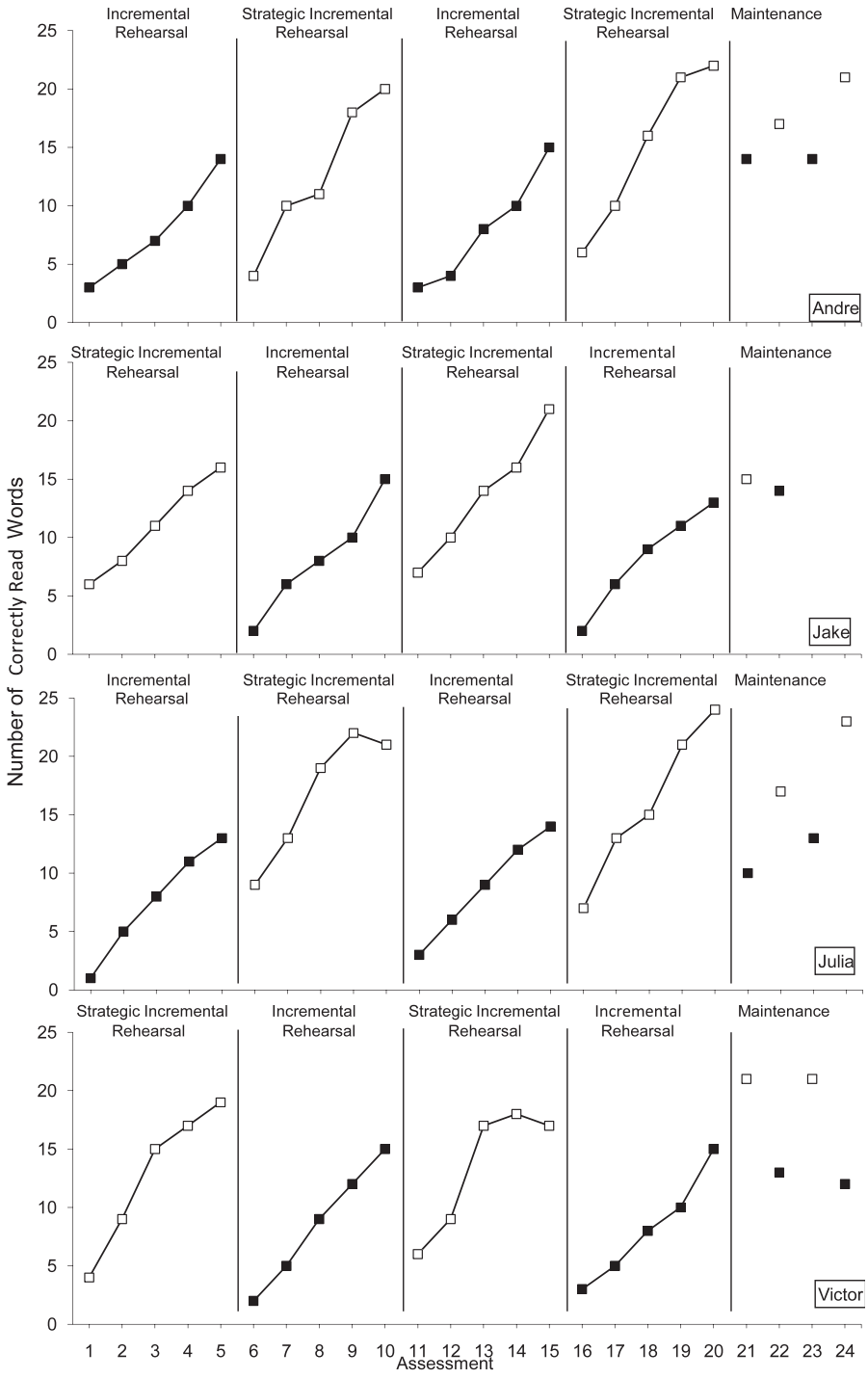

Figure 1. Total correctly read words in strategic incremental rehearsal and incremental rehearsal conditions across students.

words in both conditions, SIR resulted in (a) more rapid increases in responding, (b) higher response rates at the end of each phase of instruction, and (c) higher response rates during maintenance sessions. The findings are consistent with previous research suggesting that opportunities to respond, not the ratio of known to unknown items, account for the effectiveness of flash-card methods (Daly et al., 2000; Szadokierski & Burns, 2008). SIR allowed more response opportunities per time unit for unknown words in part because it eliminated unnecessary known items at the beginning of instruction (Cooke et al., 1993; Cooke & Reichard, 1996). The high ratio of known (90%) to unknown (10%) items used in IR, on the other hand, resulted in a greater amount of instructional time being spent practicing previously acquired words, thereby decreasing the efficiency of the method.

Strategic incremental rehearsal represents a procedural refinement of an already strong flash-card method. Several features of SIR permit more response opportunities (and presumably more learning) than IR. For example, although we recognize the importance of tailoring instructional material to appropriate difficulty levels for the learner (Daly, Martens, Kilmer, & Massie, 1996), controlling the ratio of known to unknown words (IR) is less effective than using a method that introduces instructional items based on prior student responding (SIR). The fact that SIR incorporates new items based on student responding (as opposed to a predetermined number of stimulus presentations) means that new instructional items can be introduced more quickly if the learner acquires prior items more rapidly. Furthermore, although fewer new stimulus items may be presented in a session for the learner who is acquiring material more slowly, stronger stimulus control and maintenance effects can be expected for SIR than for a condition that presents too few opportunities to respond. Therefore, systematic decision rules regarding introduction of new stimulus materials based on learner responding holds multiple advantages for the learner.

The use of a predetermined number of stimulus presentations creates yet another disadvantage for IR relative to SIR. To keep session length reasonable, an artificial ceiling typically is imposed by IR on the number of words a participant can learn per instructional session (three per instructional session as operationalized by MacQuarrie et al., 2002, and replicated in the current study). For SIR, the only factors that limit the number of words that can be presented in an instructional session are the session length and the learner's responding. SIR promotes rather than restricts response opportunities and is therefore more flexible than IR. This advantage also appears to have benefits for maintenance. Although the conditions produced approximately equivalent percentages of correct responding, SIR consistently produced a greater overall number of correct responses during the maintenance phase, in part because participants learned a greater number of words during SIR.

The SIR condition led to more new word presentations than did the IR condition. IR allowed a maximum of 15 word presentations for each phase for all participants (with the exception of Victor who, for the second phase of IR, received only 12 presentations). In the SIR condition, however, 21 (first phase) and 24 (second phase) new words were presented to Andre; 24 and 26 new words were presented to Julia; 20 and 21 new words were presented to Victor; and 17 and 42 new words were presented to Jake. It appears therefore that opportunities to respond (number of stimulus presentations) are likely to be the critical factor in the amount of learning produced (Heward, 1994; Skinner et al., 1996). Having increased opportunities to respond also creates more opportunities to deliver reinforcing consequences and other types of feedback (e.g., error correction).

This study also extends the research literature on flash-card instruction methods by including antecedent prompting strategies that presumably

speed up the transfer of stimulus control to the natural stimuli (i.e., letter configuration for word reading). SIR retains useful features of IR, because it provides a high rate of reinforcement, introduces one unknown word at a time (to control for task difficulty level), and allows repeated practice over time. However, by capitalizing on antecedent prompting, SIR strengthens opportunities to respond to unknown words by increasing the likelihood of a correct response so that it can then be reinforced. Antecedent prompts also reduce errors during initial stimulus presentations, which reduces the time needed to correct errors and increases reinforcement, which in turn increases the probability of a future correct response during subsequent stimulus presentations. Because of these features and the fact that students learned more words, SIR appears to be more efficient per unit of instructional time. SIR may be even more appealing to educators who work under serious time constraints because a session can be conducted in less than 10 min.

Results of this study suggest several areas for future research. First, it is not possible to determine whether the antecedent prompts, the systematic decision rule for introducing new items, the increased presentation of unknown items (as a result of eliminating known items), or an interaction effect among them contributed to the effectiveness of SIR. Further examination of the individual components of SIR would help to identify those that lead to the greatest amount of learning per unit of time. A component analysis of the contribution of the antecedent prompting procedure would be particularly interesting in light of the emphasis SIR placed on unknown words. It is also possible that SIR may be even more effective with further additions or modifications to the procedures. For example, other variables likely to affect the efficiency of flash-card instruction include the seating arrangement and type of error correction, both of which could be systematically examined in future research (Skinner, Belfiore, Mace, Williams-Wilson, & Johns, 1997; Van Houten & Rolider, 1989). Because the purpose of the study was to compare treatments, a separate baseline or an all-unknown condition was not included in the analysis. A baseline condition might have indicated how well the treatments did relative to no treatment. However, given the robust research that supports the efficacy of IR, additional conditions might have unnecessarily complicated the analysis. Furthermore, the screening method proved to be very effective for identifying words the participants were unable to read prior to instruction. Nonetheless, future studies might include a baseline condition to determine just how much learning occurred as a function of a flash-card instruction.

Instructional sessions in this study were carried out in a one-on-one format, as is common with flash-card instruction. Given that teachers usually have limited time to devote to one-on-one instruction, future research might examine whether SIR can be adapted to other instructional formats. For example, it may be possible to apply SIR to small-group instruction when combined with strategies that evoke unison responding from the whole group, as is done with response cards (Munro & Stephenson, 2009) and choral responding (Heward, 1994). Also, it may be worthwhile to examine whether students can learn to self-administer SIR or some variation of it as a peer-mediated intervention (Mayfield & Vollmer, 2007).

Future studies might also examine the generalizability of SIR to other academic areas such as math, spelling, alphabet naming, and vocabulary instruction. Cooke et al. (1993) found differences in effectiveness of flash-card methods depending on the academic domain. Therefore, it is premature to recommend generalized use of SIR for other academic skills until additional research has been conducted. It is also not clear whether response improvements using SIR will lead to better text reading, the natural context for responding. Although some studies have demonstrated that drill tasks can lead to generalized improvements in more advanced reading skills

(Nist & Joseph, 2008; Roberts & Shapiro, 1996; Roberts et al., 1991), other studies have demonstrated that teaching words in isolation does not lead to improved text reading (Daly & Martens, 1994; Skinner & Shapiro, 1989).

Although all participants acquired more words in SIR, the procedures could be strengthened by incorporating criteria for word removal based on correct responding. The absence of such criteria appeared to limit the number of words taught in later sessions more than might have been necessary. Participants demonstrated the largest gains in correct word reading after the initial sessions compared to later sessions within phases. In addition, Julia and Victor demonstrated a slight decrease in CRW during the final acquisition assessment in the first and second SIR phases, respectively. It is likely that the inclusion of a greater number of items in the instructional pool of words required more time to present the words in sequence (i.e., the most recent unknown item and all previous known items for the phase) and decreased opportunities to respond to new unknown stimuli. Therefore, future research might determine the number of opportunities needed for acquisition, maintenance, and generalization with SIR, which should help guide future revisions of the method to maximize the effects of later instructional sessions.

SIR appears to be an efficient and effective method for teaching sight words to elementary school students. The method allows for more words to be taught per time unit than the current benchmark method (IR) and uses student responding as the basis for instructional decision making and regulating difficulty level. Thus, it may be useful to educators looking for simple but effective instructional strategies as they strive to live up to mandates to use evidence-based practices in schools.

REFERENCES

Adams, G., & Carnine, D. (2003). Direct instruction. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 403–416). New York: Guilford.

Belfiore, P. J., Skinner, C. H., & Ferkis, M. (1995). Effects of response and trial repetition on sight-word training for students with learning disabilities. *Journal of Applied Behavior Analysis*, *28*, 347–348.

Burns, M. K. (2004). Empirical analysis of drill ratio research: Refining the instructional level for drill tasks. *Remedial and Special Education*, *25*, 167–175.

Burns, M. K. (2007). Comparison of opportunities to response within a drill model when rehearsing sight words with a child with mental retardation. *School Psychology Quarterly*, *22*, 250–263.

Catania, A. C. (2007). *Learning* (interim ed.). Cornwall-on-Hudson, NY: Sloan.

Cates, G., Skinner, C. H., Watson, S., Meadows, T., Weaver, A., & Jackson, B. (2003). Instructional effectiveness and instructional efficiency as considerations for data-based decision making: An evaluation of interspersing procedures. *School Psychology Review*, *32*, 601–616.

Coleman-Martin, M., & Wolff-Heller, K. (2004). Using a modified constant prompt-delay procedure to teach spelling to students with physical disabilities. *Journal of Applied Behavior Analysis*, *37*, 469–480.

Cooke, N., Guzaukas, R., Pressley, J., & Kerr, K. (1993). Effects of using a ratio of new items to review items during drill and practice: Three experiments. *Education and Treatment of Children*, *16*, 213–235.

Cooke, N., & Reichard, S. (1996). The effects of interspersal drill ratios on acquisition and generalization of multiplication and division facts. *Education and Treatment of Children*, *19*, 124–142.

Daly, E., Hintze, J., & Hamler, K. (2000). Improving practice by taking steps toward technological improvements in academic intervention in the new millennium. *Psychology in the Schools*, *37*, 61–72.

Daly, E. J., III, & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis*, *27*, 459–469.

Daly, E. J., III, Martens, B. K., Kilmer, A., & Massie, D. (1996). The effects of instructional match and content overlap on generalized reading performance. *Journal of Applied Behavior Analysis*, *29*, 507–518.

Gersten, R., Carnine, D., & White, W. (1984). The pursuit of clarity: Direct instruction and applied behavior analysis. In W. L. Heward, T. E. Heron, D. S. Hill, & J. Trapp-Porter (Eds.), *Focus on behavior analysis in education* (pp. 28–57). Columbus, OH: Merrill.

Gickling, E. E., & Armstrong, D. L. (1978). Levels of instructional difficulty as related to on-task behavior, task completion, and comprehension. *Journal of Learning Disabilities*, *11*, 32–39.

Gickling, E. E., & Havertape, S. (1981). *Curriculum-based assessment (CBA)*. Minneapolis, MN: School Psychology Inservice Training Network.

Gickling, E. E., & Rosenfield, S. (1995). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 3, pp. 587–596). Washington, DC: National Association of School Psychologists.

Goldiamond, I., & Dyrud, I. E. (1966). Some applications and implications of behavior analysis for psychotherapy. In J. Schlein (Ed.), *Research in psychotherapy* (Vol. 3, pp. 54–89). Washington, DC: American Psychological Association.

Heward, W. L. (1994). Three "low-tech" strategies for increasing the frequency of active student response during group instruction. In R. Gardner, III, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 283–320). Pacific Grove, CA: Brooks/Cole.

MacQuarrie, L., Tucker, J., Burns, M., & Hartman, B. (2002). Comparison of retention rates using traditional, drill sandwich, and incremental rehearsal flash card methods. *School Psychology Review*, *31*, 584–595.

Mayfield, K. H., & Vollmer, T. R. (2007). Teaching math skills to at-risk students using home-based peer tutoring. *Journal of Applied Behavior Analysis*, *40*, 223–237.

Munro, D. W., & Stephenson, J. (2009). The effects of response cards on student and teacher behavior during vocabulary instruction. *Journal of Applied Behavior Analysis*, *42*, 795–800.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute for Literacy.

Neef, N. A., Iwata, B. A., & Page, T. J. (1977). The effects of known-item interspersal on acquisition and retention of spelling and sight reading words. *Journal of Applied Behavior Analysis*, *10*, 738.

Neef, N., Iwata, B., & Page, T. (1980). The effects of interspersal training versus high-density reinforcement on spelling acquisition and retention. *Journal of Applied Behavior Analysis*, *13*, 153–158.

Nist, L., & Joseph, L. (2008). Effectiveness and efficiency of flashcard drill instructional methods on urban first-graders word recognition, acquisition, maintenance, and generalization. *School Psychology Review*, *37*, 264–308.

Roberts, M., & Shapiro, E. (1996). Effects of instructional ratios on students' reading performance in a regular education program. *Journal of School Psychology*, *34*, 73–91.

Roberts, M., Turco, T., & Shapiro, E. (1991). Differential effects of fixed instructional ratios on students' progress in reading. *Journal of Pscyhoeducational Assessment*, *9*, 308–318.

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York: Guilford.

Skinner, C., Belfiore, P., Mace, H., Williams-Wilson, S., & Johns, G. (1997). Altering response topography to increase response efficiency and learning rates. *School Psychology Quarterly*, *12*, 54–64.

Skinner, C. H., Fletcher, P. A., & Henington, C. (1996). Increasing learning rates by increasing student response rates: A summary of research. *School Psychology Quarterly*, *11*, 313–325.

Skinner, C. H., & Shapiro, E. S. (1989). A comparison of taped-words and drill interventions on reading fluency in adolescents with behavior disorders. *Education and Treatment of Children*, *12*, 123–133.

Szadokierski, I., & Burns, M. (2008). Analogue evaluation of the effects of opportunities to respond and ratios of known items within drill rehearsal of Esperanto words. *Journal of School Psychology*, *46*, 593–609.

Van Houten, R., & Rolider, A. (1989). An analysis of several variables influencing the efficacy of flash card instruction. *Journal of Applied Behavior Analysis*, *22*, 111–118.

Wolery, M., Bailey, D. B., Jr., & Sugai, G. M. (1988). *Effective teaching: Principles and procedures of applied behavior analysis with exceptional children*. Boston: Allyn & Bacon.

Received January 26, 2010
Final acceptance February 11, 2011
Action Editor, Claudia Dozier