JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2005, **38,** 485–497 NUMBER 4 (WINTER 2005)

REFINING THE EXPERIMENTAL ANALYSIS OF ACADEMIC SKILLS DEFICITS: PART I. AN INVESTIGATION OF VARIABLES THAT AFFECT GENERALIZED ORAL READING PERFORMANCE

EDWARD J. DALY, III

UNIVERSITY OF NEBRASKA–LINCOLN

CHRISTINE M. BONFIGLIO

ELKHART COMMUNITY SCHOOLS

AND

TARA MATTSON, MICHAEL PERSAMPIERI, AND KRISTIN FOREMAN-YATES

UNIVERSITY OF NEBRASKA–LINCOLN

Experimental analyses for improving reading fluency deficits have rarely targeted generalized increases in academic responding. As a consequence, the variables that may help students to generalize newly learned forms of academic responding like reading are not well understood. Furthermore, experimental analyses of reading fluency interventions have not systematically examined difficulty level as a variable that may affect instructional outcomes. The experiment reported in this paper expands (a) the measurement of the dependent variables to include generalized increases across tasks (reading passages) and (b) the combination of independent variables used to produce measurable generalized increases. The results demonstrate the importance of combining reward and instructional variables (including difficulty level) to produce generalized increases and how those variables can be meaningfully investigated prior to making treatment recommendations.

DESCRIPTORS: academic performance, experimental analysis, generalization, reading fluency

---

Millions of Americans have difficulty with common literacy tasks (Office of Technology Assessment, 1993). Reporting on the results of the National Assessment of Educational Progress, the National Center for Educational Statistics (2002) found that less than one third of a national sample of 7,900 fourth graders were reading at or above the "proficient" level in 2002. The results also indicated that whereas high-scoring students have improved on average, low-performing students' proficiency has worsened. It appears that the 100,000 research studies that have been published on reading since 1966 (National Reading Panel, 2000) may actually have had little real impact on the lives of children, especially the disadvantaged.

Behavior-analytic approaches to reading, such as direct instruction (Gersten, Carnine, & White, 1984) and precision teaching (Johnson & Layng, 1992), have achieved a measure of success with poor readers thanks to rigorous measurement and careful sequencing of stimulus conditions. To sequence stimulus conditions effectively requires knowledge of the basic principles that govern responding. Within a behavior-analytic paradigm, an oral reading response comes under the control of appropriate visual stimuli (i.e., the formation and configuration of letters and spaces in sequence). This stimulus control is brought about by differential reinforcement (Catania, 1998). With the student who is unable to read, the teacher uses antecedent strategies (e.g., modeling, prompting, choice of appropriate texts) to

Address correspondence to Edward J. Daly, III, University of Nebraska–Lincoln, Educational Psychology, 33 Teachers College, Lincoln, Nebraska 68588 (e-mail: edaly2@unl.edu).

doi: 10.1901/jaba.2005.113-04

establish new responses that can then be reinforced. Repeated readings (Rashotte & Torgenson, 1985), listening passage preview (Daly & Martens, 1994), and error correction (O'Shea, Munson, & O'Shea, 1984) are all examples of strategies that have been shown to improve reading. Ultimately, effective instruction increases the frequency of accurate and fluent reading. Reading responses (oral or covert) presumably compete more effectively with other responses (e.g., an incorrect word, looking away from the text; Wolery, Bailey, & Sugai, 1988), at which point the student is characterized as a fluent reader.

Students demonstrate that they are proficient with a word when they can read it in connected text. In text, however, a given word may be embedded in a host of different words, with grammatical conventions providing the only real constraints on the sequences in which words are likely to appear. Reading text is more difficult than reading words in isolation (Howell & Nolet, 2000), which is not surprising in light of the fact that each reading response must compete effectively irrespective of the preceding words and all of this must be coordinated rapidly for comprehension to occur. Generalized reading responses to texts and across texts is therefore the key to proficient reading. It is more likely to occur when instruction is conducted in texts (i.e., the criterion setting) than when it is carried out in isolation (e.g., words appear on word lists or flashcards; Daly & Martens, 1994). Nonetheless, the process by which students come to generalize reading and the variables that can be brought to bear on reading instruction to improve generalized reading are poorly understood (Lyon & Moats, 1997). Too few analytical investigations have directly targeted generalized responding.

One exception is a study conducted by Daly, Martens, Kilmer, and Massie (1996), in which generalized responding to novel texts was measured. Word overlap (between instruction and assessment) and difficulty level (i.e., degree of instructional match) were manipulated. The results indicated that the combination of instructional match and a high degree of word overlap between what was taught and what was tested led to greater generalized increases in reading accuracy and fluency. In other words, when instruction was carried out in instructional-level reading material (as opposed to more difficult reading material) and assessment reflected what was actually taught, reading improvements were observed. The students in the study, however, were very poor readers (i.e., reading between 8 and 52 correctly read words [CRW] per minute prior to the experiment). Daly et al. acknowledged that the participants were probably at a frustrational level even with the materials deemed to be an instructional match. It is unclear whether these results are applicable to students at other proficiency levels.

Although word overlap and instructional match are clearly important instructional variables, there are a multitude of other variables that can be examined for their potential ability to contribute to generalized increases in reading. It stands to reason that the right combination of instructional and reward variables might function to make reading responses compete more effectively in criterion settings (i.e., novel texts), yielding generalized responding. For example, if the student was highly motivated to read words correctly and fluently in the criterion setting, he or she might benefit more from prior instruction. To this end, rewards might be offered for reading as fluently as possible in novel texts. Indeed, contingent reinforcement (Lovitt, Eaton, Kirkwood, & Pelander, 1971) has been shown to improve reading. Measured increases in reading due to positive reinforcement are probably an indication of the degree to which reinforcement strengthened already acquired (but not yet fluent) responses. However, no amount of motivation alone can help a student to read words that he or she has not yet acquired

(Haring, Lovitt, Eaton, & Hansen, 1978). Therefore, if even greater increases are achieved when prior instruction precedes opportunity for reward than when no instruction precedes opportunity for reward, one can infer that the combination of instruction and reward led to generalized increases. Effective instruction should enable the reader to produce accurate responses that can then be strengthened through reinforcement. The case for generalization would be clearer if the instruction was carried out in one or more passages that differed from the criterion texts to which contingencies were applied and in which outcomes were measured.

The purpose of this experiment was twofold. First, this study sought to examine the generality of Daly et al.'s (1996) findings with respect to difficulty level and its effect on students' ability to generalize reading responses. There were two levels of text difficulty—easy and hard. Participants, however, were more proficient than the participants in Daly et al.'s study, allowing an investigation of the effects of difficulty level at different parameters. Second, this study sought to examine whether instruction and reward could be combined to produce generalized increases in reading fluency. In half the sessions, instruction was carried out on passages that contained approximately 80% of the same words as the passages in which measurements were taken (generalization passages). In the other half, no instruction was carried out prior to reading in generalization passages. There were four conditions in all: hard/no instruction (H/NI), hard/instruction (H/I), easy/no instruction (E/NI), and easy/instruction (E/I). Reward and word overlap were held constant across all conditions. Therefore, the study was designed to provide an analysis of the interactive effects of three broad classes of variables—difficulty level, instruction, and reward. It was expected that the combination of instructional match (i.e., easier rather than harder) and systematic instruction would produce a greater degree of generalization than the absence of instructional match (i.e., harder rather than easier) and systematic instruction when reward for achieving a predetermined performance criterion across conditions was held constant.

METHOD

Participants and Setting

Participants in this project were 3 boys who had been referred by their teachers for reading problems. Foster and Dustin were in the third grade, and Austin was in the fourth grade. The students ranged in age from 9 years 6 months to 11 years 8 months. All students attended a public elementary school where data were collected. Students were receiving special education services for students with learning disabilities. Parent consent and child assent were obtained before participation. Experimental sessions were conducted in the school psychologist's office (3 m by 3 m) located in the main office of the school.

Materials

Instructional passages. Thirteen second-grade and 13 third-grade reading passages were chosen randomly from the Silver, Burdett, and Ginn basal reading series (Pearson et al., 1989). Only narrative and expository texts were used. Passages contained approximately 170 words. Readability scores for the passages were computed using the Spache formula (Spache, 1953). The average readability of the second-grade passages was 2.3 (range, 2.0 to 2.6), and the average readability of the third-grade passages was 3.4 (range, 3.0 to 3.9).

Passages with high content overlap (HCO). HCO passages were the generalization passages that contained a high percentage of the same words in a corresponding instructional passage (Daly et al., 1996). The HCO passages were created by rewriting the original passages, using the majority of words from the original instructional story. Therefore, there was one HCO passage for each instructional passage.

The average readability for the second-grade passages was 2.9 (range, 2.3 to 3.9), and the average readability for the third-grade passages was 3.2 (range, 2.3 to 4.0).

The percentage of word overlap with instructional passages was calculated by dividing the number of words that appeared in both passages by the total number of words in the HCO passage. The mean word overlap was 87% (range, 80% to 95%) for all passages.

Dependent Variables

CRW and errors per minute in the HCO passages were the dependent variables. Students were instructed to read aloud to the examiner who followed along on a separate copy of the story. As the student read a passage, the examiner recorded CRW and errors. A correctly read word was defined as a word that is pronounced correctly in 3 s (Shinn, 1989). An error was recorded if the student mispronounced a word, omitted a word, substituted another word, or did not read a word within 3 s (Shinn). Each student read the passage for 1 min. The number of CRW and errors were calculated for each passage. An audiocassette recorder was used during all sessions to assess interobserver agreement.

Independent Variables and Treatment Conditions

Three variables were manipulated in the study. Procedural details can be found in the next section. What follows is a brief description of each of the variables.

Reward. Rewards were available during all HCO sessions. Participants chose a tangible item from a variety of rewards (e.g., pocket games, pencils, and baseball cards) and were told that they could earn the reward if they met a predetermined criterion for fluency (expressed as CRW per minute) and accuracy (expressed as number of errors) on the HCO passage. The criterion was set based on preexperimental screening results. To earn a reward, the student had to read at least one more word correctly and make three or fewer errors. Students were told that they "could earn the reward if they beat their last score while making no more than three errors."

Instruction. Instruction consisted of listening passage preview (Daly & Martens, 1994), repeated readings (Rashotte & Torgenson, 1985), phrase drill error correction (O'Shea et al., 1984), and a syllable segmentation and blending lesson (when errors were repeated after the second student reading of the passage). Listening passage preview provided modeling of accurate and fluent reading in the instructional passage. Repeated readings increased students' opportunities to respond by having them read a passage twice. Error correction provided corrective feedback and practice on error words by having the student reread phrases containing error words three times after the experimenter modeled correct reading. The syllable segmentation and blending procedure provided additional practice with words that the student (a) read incorrectly during the first student reading, (b) were corrected as a part of the phrase drill error correction procedure, and (c) read incorrectly on the second reading of the passage. Words were broken into individual syllables while the experimenter and then the student read the syllables in order and together as a word. No rewards were offered or delivered during instruction.

Difficulty level. Identification of hard and easy passages was based on preexperimental screening results (described below). Hard passages were the eight most difficult HCO passages for each participant, and easy passages were the eight least difficult HCO passages for each participant.

Experimental Design and Procedure

An alternating treatments design was used to compare the effects of reward only with instruction-plus-reward conditions in easy and difficult HCO passages (Kazdin, 1982; Sindelar, Rosenberg, & Wilson, 1985). Sessions were conducted once daily by a trained doctoral-level school psychologist (second author) for 4

Table 1
Preexperimental Correctly Read Words per Minute and Error Rates (in Parentheses) in HCO Passages for Each Participant

Participant	Condition			
	E/NI	E/I	H/NI	H/I
Foster				
M	46.75 (6.25)	49.75 (6.25)	29.75 (7.25)	30.75 (6.75)
SD	4.86 (2.63)	4.57 (2.22)	5.56 (2.22)	2.22 (2.22)
Range	44–54 (4–9)	45–55 (4–9)	22–35 (4–9)	28–33 (4–9)
Dustin				
M	52.5 (6.75)	53.75 (5.5)	34.25 (9)	31.75 (10.25)
SD	4.20 (1.71)	6.13 (1.73)	2.87 (2.83)	5.5 (3.59)
Range	47–57 (5–9)	48–60 (3–7)	31–38 (7–13)	27–37 (7–15)
Austin				
M	55.5 (7.25)	52 (6.75)	31 (9.5)	34.5 (7.75)
SD	7.59 (1.26)	7.16 (2.22)	2.16 (3.11)	2.89 (2.75)
Range	48–63 (6–9)	45–62 (4–9)	28–33 (7–14)	31–38 (5–11)

consecutive days each week. All sessions were counterbalanced semirandomly so that all conditions occurred once every four sessions.

Preexperimental screening and assignment of passages to conditions. Pretreatment screening was conducted to identify hard and easy passages in the following manner. Students read all 26 HCO passages (13 second-grade and 13 third-grade passages) aloud for 1 min in several sessions within a 4-day span as a part of an initial screening. Results were scored as CRW per minute. Based on the results of the initial screening, passages were divided on an individual basis into easy and hard categories according to the number of CRW per minute. The eight passages with the lowest CRW per minute were assigned to the hard condition, and the eight passages with the highest CRW per minute were assigned to the easy condition. Passages were assigned randomly to I and NI conditions within their respective difficulty levels. Average preexperimental fluency and error rates in HCO passages for each participant are shown in Table 1. The participants were also screened for responsiveness to reinforcement criteria. This was done by asking each participant to reread two different passages from the same reading series while the reward contingency was applied. If a student failed to increase his performance by 20%, he would have been excluded from the study. All of the students, however, met the criterion for performance increase under the reward contingency.

Experimental procedure. Sessions were conducted 4 days per week until each condition had been administered four times to all participants. Each session lasted approximately 10 to 25 min. Only one condition was administered per day. In all sessions, the instructional passage was administered before the HCO passage. At the beginning of each session, the experimenter first explained to the child that he would have an opportunity to earn a reward for reading later in the session. The student chose a reward, and it was displayed on the desk in full view. For no-instruction sessions (H/NI and E/NI), the student was then told to read the instructional passage aloud in its entirety once. If the child did not respond to a word within 3 s, the experimenter said the word to the child. The experimenter told the student how accurately and fluently he read the passage (e.g., "You read 63 words with 4 errors in 1 minute"). For instructional sessions (H/I and E/I), the experimenter first read the passage aloud to the student at a comfortable reading rate and then had the student read the instructional passage aloud in its entirety twice. After the first reading, however, the examiner had the student practice reading phrases containing error words three times (phrase drill) before rereading the passage one more time. If the participant repeated an error, the experimenter then did the syllable segmentation and blending lesson. The experimenter covered all but the first syllable of the word in the passage, read the syllable to the student, and prompted the student to read the syllable. The experimenter continued to expose each syllable sequentially and followed the same procedure of modeling and prompting correct responses. When all of the error words had been read in this fashion, the experimenter started again with the first error word and had the student reread each word in syllables and then as a single word without modeling. If the student answered incorrectly, the experimenter corrected him and had him repeat the response.

When the student was done with the instructional portion of the session, the experimenter told the student that he could now earn a reward for reading the next passage at the predetermined fluency and accuracy levels. The student then read the HCO passage for 1 min. If the student met the criterion, the experimenter praised him and gave him the reward. If the student did not meet the criterion, the experimenter made an encouraging statement ("nice try") and told the child that he could try to earn the reward the next time they read together. The student was not allowed to take the reward.

Interobserver Agreement

Experimenters audiotaped all sessions to score student performance. An independent observer scored 50% of sessions to assess interobserver agreement. Interobserver agreement was computed by dividing the number of agreements (i.e., on words read correctly and words read incorrectly) on a word-by-word basis by the total number of words. The mean interobserver agreement was 99.7% (range, 97% to 100%).

Treatment Integrity

Experimenters followed a protocol in the form of a checklist (available from the first author) for each condition to prompt them to follow each step in the correct order. Verbal instructions and feedback were scripted. All sessions were audiotaped. An independent observer assessed treatment integrity on 50% of the experimental sessions. Using the same checklist, the independent observer scored the checklist while listening to the audiotaped session. The mean treatment integrity was 99% (range, 96% to 100%) across all participants.

RESULTS

CRW per minute for all participants are displayed in Figure 1. Clearly discernible differences between instruction (E/I and H/I) and control (E/NI and H/NI) conditions would be interpreted as indicating an experimental effect. If the difference between conditions was larger for one difficulty level (e.g., E/I vs. E/NI) than for the other difficulty level (H/I vs. H/NI), then results would be suggestive of an interaction effect; that is, the experimental effect is moderated by difficulty level. Also, differences between difficulty levels (E/I vs. H/I) can be examined to identify the difficulty level in which performance was most improved. Interpretation of the strength of treatment (i.e., the difference between effects in E/I and H/I) based on differences between the treatment conditions alone, however, would need to be qualified by existing preexperimental treatment differences. Means and standard deviations for each condition can be found in Table 2.

Foster displayed the highest level of responding in the E/I condition for CRW per minute. Results were superior to all but one of the E/NI sessions (see the third E/NI data point). Differences between H/I and H/NI were not as evident. Data series were overlapping for the hard passages across conditions (Figure 1). Mean differences between conditions (Table 2)

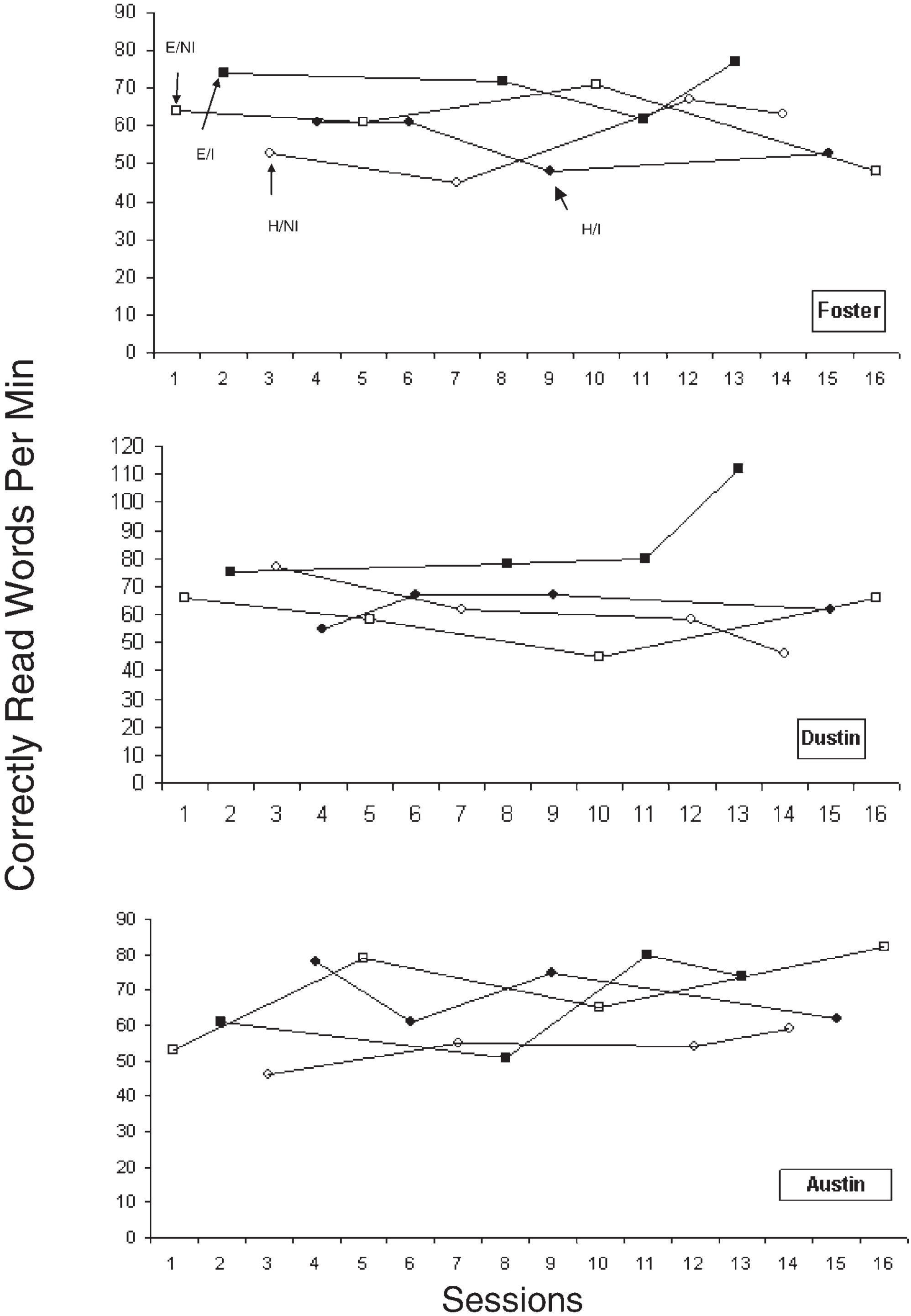

Figure 1. The number of CRW per minute for Foster, Dustin, and Austin across all experimental conditions in HCO passages.

Table 2
Correctly Read Words per Minute and Error Rates (in Parentheses) in HCO Passages for Each Participant

	Condition			
Participant	E/NI	E/I	H/NI	H/I
Foster				
M	61 (3.5)	71.25 (1.25)	57 (4.75)	55.75 (2)
SD	9.63 (1.29)	6.50 (1.5)	9.93 (1.71)	6.40 (1.41)
Range	48–71 (2–5)	62–77 (0–3)	45–67 (3–7)	48–61 (1–4)
Dustin				
M	58.75 (4)	86.25 (0.75)	60.75 (4)	62.75 (2.75)
SD	9.91 (3.37)	17.29 (0.5)	12.79 (4.08)	5.68 (1.71)
Range	45–66 (0–8)	75–112 (0–1)	46–77 (1–10)	55–67 (1–5)
Austin				
M	69.75 (5)	66.5 (2)	53.5 (3.5)	69 (2.75)
SD	13.40 (2.71)	13.02 (0.81)	5.45 (2.08)	8.76 (0.5)
Range	53–82 (3–9)	51–80 (1–3)	46–59 (1–6)	61–78 (2–3)

also indicate the same interaction effect. Therefore, greater increases in generalized reading fluency (when controlling for reward) were obtained in easier passages than in harder passages. Prior to experimental sessions, Foster was reading just under 50 CRW per minute in the easy passages and approximately 30 CRW per minute in the hard passages (Table 1).

The pattern of results for CRW per minute was very similar for Dustin. However, differences between E/I and E/NI were even more pronounced. There are no overlapping data points for these two conditions. On the other hand, as for Foster, the data series for H/I and H/NI were overlapping. Interestingly, one of the data points for H/NI fell within the range of data points for E/I. The largest mean differences were in the same direction (Table 2), favoring E/I relative to the H/I condition. Dustin's preexperimental fluency levels were similar to Foster's. Dustin was reading just over 50 CRW per minute in the easy passages and just over 30 CRW per minute in the hard passages (Table 1).

For Austin, the opposite pattern was found for CRW per minute. Clearly discriminable data series are evident for the H/I condition relative to the H/NI condition, whereas data series for E/I and E/NI are overlapping. There are two interesting aspects to this finding. First, Austin was reading at approximately the same levels as Foster and Dustin for easy and hard passages prior to experimental sessions (approximately 55 CRW per minute in easy materials and just over 30 CRW per minute in hard materials; Table 1). Therefore, preexperimental fluency differences do not account for the opposite interaction effects across students. The other finding that is noteworthy is that Austin achieved the same level of reading fluency following instruction in harder materials as he achieved in easier materials. H/I data points overlap entirely with both E/I and E/NI data points. Therefore, Austin actually displayed greater generalized performance increases in harder passages than in easier materials.

To give a complete account of the effects of treatment on generalized reading fluency, it is necessary to consider not only the absolute performance levels but also the relative performance increases from the preexperimental screening to experimental conditions. This can be done by comparing the results in Table 2 to those in Table 1. Increases in all conditions from preexperimental screening to the experimental analysis results are evident from an inspection of mean differences across the two tables. This finding applies to the NI conditions as well as the I conditions. Differences in NI conditions between the preexperimental screening and the experimental analysis results may be a result of maturation or effects of the contingency applied to these passages, increasing the importance of their function as a standard of comparison as control passages.

Although all participants were reading at about the same fluency levels prior to the experimental analysis, there were individual differences in just how much relative growth was observed and which conditions were associated with the greatest changes. Comparison of means between experimental analysis results and preexperimental screening shed some light on these differences. Patterns of absolute differences across conditions (e.g., E/I being superior for Foster) did not necessarily correlate with relative performance changes across conditions for each participant. For Foster, larger relative changes in CRW per minute were found for the hard conditions (H/NI = 27.25 CRW per minute; H/I = 25 CRW per minute) than for the easy conditions (E/NI = 14.25 CRW per minute; E/I = 21.5 CRW per minute), even though the greatest absolute change for Foster was in the E/I condition relative to all the other conditions (described earlier). For Dustin, large relative changes in CRW per minute were also obtained with the hard conditions (H/NI = 26.5 CRW per minute; H/I = 31 CRW per minute), but were eclipsed by an even larger relative change in the E/I condition (32.5 CRW per minute). The smallest change was in the E/NI condition (6.25 CRW per minute). Similarly, for Austin, larger effects were observed for the hard conditions, with H/I producing the largest relative performance increase (H/NI = 22.5 CRW per minute; H/I = 34.5 CRW per minute) and smaller relative changes in performance were observed in the easy conditions (E/NI = 14.25 CRW per minute; E/I = 14.5 CRW per minute).

These findings attest to the importance of attending to both relative and absolute performance increases in evaluating performance changes for a skilled behavior like oral reading fluency as a function of experimental manipulations.

Finally, errors were reduced in all of the instruction conditions relative to the no-instruction conditions and preexperimental screening results (Table 2), perhaps largely as a function of repeated exposure (from preexperimental screening to experimental analysis) and the availability of a contingency that specified a criterion for accuracy as well as fluency.

DISCUSSION

One purpose of the study was to examine whether instructional and reward variables could be combined to produce generalized increases in reading fluency, an uncommon finding in the literature. Based on differences between conditions with no prior instruction and those with prior instruction (with difficulty level controlled), the results indicate that generalization was obtained for all participants. Therefore, it appears that the combination of instructional and reward variables (compared to reward variables only) was helpful to facilitating generalized increases. These findings should contribute to our understanding of how to produce generalization of responding for a skilled behavior that is under the control of textual stimuli. This study serves to broaden the scope and configuration of variables that have been examined to produce generalized increases in reading fluency.

The strategy involved a contingency in the natural setting (i.e., HCO passages) that was preceded by instruction that established new responses in the students' repertoires. Prior instruction involved programming common stimuli (Stokes & Baer, 1977), which permitted the contingency applied to the HCO passages to strengthen the newly acquired responses. Those responses were brought under stimulus control and therefore were able to compete when students were reading the generalization (HCO) texts when their motivation to read the words correctly and rapidly was presumably high. It is important to note, however, that under the current experimental arrangement no programmed reinforcement was provided in the instructional passage. Therefore, the influence of reinforcement as a variable occurred during responding in the HCO passages in conjunction with antecedent instruction only.

It is possible that future researchers could choose different combinations of instructional and reward variables to produce even greater increases in responding in criterion settings. In the future, investigators should look more extensively at whether generalization is observed when responding is introduced to natural maintaining contingencies or when the contingencies for responding are indiscriminable (Stokes & Baer, 1977). Investigators might examine whether the same effects are achieved in classrooms and with various contingencies in place (e.g., praise, goal setting, privileges in the classroom).

To date, very few studies have carefully operationalized the conditions under which generalization in reading passages is measured. The study reported here presents one method for doing just that. Passages were rewritten, controlling the amount of word overlap between instructional and criterion settings. In this way, criterion settings still reflected the way in which words may appear in different sequences across passages (generalization) while a reasonable level of sensitivity to potential treatment effects was maintained. The HCO passages represent one form of generalization related to the degree to which textual stimuli control responding. In this case, textual control is reduced to the degree of word overlap across passages. Therefore, conclusions of this study are limited to this form of generalized responding only. Other forms of generalization are probably important as well. For instance, a learner could begin to generalize fluent reading to nontaught but equally difficult words, a scenario in which the correspondence of training to nontraining stimuli is even more discrepant (and hence constitutes an even greater degree of generalization). There may also be other dimensions of the stimuli (e.g., linguistic structures) that affect responding in other ways, representing possible dimensions of yet other forms of generalization of responding that could be measured. Therefore, in the future investigators may want to devise other measures of generalization for experimental analyses of academic interventions based on theoretical or empirical considerations. The findings of this and future investigations that target different dimensions of generalization will probably lead to the development of stronger and more robust treatments for students who experience chronic academic performance problems.

The other purpose of the study was to determine whether difficulty level could moderate the effects of a strong reading intervention. Relative difficulty level was determined through a preexperimental screening with passages that were at or close to the level at which classroom teachers were carrying out instruction. The results indicated that there were interaction effects for all 3 participants. Interpretation is complex because interaction effects can be examined in two different ways—either in an absolute sense or a relative sense. With respect to absolute differences across difficulty levels, for Foster and Dustin E/I led to higher levels of responding relative to its control condition (E/NI) than did H/I relative to its control condition (H/NI). The opposite was found for Austin, who displayed a larger difference between treatment (I) and control (NI) for harder than for easier passages. This finding is intriguing, especially because students had the same preexperimental reading fluency levels, and suggests that instructional match may need to be investigated on a case-by-case (i.e., idiographic) basis through an experimental analysis. On the other hand, comparisons to preexperimental screening results for each instruction condition suggested that larger relative improvements were observed overall in the harder materials for Foster and Austin. Although the relative difference between E/I and H/I conditions favored the E/I condition, the discrepancy was very small (i.e., amounting to a difference of 1.5 CRW per minute). The comparison of absolute and relative performance increases further reinforces the importance of examining individual differences, an issue dealt with in more detail in the next paper in this series. At the very least, results indicate that not all students will respond in the same way to recommendations for instructional levels that are commonly found in the literature.

The issue of instructional level has received a fair amount of attention (if not experimental analysis) in the academic intervention literature. Recommendations for students' instructional levels based on measured accuracy and fluency abound (Fuchs & Deno, 1982; Gickling & Rosenfield, 1995; Good, Simmons, & Kame'enui, 2001; Howell & Nolet, 2000; Shapiro, 2004). The premise of these recommendations is that a student's learning will progress more rapidly if he or she is taught at the appropriate instructional level (as opposed to being taught at the frustrational or mastery level). Instructional levels are described as ranges of responding (e.g., 40 to 60 CRW per minute), whereas mastery and frustrational levels are described as cut-offs (e.g., mastery is greater than 60 CRW per minute). There are some data in the literature to support this notion. For instance, Gickling and Armstrong (1978) obtained the highest rates of on task, task completion, and task comprehension for instruction carried out in the instructional level when compared to the frustrational and independent levels. As noted earlier, Daly et al. (1996) obtained a greater degree of generalization in reading texts when passages were closer to an appropriate instructional level than when they were more difficult.

One problem with current practice is that the actual recommendations for instructional levels can vary widely. For example, whereas Fuchs and Deno (1982; echoed by Shapiro, 2004) recommend 40 to 60 CRW per minute as an appropriate instructional level at the second grade, Howell and Nolet (2000) recommend 70 to 100 CRW per minute for the same grade. Consequently, educators may be in a quandary as to what the appropriate instructional placement might be for each student. The concept of instructional level as interpreted by educators also will probably affect decisions regarding effectiveness of an intervention. By implication, the determination of effectiveness of a reading intervention with a second-grade student who achieves 61 CRW per minute will be different depending on whether Fuchs and Deno's (1982) or Howell and Nolet's (2000) recommendations are consulted. One is encouraged in the first case because the student's proficiency has risen above the instructional level. In the second case, one is discouraged to find out that the student has not yet even reached the instructional level and should have been placed lower in the curriculum in the first place.

Published instructional placement recommendations appear to be based on authority in some instances and on normative estimates for the population in question in other instances. However, normative estimates (e.g., the average second-grade reader's fluency) may not be the strongest basis for establishing the level at which instruction can have the largest effect on learning for an individual student. The real question is the level of proficiency with instructional tasks that is optimal for achieving generalized responding. When instructional material is neither too hard nor too easy, good instruction will presumably be more effective at helping the student to build stronger response repertoires that can generalize more readily to novel material, be it material at the same difficulty level or at higher difficulty levels. Cast in this light, instructional level is a construct that needs to be investigated experimentally and not just based on average proficiency levels (unless experimental analyses substantiate the use of average proficiency levels—a type of nomothetic approach to the question—for determining appropriate instructional match). Moreover, the instruction that is used is likely to interact with the difficulty level of the material. It is possible that stronger instruction in more difficult material can achieve the same or a greater effect as weaker instruction in not-too-difficult material, creating difficulties in operationalizing and interpreting instructional match.

The results of the study reported here suggest that it may be time to reframe the issues of instructional match, difficulty level, and recommended instructional levels. The results suggested that difficulty level may mediate the effects of instructional and reward variables, speaking to the importance of this variable in producing generalized reading fluency. The results also indicate, however, that the effects may be idiosyncratic and that under some conditions greater relative improvements can be achieved in harder materials than in easier materials. Students with the same preexperimental reading levels may respond differentially to harder or easier instructional passages, necessitating an individualized analysis of the effects of difficulty level when appropriate instructional placement is sought. Unfortunately, the current study is unable to answer the broader question of the optimal level of proficiency necessary for students to benefit most from instruction. Difficulty level was categorized dichotomously (hard vs. easy). Future investigations could examine various parameters of responding to determine whether a nomothetic or an idiographic approach is the most effective and efficient way to identify the instructional level at which a student's progress will be accelerated. The finding of idiosyncratic effects across participants strongly suggests that an experimental analysis is probably the best way to proceed. The findings of the current study also strongly suggest the need for examining both relative and absolute performance increases.

Interpretation of the results must be tempered in light of limitations of the current study. First, instructional sessions were carried out individually and with strong instructional and reward variables that are known to improve reading fluency. The results may not represent what might be achieved at similar fluency levels if weaker instructional or reward variables are used or in the context of group instruction, which is more common in the classroom setting. Future investigations should examine instructional effects in different grouping formats (e.g., small-group reading instruction) and in more naturalistic classroom conditions (e.g., with teachers carrying out at least some part of the analysis). At the very least, the methods used in the current study should be further validated by examining the effects of recommended treatment packages in classrooms over time. These limitations notwithstanding, the results have direct implications for conducting brief experimental analyses to identify potentially effective reading interventions.

REFERENCES

Catania, A. C. (1998). *Learning* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

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.

Fuchs, L. S., & Deno, S. L. (1982). *Developing goals and objectives for educational programs.* Minneapolis: Institute for Research in Learning Disabilities, University of Minnesota.

Gersten, R., Carnine, D., & White, W. A. T. (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. 38–57). 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., & 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.

Good, R. H., III, Simmons, D. C., & Kame'enui, E. J. (2001). The importance of decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, *5*, 257–288.

Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). *The fourth R: Research in the classroom.* Columbus, OH: Merrill.

Howell, K. W., & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making* (3rd ed.). Belmont, CA: Wadsworth.

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist*, *47*, 1475–1490.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford.

Lovitt, T. C., Eaton, M., Kirkwood, M., & Pelander, J. (1971). Effects of various reinforcement contingencies on oral reading rate. In E. Ramp & B. Hopkins (Eds.), *A new direction for education: Behavior analysis* (pp. 54–71). Lawrence, KS: University of Kansas.

Lyon, G. R., & Moats, L. C. (1997). Critical conceptual and methodological considerations in reading intervention research. *Journal of Learning Disabilities*, *30*, 578–588.

National Center for Education Statistics. (2002). *National assessment of educational progress* [online]. Available at http://nces.ed.gov/nationsreportcard/

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.* Retreived January 21, 2002, from http://www.nichd.nih.gov/publications/nrp/smallbook.htm

Office of Technology Assessment. (1993). *The scope of illiteracy in this country* [online]. Available at http://Indianriver.fl.us/living/services/als/scope.html

O'Shea, L. J., Munson, S. M., & O'Shea, D. J. (1984). Error correction in oral reading: Evaluating the effectiveness of three procedures. *Education and Treatment of Children*, *7*, 203–214.

Pearson, P. D., Johnson, D. D., Clymer, T., Indirsano, R., Venezky, R. L., Baumann, J. F., et al. (1989). *Silver, Burdett, and Ginn.* Needham, MA: Silver, Burdett, and Ginn, Inc.

Rashotte, C. A., & Torgenson, J. K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly*, *20*, 180–188.

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

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York: Guilford.

Sindelar, P. T., Rosenberg, M. S., & Wilson, R. J. (1985). An adapted alternating treatments design for instructional research. *Education and Treatment of Children*, *8*, 67–76.

Spache, G. (1953). A new readability formula for primary grade materials. *Elementary English*, *53*, 410–413.

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

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 July 29, 2004
Final acceptance July 14, 2005
Action Editor, Jennifer McComas