JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1996, **29,** 507–518 NUMBER 4 (WINTER 1996)

THE EFFECTS OF INSTRUCTIONAL MATCH AND CONTENT OVERLAP ON GENERALIZED READING PERFORMANCE

EDWARD J. DALY III
UNIVERSITY OF CINCINNATI

AND

BRIAN K. MARTENS, ANNA KILMER, AND DOREEN R. MASSIE
SYRACUSE UNIVERSITY

This study examined the effects of instructional match and content overlap on students' ability to generalize from passage reading instruction. Four students with mild disabilities served as participants. Using a multielement design, students were instructed with passages at two levels of text difficulty (instructionally matched vs. instructionally mismatched), and generalization was assessed with passages at two levels of similarity to those instructed (low vs. high content overlap). Results indicated that students' oral reading accuracy and fluency showed the greatest degree of generalization when instructional materials were matched to the students' skill level and assessment materials were similar to those used during instruction. Moreover, these results were maintained at 1-month follow-up. The implications of these findings for classroom reading instruction and the assessment of students' reading skills are discussed.

DESCRIPTORS: content overlap, instructional match, passage reading, generalization

Factors that can affect students' reading include how well prerequisite skills have been taught, the use of prompts and models, the amount of drill, and whether proficient reading is reinforced. These factors might be categorized as instructional presentation variables because they determine how reading is presented or taught. Other factors that may affect reading performance include the overlap between what is taught and what is tested (i.e., content overlap) and how closely a student's skills match the difficulty of what is to be read (i.e., instructional match). These latter two factors, content overlap and instructional match, might be classified as instructional planning variables because they determine what is to be taught and when it is to be introduced into an instructional sequence. Instructional presentation variables have received considerable attention in the research literature (e.g., Gersten, Woodward, & Darch, 1986; Reid, 1986), whereas instructional planning variables have infrequently been subjected to experimental manipulation.

Although few investigations have experimentally manipulated the amount of overlap between what is taught and what is tested, several studies have demonstrated that systematic differences exist in the degree of overlap between standardized, norm-referenced achievement tests and commercially available reading programs (e.g., Armbruster, Stevens, & Rosenshine, 1977; Jenkins & Pany, 1978a, 1978b; Shapiro & Derr, 1987; Webster, McInnis, & Craver, 1986). In these studies, overlap was assessed by comparing word lists from various reading programs with word lists from the word-recognition subtests of standardized instruments. Results

The authors thank Eric J. Dool and Jennifer Ostrom for their assistance in the preparation of the table and the figures.

Reprints can be obtained from Edward J. Daly III, School Psychology Program, 522 Teachers College, University of Cincinnati, P.O. Box 210002, Cincinnati, Ohio 45221-0002.

showed clear discrepancies between tests for a single program and between programs for a single test, leading to conclusions of curriculum bias or poor content overlap between what is taught and what is tested for many standardized achievement measures. Similarly, Good and Salvia (1988) found that the number of exact word matches between a reading curriculum and test predicted differences in reading performance among third and fourth graders, and these findings were replicated by Bell, Lentz, and Graden (1992) with first and second graders. The obvious consequence of poor content overlap is that test scores and decisions made on the basis of these scores may not accurately reflect what a student has learned.

From a behavior-analytic perspective, content overlap refers to generalization as a function of item similarity. Effective generalization strategies such as (a) teaching skills in the natural environment (Stokes & Osnes, 1988), (b) using multiple exemplars (Stokes & Baer, 1977), and (c) using general case instruction (Albin & Horner, 1988) are based on the assumption that the more overlap there is between teaching conditions and testing conditions, the more likely students are to generalize. We have been unable to find in the literature experimental demonstrations of the effects of content overlap (i.e., using item similarity between testing and teaching conditions) on students' ability to generalize from reading instruction in reading passages. Rather, findings in this area have been based on either the calculation of scores that would be obtained by a hypothetical student or post hoc statistical analyses of grouped test performances (e.g., Bell et al., 1992; Martens, Steele, Massie, & Diskin, 1995).

Good and Salvia (1988) have suggested that there may be individual differences in students' ability to generalize, which would systematically affect their performance on achievement tests. An alternative hypothesis is that there are other variables that may interact with item similarity to affect generalization. One such variable is instructional match or the correspondence between a student's skill level and the difficulty level of the materials used for reading instruction (Hunt, 1961). Presumably, if students' skill levels are better matched to instructional materials, their performance during instruction should be more accurate and fluent and therefore more likely to generalize. Indeed, in a study with 8 first- and second-grade students, Gickling and Armstrong (1978) manipulated instructional materials only and observed differences in on-task behavior, task completion, and task comprehension (i.e., responses to questions about the instructional content of the materials) across three difficulty levels: frustration, instruction, and independent. The frustration-level assignments contained less than 70% known elements during seat work and less than 90% during reading; the instruction-level assignments contained between 70% and 85% known elements during seat work and between 93% and 97% during reading assignments; and the independent-level assignments contained more than 90% known elements during seat work and 97% during reading. They obtained the highest rates of all three behavior categories (on-task, task completion, and task comprehension) when instruction-level material was used.

Gickling advocated controlling daily task difficulty level relative to students' skill levels, observing that "grade level conversions are insensitive to this type of daily assessment" (Gickling & Thompson, 1985, p. 210). There is as yet, however, limited empirical support for this view. Roberts, Turco, and Shapiro (1991) examined the effects of controlling instructional ratios of known to unknown vocabulary words (e.g., 50% known to 50% unknown vs. 90% known to 10% unknown) on acquisition of vocabulary words, passage reading fluency, and passage

comprehension. The intervention included modeling of unknown words and frequent response repetitions (drill). The authors concluded that the group with the greatest ratio of known (90%) to unknown words (10%) showed the most generalization. They based this assertion on the fact that this group showed the greatest gains from pretest to posttest in number of words read correctly relative to the number of words learned during drill. However, it is unclear whether this finding was an artifact, given that this group was presented with (and thus learned) the fewest number of words during drill, which may have artificially inflated the ratio used as a measure of generalization. Thus, it remains unknown whether presenting instructional materials at a difficulty level that matches a student's skill level results in greater generalization.

Again, from a behavior-analytic perspective, good instructional match between student skill level and instructional materials should be expected to promote generalization as a result of better reading proficiency (i.e., higher levels of correct responding). In other words, if students are making more correct responses as a function of well-matched instructional materials, such instruction should result in better generalization. Along these lines, Haring, Lovitt, Eaton, and Hansen (1978) described a learning hierarchy that offers a useful heuristic for conceptualizing the various stages in performing a skill. With proper instruction, students first learn how to perform a skill accurately before they perform it fluently. The fluent performance of a skill, in turn, is believed to promote generalization to novel items, times, and settings by increasing response competition (Wolery, Bailey, & Sugai, 1988) when diverse opportunities for responding are provided as a generalization strategy (Stokes & Baer, 1977). For example, Daly and Martens (1994) compared the effects of three interventions for improving oral reading performance in children with learning disabilities. The 2 children who showed the greatest gains from treatment evidenced higher accuracy rates at baseline, suggesting good instructional match. This study did not, however, explicitly manipulate the degree of instructional match. We have not found any experimental studies that have examined the role of instructional match on passage reading.

The goal of this study was to conduct an experimental analysis of the effects of instructional match and content overlap on students' ability to generalize from passage reading instruction. As in previous research, instructional match was a student performance variable measured as the degree of accuracy and fluency in instructional materials at baseline, whereas content overlap was defined as the proportion of words in assessment passages that overlapped with instructional passages. We examined the effects of (a) students' entry levels of accuracy and fluency in instructional materials prior to instruction (i.e., matched or mismatched), (b) the effects of word similarity between what was instructed and what was tested (i.e., high content overlap or low content overlap), and (c) any interaction between these two variables. The generalized effects of these two instructional planning variables (instructional match and content overlap) were assessed by evaluating the students' reading accuracy and fluency on passages that were not presented during instruction. Based on previous correlational research (Good & Salvia, 1988), we predicted greater generalization from instructional passages that (a) were matched to students' skill levels and (b) contained a high degree of content overlap with what was tested.

METHOD

Participants

Participants were 4 students who received special education services in two self-con-

Table 1
Content of Assessment Passages

Condition	Passage number	Instructional target	Number of words	Percentage of overlap	Percentage of target words	Example
	1	short *i*	48	90	44	"The rig has quit!"
High	2	short *u*	57	77	40	"Big boat puts up a fuss. . ."
	3	short *e*	47	94	51	"He yells and yells. . ."
	4	short *a*	48	98	54	"The rat is glad. . ."
Match						
	1	short *i*	50	24	0	"Pat and Ann are pals. . ."
Low	2	short *u*	46	46	10	"It is hot. . ."
	3	short *e*	51	29	18	"Nan is sad. . ."
	4	short *a*	57	37	25	"Tom can fix a tag. . ."
	1	long *i*	54	83	37	"Mike's five pals wave. . ."
High	2	long *e*	57	88	52	"Dee feeds Zeke. . ."
	3	long *e*	50	80	38	"Pete weeds the seeds. . ."
	4	long *a*	45	82	38	"Jake will not take a nap. . ."
Mismatch						
	1	long *i*	49	43	0	"Dave has a raft. . ."
Low	2	long *e*	54	24	6	"Dune Bug is a cute hot rod. . ."
	3	long *e*	52	37	0	"Sue and June set up a fair. . ."
	4	long *a*	49	18	4	"Joe and Moe are toads. . ."

tained special education classrooms. These students were selected because each was below grade level and had individualized education plan objectives in the area of reading. Their typical reading instruction consisted of passage reading in both basal readers and children's literature. The participants also received skill instruction in basal reading workbooks. Elizabeth was classified as having an emotional disability, Jim and Jon were classified as having a learning disability, and Willie was classified as having mild mental retardation. All ranged in age from 8 years 6 months to 12 years, and had a mean reading achievement standard score of 64 (range, 50 to 80).

Materials

Instructional passages. Eight passages (M = 50 words, range, 49 to 54) were selected from a phonics reading series for instruction (Modern Curriculum Press, 1986). All eight passages were composed of single-consonant words only. The four passages for the instructional match condition contained a large percentage of single-consonant/short-vowel words (M = 46%, range, 27% to 67%). Target phonics skills (in the sequence in which they were instructed) for the instructional match passages were short *i,* short *u,* short *e,* and short *a.* The four passages for the instructional mismatch condition contained a large percentage of single-consonant/long-vowel words (M = 32%, range, 29% to 38%). Target phonics skills (in the sequence in which they were instructed) for the instructional mismatch passages were long *i,* long *e* (twice consecutively), and long *a.*

Assessment passages. Sixteen passages (M = 51 words, range, 45 to 57) were selected from the same phonics reading series for assessment. Table 1 provides the instructional target, the number of words per passage, the percentage of words overlapping with the instructional texts, the percentage of target phonics words, and an example of a typical sentence. The high-overlap passages were the second half of the instructional passages, assuring a high degree of overlap between in-

structional and assessment conditions. Although the low-overlap passages were selected from the same reading series, passages were chosen that included only a small percentage of words containing the target phonics skill from the instructional condition. For example, if the target skill was short *i,* the assessment passage contained a small percentage of short *i* words but a higher percentage of words of equal difficulty (e.g., short *o*). Difficulty level was based on Beck and McCaslin's analyses of phonics sequences for beginning reading programs (as cited in Adams, 1990).

Percentage overlap was calculated by counting the number of words in the assessment passage that either appeared in the instructional passage or contained the target instructional phonic skill and dividing by the total number of words in the assessment passage. The percentage of target phonics words was calculated by counting the number of words containing the instructional target (e.g., short *i*) and dividing by the total number of words in the assessment passage. The instructional match/high-overlap assessment passages contained an average of 90% words that overlapped with their respective instructional passages. The instructional match/low-overlap assessment passages contained an average of 34% words that overlapped with their respective instructional passages. The instructional mismatch/high-overlap passages contained an average of 83% words that overlapped with their respective instructional passages. The instructional mismatch/low-overlap assessment passages contained an average of 31% words that overlapped with their respective instructional passages.

Maintenance passages. Eight of the assessment passages were selected in a semirandom fashion as maintenance probes for a 1-month follow-up assessment. Two high-overlap and two low-overlap assessment passages were selected randomly for each condition: instructional match and instructional mismatch. The high-overlap and low-overlap assessment passages were chosen as corresponding pairs from the initial assessment (hence the semirandom nature of probe selection).

Word lists. Word lists for the instructional treatment were generated from target words in the instructional passages. For each instructional passage, all words containing the target phonics skill (e.g., short *e*) were put on a word list in a single column. The average number of words on the instructional match word lists was 12. The average number of words on the instructional mismatch word lists was nine.

Dependent Measures

The effects of treatment conditions on students' ability to generalize to novel passages were assessed using both accuracy and fluency measures. If the student read the word correctly in 3 s, the word was counted as correct. If the student mispronounced the word, substituted another word, or did not read a word in 3 s, the experimenter said the word to the student and marked the word as incorrect. The experimenter scored student performance on a word-by-word basis. That is, if the student read the word accurately within 3 s, it was counted as a word read accurately and fluently. In this way, each word was evaluated independently of the other words. Oral reading fluency, scored as the number of correctly read words per minute, is a robust measure of reading skill (Shinn, Good, Knutson, & Tilly, 1992). Accuracy was computed by dividing the number of correctly read words by the total number of words in the passage. Fluency was calculated by dividing the number of correctly read words by the number of seconds it took the student to read the passage and multiplying by 60 (to obtain the number of correctly read words per minute). An audiocassette recorder was used to tape the students'

passage reading samples for the purpose of assessing interscorer agreement.

Screening and Preexperimental Training

Students were screened in the instructional passages based on accuracy and fluency. Initially, students displayed very poor accuracy and fluency rates in all instructional passages (i.e., all reading below 60% accuracy and less than 20 correctly read words per minute, on average), indicating poor instructional match. As a result, it was necessary to improve students' reading proficiency in half of the instructional passages in order for these passages to be considered instructionally matched. The distinction between instructional match and instructional mismatch was that students' initial accuracy and fluency rates were higher in the instructional match passages (Gickling & Rosenfield, 1995). Subsequent instruction with these passages (which in this study contained a large percentage of single-consonant/short-vowel words) should promote better generalization as a function of more accurate responding. A preexperimental treatment phase was therefore introduced to increase reading accuracy and fluency on the instructional match passages. The first and third authors read aloud repeatedly to the students those instructional passages intended to constitute the instructional match condition, an instructional technique referred to as *listening passage preview* (Daly & Martens, 1994). Elizabeth received five training sessions, Jim received six, and Willie and Jon each received one. After these preexperimental training sessions, reading accuracy improved to above 80% (M = 84.5%; range, 83% to 87%) and reading fluency improved to above 25 words per minute (M = 36.3; range, 26 to 52) on the matched passages. Reading accuracy (M = 54.3%; range, 36% to 75%) and fluency (M = 13; range, 26 to 52) remained lower for the mismatched passages following the preexperimental training sessions.

Experimental Design and Procedures

A multielement design was used to compare instructionally matched and mismatched conditions, with each session being associated with different instructional materials (Sindelar, Rosenberg, & Wilson, 1985). Sessions were conducted once or twice daily in an office (3 m by 3 m). The student and experimenter were seated next to one another at a table. The experimenter administered the instructional treatment in the target phonic skill for the session and assessed the student's oral reading accuracy and fluency in assessment passages immediately following instruction. The order of the high-overlap and low-overlap passages was counterbalanced across sessions for both instructional conditions (matched and mismatched). Therefore, for half of the sessions, a high-overlap assessment passage was read first following instruction, and, for the other half of the sessions, a low-overlap assessment passage was read first following instruction.

Instruction. The first part of each experimental session was an instructional lesson about a phonics skill. The lesson included a verbal statement of the phonics rule, modeling and drill of target words in isolation, and modeling and drill of target words in the context of a reading passage. The lesson provided instruction of the target skill in isolation with opportunities to apply and practice the skill in context, an important sequence for reading instruction (Grossen & Carnine, 1991). The experimenter read the directions and stated the phonics sound and rule while presenting the word list to the student (e.g., "These words all contain the *i* sound because the vowel stands by itself in the word and is short"). The experimenter then read the word list and asked the student to read the entire word list. If the student did not read a word correctly in 3 s, the

experimenter said the word. Next, the experimenter read the entire instructional passage and asked the student to read the entire instructional passage. If the student did not say a word in 3 s, the experimenter stated the correct word. Sessions (i.e., instruction followed by assessment in a high-overlap passage and a low-overlap passage) lasted approximately 10 to 15 min.

Assessment. Students were assessed immediately following instruction in a phonics rule. The student read one high-overlap and one low-overlap passage. The student read the entire passage while the experimenter marked incorrectly read words and hesitations of more than 3 s and the time needed to complete reading the passage.

Maintenance. Students also were assessed 1 month after the last experimental session with maintenance probes to assess generalization. In one session, the student read the eight maintenance passages (i.e., two instructional match/high-overlap, two instructional match/low-overlap, two instructional mismatch/high-overlap, and two instructional mismatch/low-overlap passages) in their entirety in the same order in which they were read during the experimental phase of the study.

Interobserver Agreement

Two observers (naive to the nature and purpose of the experiment) independently scored 24 (75%) of the assessment passages to assess interobserver agreement. The observers listened to an audiocassette recording of the session to score student performance on the assessment passages. Each word was scored as correct or incorrect. 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 in the passage (which represented all possible agreements and disagreements). The mean interobserver agreement was 95% (range, 79% to 100%).

Treatment Integrity

The two independent observers also assessed treatment integrity during 24 (75%) of the treatment-assessment sessions. The observers recorded whether the experimenter (a) read the directions to the student, (b) stated the phonics sound and rule, (c) read the word list to the student, (d) asked the student to read the entire word list, (e) said the words if the student did not read the words correctly in 3 s, (f) read the entire instructional passage to the student, (g) asked the student to read the entire instructional passage, (h) stated the correct word if the student hesitated for more than 3 s, (i) asked the student to read two assessment passages in the correct counterbalanced order in their entirety, and (j) stated the correct word if the student hesitated for more than 3 s. The correct implementation of experimental conditions was 100% on all sessions but one (during which the experimenter neglected to ask the student to reread the word list).

RESULTS

The percentage of words read correctly (accuracy) on the four types of passages for each of the 4 participants is presented in Figure 1. The highest level of reading accuracy for each participant occurred with the high-overlap/matched passages (M = 76.5; range, 62 to 91), followed by the high-overlap/mismatched passages (M = 64; range, 48 to 79), and then the low-overlap/matched passages (M = 56.3; range, 45 to 70), and finally the low-overlap/mismatched passages (M = 51; range, 34 to 68). These results indicated that both instructional match and content overlap affected reading accuracy and that these two variables interacted to produce the superior effects of the high-

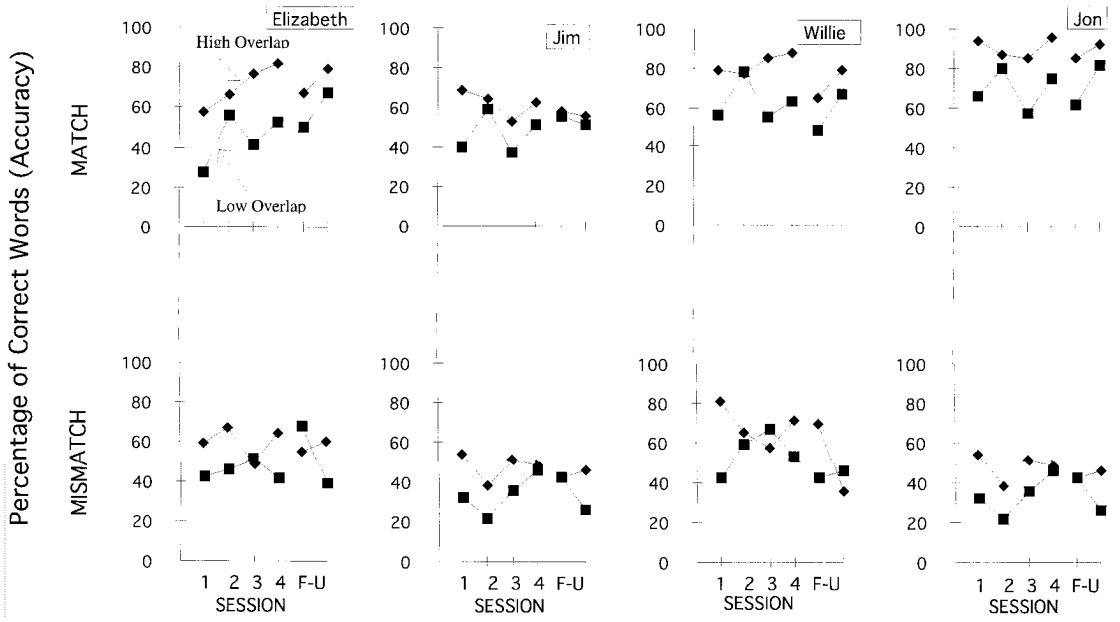

Figure 1. Percentage of words read correctly (accuracy) in instructional match (top panel) and instructional mismatch (bottom panel) conditions for all participants. F-U = follow-up.

overlap/matched passages. All 4 children showed the same pattern of a higher percentage of correct words in the high-overlap/matched condition, although the difference between this condition and the others was small for Jim.

Figure 2 shows the corresponding results for the other dependent measure, the num-

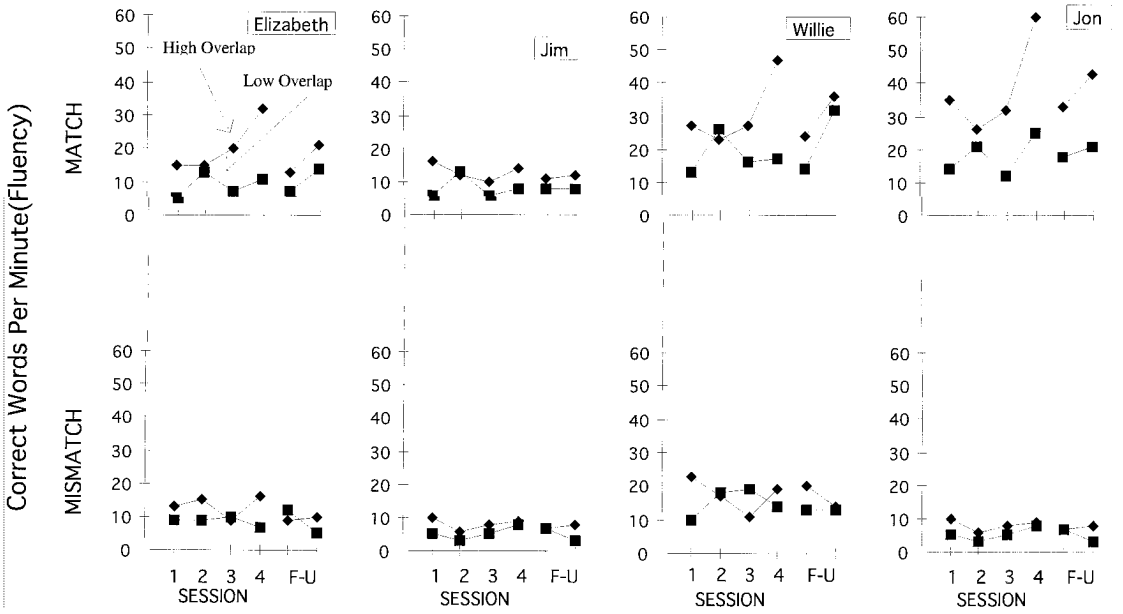

Figure 2. Number of words read correctly per minute (fluency) in instructional match (top panel) and instructional mismatch (bottom panel) conditions for all participants. F-U = follow-up.

ber of correct words per minute (fluency). The pattern of results was similar for the number of correct words read per minute (fluency). That is, the highest number of correct words read per minute was in the high-overlap/matched condition for all 4 children, but the difference between this condition and the others was small for Jim.

Finally, the last two data points in each panel for both accuracy (Figure 1) and fluency (Figure 2) show the data collected at 1-month follow-up. Better reading accuracy and fluency associated with the high-overlap/matched condition were also evident at this follow-up.

DISCUSSION

These results support the importance of content overlap in assessing the effects of phonics skill instruction. More specifically, in the low-overlap condition, only 32% of the words in assessment passages appeared in the instructional passages and only 8% of the words contained the phonics skills targeted by the instructional passages. This low level of overlap resulted in students reading eight fewer words per minute with 16.6% lower accuracy on average following instruction. These results are consistent with research on curriculum bias in standardized tests of reading decoding (e.g., Martens et al., 1995) in suggesting that as the assessment materials provide fewer opportunities to demonstrate the target skill (i.e., low content overlap), the assessment results will likely underestimate the actual impact of instruction. Although Fuchs and Deno (1992) found that students' performance on a standardized reading measure correlated similarly with their oral reading rate as measured by two different curricula, the reading rates themselves differed across curricula by as much as 31 words per minute for students in Grades 3 and 6. This suggests that although students' relative standings on assessment passages may not differ as a function of curriculum content, their absolute performance levels may, leading to different conclusions about instructional effectiveness.

Fuchs and Deno (1992) argued that if assessments are conducted repeatedly over time on material of comparable difficulty with valid outcome indicators that allow for error analysis, then it is not essential to use actual curricular materials to assess global outcomes. Although this may be the case when comparing students' relative standings on global measures, the present findings suggest that instructional decisions about specific skills for any single student will differ as a function of content overlap.

Fuchs and Deno (1991) identified a number of potential limitations of subskill mastery measurement, including arbitrary instructional hierarchies, failure to assess generalization over time, frequent measurement shifts, and teacher-made tests of questionable technical adequacy. The procedures used in the current investigation provide ways of addressing most of these limitations. First, the assessment passages came from the same phonics curriculum (Modern Curriculum Press, 1986) and were similar in terms of global difficulty (i.e., the grade levels of the passages were approximately equivalent across conditions). Second, the procedures used for measuring reading accuracy (percentage of correct words) and fluency (correct words per minute) were constant over time. Selecting passages from a sequenced reading curriculum and using consistent measures of reading accuracy and fluency over time should limit the arbitrariness of the instructional sequence, eliminate large measurement shifts, and provide teacher-made tests with an outcome indicator that has substantial empirical support for its reliability and validity (Shinn, 1989; Shinn et al., 1992). Third, generalization was assessed over time using passages that varied systematically in terms of their level of similarity to

the passages used during instruction. Finally, these measures were highly sensitive to the effects of brief instructional sessions, thus providing information relevant to day-to-day instructional decisions, a drawback of global outcome measures.

The current findings also suggest that good instructional match prior to instruction promotes the generalization of reading to passages that are both similar to or different from those instructed. Under instructional match conditions (an average of 85% accuracy and 36 words per minute pretreatment), students read an average of six words more per minute with 9% higher accuracy on assessment passages than under mismatch conditions. Increased generalization under the instructional match condition might be attributed to greater response competition for correct decoding as a function of higher accuracy and fluency during instruction (Wolery et al., 1988). An alternative explanation is that the differences in generalization of reading performance between the matched and mismatched passages were due to the preexperimental training provided on the matched but not on the mismatched instructional passages. Had we conducted baseline measurements with the assessment passages after the preintervention training and prior to the experimental instruction, we would have been able to differentiate the generalized effects of the preintervention training from the effects of the instruction provided at the beginning of each experimental session.

A potential limitation of this study concerns the nature of the instructional conditions. Specifically, to provide a direct experimental demonstration of the variables in question, students received individual instruction involving modeling and drill of phonics skills in isolation with opportunities to practice the skills in context. This approach was consistent with prescriptions for quality reading instruction (Grossen & Carnine, 1991) and was believed to produce strong learning. Because variables such as modeling and practice are often confounded with the way students are grouped for instruction, the present findings need to be extended to natural classroom settings.

Another potential limitation of this study is that the students displayed such low response rates (even in the instructional match conditions) that it is difficult to say that the students were well matched to the instructional materials. In fact, some may argue that the students were at a frustration level even in the instructional match passages. It is not possible at this time, however, to make inferences about how well students who are reading at higher rates would generalize under the same conditions or even how well the same students would generalize if given easier instructional material.

These results suggest that instructional match and content overlap produce an interaction effect on reading performance. If students' skill levels are well matched to the instructional materials and assessment is conducted on materials that have a high degree of overlap with instructional materials, generalized reading performance may be more likely to occur. This interaction has important implications for instructional planning. We identified a range of accurate and fluent responding that promoted greater generalization to high-overlap assessment passages. Specifically, students who were well matched to the instructional materials read the instructional passages with 85% accuracy at a rate of 36 correctly read words per minute prior to treatment. These results are consistent with previous research that has established instructional levels for oral reading fluency (Deno & Mirkin, 1977; Shapiro, 1989). The current findings indicate the importance of screening students and selecting materials so that all students in a reading group are performing within a similar, optimal range of accuracy and fluency. Al-

though teachers have long relied on ability grouping to facilitate instruction, these findings emphasize the potential benefits for generalization. The direct benefit to both pupil and teacher is that students may require exposure to fewer passages at their instructional level before displaying sufficient generalized reading performance to move on in the curriculum. Teachers who take into account the potential effects of instructional match and content overlap on instructional outcomes should consider assessing student proficiency in instructional materials prior to assignment of instructional tasks and assessing generalized outcomes across instructional tasks.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Albin, R. W., & Horner, R. H. (1988). Generalization with precision. In R. H. Horner, G. Dunlap, & R. L. Koegel (Eds.), *Generalization and maintenance* (pp. 99–120). Baltimore: Paul H. Brookes.

Armbruster, B., Stevens, R., & Rosenshine, B. (1977). *Analyzing content coverage and emphasis: A study of three curricula and two tests*. (Technical Report No. 26). Urbana, IL: University of Illinois, Center for the Study of Reading.

Bell, P. F., Lentz, F. E., & Graden, J. L. (1992). Effects of curriculum-test overlap on standardized achievement test scores: Identifying systematic confounds in educational decision making. *School Psychology Review, 21,* 644–655.

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.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488–500.

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232–242.

Gersten, R., Woodward, J., & Darch, C. (1986). Direct instruction: A research-based approach to curriculum design and teaching. *Exceptional Children, 53,* 17–31.

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.

Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52,* 205–218.

Good, R. H., III, & Salvia, J. (1988). Curriculum bias in published norm-referenced reading tests: Demonstrable effects. *School Psychology Review, 17,* 51–60.

Grossen, B., & Carnine, D. (1991). Strategies for maximizing reading success in the regular classroom. In G. Stoner, M. R. Shinn, & H. M. Walker (Eds.), *Interventions for achievement and behavior problems* (pp. 333–356). Washington, DC: National Association for School Psychologists.

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

Hunt, J. M. (1961). *Intelligence and experience*. New York: Ronald Press.

Jenkins, J. R., & Pany, D. (1978a). Curriculum biases in reading achievement tests. *Journal of Reading Behavior, 10*(4), 345–357.

Jenkins, J. R., & Pany, D. (1978b). Standardized achievement tests: How useful for special education? *Exceptional Children, 44*(6), 448–453.

Martens, B. K., Steele, E. S., Massie, D. R., & Diskin, M. T. (1995). Curriculum bias in standardized tests of reading decoding. *Journal of School Psychology, 33,* 287–296.

Modern Curriculum Press. (1986). *Phonics practice readers*. Cleveland, OH: Author.

Reid, E. R. (1986). Practicing effective instruction: The exemplary center for reading instruction approach. *Exceptional Children, 52,* 510–519.

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

Shapiro, E. (1989). *Academic skills problems: Direct assessment and intervention*. New York: Guilford.

Shapiro, E., & Derr, T. (1987). An examination of overlap between a reading curricula and standardized achievement tests. *Journal of Special Education, 21,* 59–76.

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

Shinn, M. R., Good, R. H., III, Knutson, N., & Tilly, W. D., III. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory anal-

ysis of its relation to reading. *School Psychology Review, 21,* 459–479.
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.
Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.
Stokes, T. F., & Osnes, P. G. (1988). The developing applied technology of generalization and maintenance. In R. H. Horner, G. Dunlap, & R. L. Koegel (Eds.), *Generalization and maintenance* (pp. 5–19). Baltimore: Paul H. Brookes.
Webster, R. E., McInnis, E. D., & Craver, L. (1986). Curriculum biasing effects in standardized and criterion-referenced reading achievement tests. *Psychology in the Schools, 23,* 205–213.
Wolery, M., Bailey, D. B., & Sugai, G. M. (1988). *Effective teaching: Principles and procedures of applied behavior analysis with exceptional students.* Boston, MA: Allyn & Bacon.

Received March 8, 1996
Initial editorial decision May 6, 1996
Revision received May 30, 1996
Final acceptance June 26, 1996
Action Editor, Wayne W. Fisher

STUDY QUESTIONS

1. What distinction did the authors make between instructional planning variables and instructional presentation variables, and what examples of these variables did the authors give?

2. What is meant by the term *instructional match,* and what are its presumed effects on student performance?

3. What is *content overlap,* and what is its relevance to assessment?

4. Describe the instructional and assessment passages. How did the experimenters determine the degree of instructional match and overlap between these passages?

5. How were the instructional and assessment procedures implemented and evaluated? Also, why did the authors conduct preexperimental training?

6. What were the dependent variables, how were they measured, and how was interobserver agreement assessed?

7. According to the results, which variables produced the highest measures of accuracy and fluency? Using Figure 1 as an example: (a) Which set of data portrays the combined effects of content overlap and match? (b) Which sets of data should be compared to determine which of the two variables had greater influence on student performance? (c) For which subject did high match produce better accuracy than high overlap?

8. What are the implications of the results for the design of instructional and testing procedures?

Questions prepared by Jana Lindberg and Eileen Roscoe, University of Florida