DIANE KINDER and MARCY STEIN, University of Washington, Tacoma;
JEAN OSBORN, University of Illinois

Analyzing Beginning and Adolescent Reading Programs: Exploring Issues of Content, Coherence, and Assessment

Abstract: Two curriculum evaluation projects were conducted in response to requests from practitioners interested in determining the quality of commercially developed reading programs. The 1st project, sponsored by 3 school districts in Texas, focused on 5 1st-grade reading programs. The 2nd, conducted at the request of junior high, middle-, and high-school teachers from several school districts in the Pacific Northwest, examined 8 remedial reading programs. The evaluation criteria, used in both projects, were organized into 3 areas: program content, program coherence, and student assessment. Results of application of these criteria to beginning and adolescent reading programs are reported. These projects generated the necessary preliminary work for the development of a research-based curriculum evaluation instrument. The paper concludes with recommendations for further work in the development of reliable and valid instruments to help educators evaluate and select curricular materials.

The report on reading from the 2002 National Assessment of Educational Progress (NAEP) documents the lack of significant improvement in reading for fourth-grade students during the last decade. In 2002, only 31% of fourth graders performed at or above the proficient level. In recent years, not only did average scores fail to improve, but the gap between the highest and lowest performing students increased, with the lowest performing students scoring even more poorly than in the past (Donahue, Finnegan, Lutkus, Allen, & Campbell, 2001). The reading performance of adolescent students also remains a major concern to educators. Only 33% of 8th graders and 36% of 12th graders scored at or above the proficient level on their NAEP reading assessment (NAEP, 2002). In addition, 25% of 8th-grade students and 26% of 12th-grade students were functioning below the basic level, demonstrating only partial mastery of the skills of proficient readers.

To meet the literacy challenges documented by these assessments, many educators are reconsidering the role that commercially developed reading programs play in the classrooms of American schools (Baker, Kame'enui, Simmons, & Stahl, 1994; Snow, Burns, & Griffin, 1998). This interest in published reading programs follows the move away from these programs in the late 1980s and early 1990s when whole language proponents encouraged the use of trade books and teacher-developed lessons rather than systematic reading programs.

When student performance lagged in the 1990s (NAEP, 2002), Ball and Cohen (1996)

were among the first to suggest that teachers needed curriculum materials for instructional support. While all teachers deserve the best designed materials available, many educators now agree that these materials are especially important for teachers who have not had adequate teacher preparation or who do not have adequate planning time to adapt and modify their materials. Heibert and Taylor (2000) noted that paraprofessionals, who are becoming increasingly responsible for teaching reading to struggling students in Title I and special education classes, also need well designed materials.

Currently, No Child Left Behind (2001) has generated an increased emphasis on the use of scientifically based reading programs. As a result, expertise in curriculum analysis based on current reading research has become critical to selecting these programs. Historically, curriculum adoption committees have developed criteria and conducted analyses of commercial programs. Yet, few teachers who serve on curriculum adoption committees have experience generating research-based criteria to help determine the quality of a reading program (Comas & Farr, 1989; Stein, Stuen, Carnine, & Long, 2001). Many schools and districts are forced to rely on teacher observation and informal rating scales, often provided by publishers (Stein et al., 2001).

Ideally, educators would have access to curriculum evaluation instruments that incorporate research-based criteria to assist them in the evaluation process. However, valid and reliable instruments for evaluating commercially developed reading programs do not exist at this time. Clearly, systematic research is necessary in order to develop valid criteria and objective procedures for the analysis of reading programs. The purpose of this paper is to contribute to the larger discussion of how best to approach the development of research-based evaluation instruments.

The two projects described in this paper each explored ways to use research to guide the development of evaluation criteria for analyzing reading programs. The projects were conducted in response to requests from school districts and practitioners interested in determining the quality of available commercially developed reading programs. The first project, sponsored by Brownsville, Fort Worth, and Houston Independent School Districts, was an analysis of decoding instruction in the first-grade reading programs approved for adoption in Texas in 1999 (Stein, Johnson, Boutry, & Bortleson, 2000). The second, an analysis of decoding instruction in remedial reading programs designed primarily for adolescent readers, was conducted at the request of junior high, middle-, and high-school teachers from several school districts in the Pacific Northwest.

Although the target student population of the two projects was quite different (i.e., beginning and remedial readers), the needs of the target students have some important similarities. Considerable evidence suggests that many students who require special reading assistance are lacking the decoding skills typically acquired in the primary grades (Juel, 1988; Lyon, 1996; Stanovich, 1986). Therefore, many remedial programs for adolescent readers are designed to provide instruction in the decoding skills these students failed to acquire. Clearly, struggling adolescent readers experience difficulty with vocabulary and reading comprehension as well as decoding. In fact, the authors have also conducted a preliminary analysis of the comprehension and vocabulary instruction in selected remedial programs. For the purposes of this paper, however, only the decoding instruction found in remedial programs is addressed.

What follows is a brief discussion of the development of a preliminary set of evaluation criteria and the subsequent application of those criteria to both beginning and remedial reading programs. The paper concludes with a discussion of what evaluators learned from both

the development and application of these criteria and some recommendations for further work in the development of research-based curriculum evaluation instruments.

Developing Evaluation Criteria

The underlying goal of this work was to use what is known from research about reading and reading instruction to generate evaluation criteria for analyzing decoding instruction in reading programs. The findings from several recent research summaries (Adams, 1990; National Reading Panel, 2000; Snow et al., 1998) provided the research base for the evaluation criteria used in both of the curriculum evaluation projects. A convergence of findings reported in these research summaries on decoding instruction suggests the importance of instruction in the use of systematic, explicit phonics strategies along with practice in oral reading fluency for both beginning and remedial readers. Therefore, evaluation criteria were designed to address these topics.

In addition to the topics of phonics and fluency highlighted by the research summaries, decodable text, considered by many researchers to be an important component of well designed phonics instruction, also was addressed in the evaluation. Mesmer (in press) defines decodable text both by the phonics regularity of the words in text selections and by the degree of match between the phonics elements taught to students and the phonics elements in words from the text selections (i.e., "lesson-to-text match"). Although, at this time, there is limited experimental research on decodable text, researchers generally agree that the application of phonics to reading text through the use of decodable text facilitates the acquisition of decoding skills in young readers (Adams, 1990; Juel & Roper-Schneider, 1985; Mesmer, in press; Snow et al., 1998; Stein, Johnson, & Gutlohn, 1999). Monitoring student progress throughout instruction also has been shown to have a positive impact on stu-

dent reading achievement (Fuchs, 1998; Good, Simmons, & Kame'enui, 2001; Kaminski & Good, 1996). Quality assessment materials help teachers identify reading problems and provide specific recommendations for subsequent instruction. Therefore, the evaluation criteria developed included questions regarding the quality of the assessment procedures in the programs examined.

The evaluation criteria used in both projects were organized into three areas: program content, program coherence, and student assessment. Program content addresses the quality and type of decoding instruction, that is, the type of phonics approach (e.g., explicit or implicit) predominant in the program as well as the availability of instruction for increasing reading fluency. Program coherence refers to the degree to which the major components of a given program are coordinated with and support one another. The evaluation of student assessment includes both the number of assessment options available to teachers as well as how the assessment options are linked to instructional recommendations based on student performance.

The evaluation criteria contained a combination of objective measures in which evaluators counted or listed specific elements in a program and evaluator ratings that required greater judgment by evaluators. Examples of more objective measures include the count of the number of phonics elements introduced during specified lessons and lists of the types of text selections recommended. Evaluator ratings addressed such topics as the clarity of directions to teachers and the quality of fluency building activities. A description of each of the evaluation criteria is included in Tables 1–8.

Although the evaluation criteria for the two projects are not identical, they do address the same three areas described above. The criteria were slightly altered to match the specific demands of comprehensive reading programs

and remedial reading programs. For example, the remedial programs examined did not contain more than one text selection option. Therefore, a tally of the number of text selections was not included in the evaluation.

Applying Evaluation Criteria to First-Grade Reading Programs

As mentioned earlier, three school districts in Texas, interested in more information about the five first-grade reading programs approved by their state for adoption, sponsored this analysis (Stein et al., 2000). The programs that were adopted by Texas in 1999 include *Collections* (2000), *Literacy Place* (2000), *McGraw-Hill Reading* (2001), *Open Court Reading* (2000), and *Scott Foresman Reading* (2000). No Direct Instruction reading programs were included in this analysis because none were adopted by the State of Texas in 1999.

Evaluators examined the instruction in Weeks 10 and 26 of each program to determine how and if the decoding instruction changed dramatically over time. Using this analysis, evaluators could determine the extent to which students were encouraged to move from sounding out strategies (Week 10) to sight reading strategies (Week 26). Week 10, selected as representative of initial reading instruction, was selected over earlier weeks (Weeks 1–9) to avoid examining what some publishers may have considered review of the kindergarten level. Week 26 was selected as representative of instruction at the end of first grade. By analyzing the instructional activities within the context of these 2 weeks, evaluators also were able to determine the overall structure of the beginning reading instruction in each program. That is, when and how phonics strategies are taught, when and how text selections are read, and when and how assessment is addressed.

Program content. Program content (see Table 1) includes the evaluation of phonics instruction and fluency building activities. To begin the

analysis of a program's approach to phonics instruction, evaluators counted the number of phonics elements introduced during Weeks 10 and 26 to get a sense of a program's rate of introduction for these elements. Evaluators were interested in whether some programs introduce these elements at a higher rate than other programs. The number of phonics elements as listed in Table 1 includes individual letter–sound correspondences (e.g., *a*), letter combinations (e.g., *ea* and *sh*), and prefixes and suffixes (e.g., *pre*, *re*).

After identifying the number of phonics elements, evaluators established whether the program offered a predominantly explicit phonics approach to learning to read. A high rating indicated that the instructional recommendations in the program included explicit teaching of letter–sound relationships and the use of explicit phonics strategies to teach word identification (e.g., sounding out). Additionally, in order to receive a high rating the program needed to contain clear and consistent instructional language.

Evaluators rated the program's instructional recommendations for building fluency. This rating is based on the number of rereading activities in a program that involve careful teacher monitoring. Also, the rating reflected whether the program provides teachers with rate and accuracy criteria for fluency activities. These ratings are summarized in Table 1.

Table 1 shows the number of phonics elements introduced in Weeks 10 and 26. As the table indicates, the programs generally introduce between two and four phonics elements per week. The exception appears to occur during Week 10 when *McGraw-Hill Reading* introduces eight new phonics elements. (It should be noted, however, that in Week 26, this program introduces only two phonics elements.) Table 1 also shows that *Open Court Reading* introduces no new phonics elements in Week 26 because at this point in the program all phonics elements have been introduced.

Table 1

First-Grade Reading Programs, 2-Week Analysis: Program Content

Content	*Collections* Harcourt, Inc.	*Literacy Place* Scholastic, Inc.	*McGraw-Hill Reading* McGraw-Hill	*Open Court Reading* SRA	*Scott Foresman Reading* Addison-Wesley
Phonics elements					
(Week 10)	2	3	8	4	1
(Week 26)	4	2	2	0	2
Explicit phonics instruction rating	3	3	2	3	2
Fluency building activities rating	1	1	2	1	1

Note. **Phonics elements.** The lists in this category include the number of phonics elements introduced during Weeks 10 and 26 of instruction of each program.

Explicit phonics instruction. Evaluators rated the presence of explicit phonics strategies on a 3-point scale using the following criteria:

3 Initial instruction in phonics elements is explicit (letter–sound); explicit phonics word identification strategies are prominent; instructional language used by the teacher is clear and consistent.

2 Initial instruction in phonics elements is explicit; explicit phonics word identification strategies are prominent, but the program contains instruction in using context clues for decoding; instructional language is inconsistent.

1 Initial instruction in phonics elements is explicit; however, word identification strategies are not consistent; instructional language is confusing.

Fluency building activities. Evaluators rated programs on a 3-point scale according to the following criteria:

3 Program provides a number of rereading activities that involve significant teacher monitoring of fluency including directions to teachers that provide rate and/or accuracy criteria.

2 Program contains rereading activities but few specific directions to teachers for monitoring fluency.

1 Program contains no specific directions for teacher monitoring of fluency.

All of the beginning reading programs included in this evaluation teach explicit phonics strategies. Although explicit phonics instruction is the predominant approach used to introduce new phonics elements in the programs, the application of these word identification strategies does not always appear to be consistent throughout each program. For example, evaluators frequently found examples of decodable words (i.e., words for which the students had previously been taught the phonics elements) that were introduced as sight words rather than as words to be sounded out.

Finally, Table 1 reports the evaluation of fluency building activities. Most of the first-grade programs provided little guidance to help teachers build their students' reading fluency. Rarely did the programs provide rate and accuracy criteria benchmarks that would help teachers monitor student progress in developing fluency. More specifically, only *McGraw-Hill Reading* provided activities for monitoring reading fluency. None of the other programs provided any specific activities or directions for fluency monitoring in the 2 weeks of instruction that were analyzed.

Program coherence. In this analysis, program coherence refers to the extent to which the recommended instructional activities are coordinated with each other. First, evaluators listed the types of text selections included in each program. They then examined the extent to which students were given the opportunity to apply the phonics that had been taught to reading the text selections. That is, they rated the extent to which the text could be considered decodable. Evaluators analyzed the decodability of the text selections using the procedures adapted by the Texas Education Agency (2000). To determine decodability, evaluators examined only core student reading materials (i.e., the student anthology), not supplementary materials. Decodable text was examined in Week 10 only, during initial reading instruction, based on the assumption that by Week 26 stu-

dents are better prepared to read less controlled text. See Percent of Decodability of Student Anthologies in Table 2.

Finally, evaluators determined ratings for program coherence by examining the extent to which the phonics instruction that was provided during the week was integrated with the text selections for that week. Most of the reading programs analyzed contained at least four types of text selections: student anthologies, decodable books or phonics readers, leveled readers, and trade books. The comparison of the decodability of the student anthology selections to that of the decodable books provided evaluators with an estimate of the level of coordination among the text selections in a program. Evaluators also rated the programs according to the clarity of directions to teachers for using these various text selections. Specifically, evaluators examined whether the programs provided directions to teachers about when, how, and with whom they should teach the selections (see Table 2).

Table 2 shows the different types of reading selections across programs. Five of the programs contained approximately six to seven different types of selections. Two of the programs, *McGraw-Hill Reading* and *Scott Foresman Reading*, included four different types of text selections. While including a variety of text selections is important to many teachers, most of the program manuals fail to indicate clearly *when* and *how* each selection should be taught and *for whom* the various text selections are most appropriate. Guidance in selecting appropriate text selections is important for several reasons. For example, it is reasonable to assume that students who need decodable texts are likely to find reading the less decodable anthology selections and leveled readers difficult and frustrating.

Table 2 also indicates that the range of decodability found in the student anthologies for Week 26 varies from 33% (*Scott Foresman Reading*) to 66% (*Literacy Place*).

Other programs were in the 50% decodable range. One program, *Open Court Reading*, was significantly different in its approach to the use of decodable text. While other programs appear to include more or less decodable text within the same lesson or unit, *Open Court Reading* does not include reading from the student anthology at Week 10. Rather, the anthology is to be read to the students until students have been introduced to more phonics elements.

In terms of coherence ratings, the relationship between phonics instruction and text selections, evaluators also found some variability. One program, *Scott Foresman Reading*, provides

Table 2

First-Grade Reading Programs, 2-Week Analysis: Program Coherence

	Collections Harcourt, Inc.	*Literacy Place* Scholastic, Inc.	*McGraw-Hill Reading* McGraw-Hill	*Open Court Reading* SRA	*Scott Foresman Reading* Addison-Wesley
Types of text selections	Student anthology	Student anthology	Student anthology	Student anthology	Student anthology
	Phonics practice readers	Phonics readers	Phonics practice readers	Predecodable texts	Phonics readers
	Leveled readers	Guided reading library	Leveled books	Decodable books	Leveled readers
	Take-home books	Trade books	Trade books	Decodable take-home stories	Trade book library
	Cut-out books	Chapter books		Big book (teacher reads aloud)	
	Library books	Big books			
	Big books				
% Decodability	54%	66%	51%	N/A*	33%
Coherence rating	3	2	2	3	1

* Students in this program do not read from the anthology during Week 10.

text selections with a wide range of decodability, coordinating phonics instruction with some text selections, but not coordinating the text selections with each other. Two programs, *Literacy Place* and *McGraw-Hill Reading*, demonstrate moderate coherence. Although these programs coordinate phonics instruction with some text selections, they do not appear to coordinate the text selections with each other. *Collections* appears to more carefully coordinate its phonics instruction with its text selections and the text selections with each other. The decodable text in *Collections* was clearly a central focus of the instruction, with the program

manuals providing explicit directions to teachers for using decodable text. *Open Court Reading* also coordinates instruction in decodable texts with instruction in the student anthology. During the first half of first grade, students read primarily from the decodable text selections that provide practice in applying the phonics instruction provided in previous lessons while teachers read the student anthologies to them. Only later in the program do the students read in their student anthologies. As students acquire more sophisticated phonics skills they are better able to read more complex text. In contrast, evalua-

Table 2 (continued)
First-Grade Reading Programs, 2-Week Analysis: Program Content

Note. **Text selection.** Evaluators listed the types of text selections included in the recommended instruction for Weeks 10 and 26.

Percent decodability of student anthologies. The decodability of the selections in the student anthology in Week 10 (only) for each program was determined by calculating the percentage of words in the text selection containing phonics elements that had been previously introduced to students.

Program coherence. Evaluators rated program coherence by examining the relationship between the phonics instruction presented during Weeks 10 and 26 and the text selections students read during those weeks. Prominence of decodable readers as a core activity of the program is considered. The relationship among the various text selections recommended for use during the 2 weeks was also considered in the rating. Evaluators rate program coherence on a 3-point scale using the following criteria:

3 The program coordinates its phonics instruction with recommended text selections as well as provides text selections that are well coordinated with each other. Decodable readers are a clear part of the instructional program, and teachers are provided with directions on how to use them in the teacher manuals.

2 The program coordinates its phonics instruction with some text selections but does not coordinate the text selections with each other. Students are asked to read some text selections with a moderate range of decodability.

1 The program coordinates its phonics instruction with some text selections but does not coordinate the text selections with each other. Students are asked to read text selections with a wide range of decodability.

tors found that during a given week, some programs direct teachers to have students read a variety of text selections; the decodability of these selections may range from over 80% (in a decodable book) to less than 50% (in the student anthology).

Assessment. One of the primary purposes of student assessment is to provide teachers with information about student progress so that they can make informed instructional decisions. All of the assessment materials provided by the publishers were reviewed. Table 3 illustrates the range of assessment options provided by each of the programs. All of the

reading programs include a number of assessment options. The simple *number* of assessment options, although interesting, does not address how useful these instruments are for improving instruction. Therefore, evaluators made an effort to conduct a more qualitative evaluation of the assessment options in the programs by closely examining five features.

The assessment rating scales and results of the evaluation for these five features are presented in Table 4. The first assessment feature was the presence of a placement test with alternative placement options for students. Two of

Table 3
First-Grade Reading Programs: Assessment Options

Collections Harcourt, Inc.	*Literacy Place* Scholastic, Inc.	*McGraw-Hill Reading* McGraw-Hill	*Open Court Reading* SRA	*Scott Foresman Reading* Addison-Wesley
Grade 1 reading inventory	Unit skills tests	Placement test	Pre/Post test	Placement tests
Mid-Year and end-of-year assessment	Unit benchmark tests	Unit assessments	Skills assessment	Unit skills tests
Selection comprehension test	End-of-Year skills test	Selection tests	Literature comprehension assessment	Unit benchmark tests
Skills assessment	End-of-Year benchmark test	IRI	Oral fluency assessment	End-of-Year skills test
Holistic reading assessment	Selection tests	Running records	Teacher observation log	End-of-Year benchmark test
Performance assessment	Spelling tests	Grammar and spelling assessment		Selection tests
Portfolio assessments		Word recognition		

the programs, *Literacy Place* and *Open Court Reading*, lack placement tests. However, *Collections* and *Scott Foresman Reading* both contain placement tests that provide a range of starting points for students based on their performance. Next, evaluators examined programs to determine the extent to which the programs contained performance standards. Two of the programs rarely specify performance standards (i.e., *Literacy Place* and *Open Court Reading*). Some of the assessment options in the remaining three programs, *Collections*, *McGraw-Hill Reading*, and *Scott Foresman*, provide performance standards. However, no program in this evaluation offers performance standards for the majority of its assessment options.

The third and fourth assessment questions addressed the recommendations for remediation and/or acceleration. That is, evaluators examined programs to determine whether they contain specific instructional recommendations for reteaching if students did poorly or recommendations for moving more quickly through the program for those students who performed particularly well. Evaluators found few programs that provide teachers with assessment procedures linked to specific recommendations for remediation or acceleration. In fact, only one program, *Collections*, provides remediation or acceleration recommendations that were coordinated with student performance based on program assessment.

Finally, the extent to which assessment directions were clearly articulated for teachers was rated. In many programs the assessment instruments appear optional. While some programs refer teachers to assessments in their resource materials, others never mention an assessment during the weekly instructional plan. Furthermore, the assessment materials are often difficult to locate and use. Only two of the five programs examined, *Collections* and *Open Court*, include clear explanations for the use of their assessment options.

Table 4
First-Grade Reading Programs: Assessment/Coherence

Rating	*Collections* Harcourt, Inc.	*Literacy Place* Scholastic, Inc.	*McGraw-Hill Reading* McGraw-Hill	*Open Court Reading* SRA	*Scott Foresman Reading* Addison-Wesley
Placement test	3	1	2	1	3
Performance standards specified	2	1	2	1	2
Recommendations for remediations	2	1	1	1	1
Recommendations for acceleration	2	1	1	1	1
Clarity of directions to teachers	3	2	2	2	3

Applying Evaluation Criteria to Adolescent Reading Programs

The purpose of the second project was to generate and apply evaluation criteria derived from reading research to help guide the analysis of commercially developed literacy programs for struggling adolescent readers. Programs were selected for analysis by examining the adolescent literacy literature, publisher's catalogs, and by consulting with teachers who work with

---

Table 4 (continued)
First-Grade Reading Programs: Assessment/Coherence

Note. **Placement tests.** Evaluators rated programs on a 3-point scale using the following criteria:

3   Program provides a placement test with specific recommendations for student placement in the program according to test results.

2   Program provides a placement test with no specific recommendations for student placement in the program.

1   Program does not provide a placement test.

Performance standards specified. Evaluators examined assessment options to determine whether student performance standards are available. Evaluators rated programs on a 3-point scale using the following criteria:

3   Performance standards are available for most assessment options.

2   Performance standards are available for some assessment options.

1   Performance standards are rarely available for assessment options.

Recommendations for remediation and acceleration. Evaluators examined assessment options to determine whether specific recommendations for remediation and/or acceleration are available. Evaluators rated programs on a 3-point scale using the following criteria:

3   Recommendations are consistently available.

2   Recommendations are sometimes available.

1   Recommendations are rarely available.

Clarity of directions to teachers. Evaluators examined the clarity of directions to determine if they specify when and how to use the various assessment tools. Evaluators rated programs on a 3-point scale according to the following criteria:

3   Directions to teachers for using assessment tools are explicit and easily accessible.

2   Directions to teachers for using assessment tools are not consistently explicit and accessible.

1   Directions to teacher for using assessment tools are vague and difficult to access.

---

adolescent students. Programs for struggling readers often target one area for remediation, such as reading comprehension or study skills. The teachers who initiated this evaluation were mostly interested in decoding programs for struggling adolescent readers. Therefore, only evaluations of programs that specify a decoding emphasis and include sequenced lessons and instructional activities are reported here. Because of time constraints and availability of materials, the analysis was limited to print programs and did not include computer-based programs. General information about the eight programs analyzed is outlined in Table 5.

Adolescent literacy programs vary greatly. For example, the number of lessons in the pro-

grams ranges from 20 (*REWARDS*) to 320 (*Corrective Reading, A–C*); therefore, the approximate time for program completion varies from 4 weeks to 3 years. The amount of daily teaching time required ranges from 30 min (*Sounder*) to 1.5 hr (*Laubach Way to Reading*). In addition, some of the programs are clearly designed for nonreaders (the lower levels of *Corrective Reading*, *Laubach Way to Reading*, *Sounder*, and *Wilson Reading System*), whereas other programs are designed for those reading at a second-grade level (*Challenger, the higher levels of Corrective Reading*) or above (*Rewards*).

As with the first-grade reading programs, the analysis of adolescent reading programs

Table 5
Adolescent Literacy Programs: General Information

Program	Number of lessons	Daily teaching time	Approximate time for completion	Minimum reading level
Challenger	160	45 min–1 hr	1 year	2nd grade
Corrective Reading (*Decoding*)	320	1 hr (period)	2 years	Nonreaders
Language!	54	Unspecified	3 years	Primer level
Laubach Way to Reading	75	1–1.5 hr	45–55 weeks	Nonreaders
REWARDS	20	40–45 min	4–6 weeks	1st grade
Sounder	92	30–50 min	4 months–2 years	Nonreaders
Voyager	124	Unspecified	1 year	1st grade
Wilson Reading System	62	40 min	1.5 years	Nonreaders

addressed program content, program coherence, and assessment. In the adolescent literacy program analysis, evaluators also examined the instruction for 1 week at the beginning and 1 week at the end of a program. Also, program content was analyzed by examining the explicitness of the phonics approach predominant in the program and the recommendations for fluency building activities. Program coherence was examined by analyzing the type of text selections available to determine how closely related the selections were to the decoding instruction provided (i.e., presence of decodable text). Also, evaluators examined the number of selections that appear early in the program and later in the program to determine whether the amount of reading changed over time. Finally, evaluators examined each program's recommendations for assessing and monitoring student performance.

Program content. Table 6 shows great variability in the number of phonics elements introduced in 1 week of a program. In three programs (*Challenger*, *REWARDS*, and *Wilson Reading System*) 18–22 phonics elements are introduced in a single week. In contrast, *Corrective Reading* introduces only 3 new phonics elements in a week of instruction. This variability also was evident in the lessons toward the end of the programs: Four programs do not introduce any elements (*Challenger*, *REWARDS*, *Voyager*, and *Wilson Reading System*), while one program (*Language!*) introduces 13 new phonics elements.

In general, most of the programs use an explicit approach to phonics instruction. *Corrective Reading*, *REWARDS*, and *Sounder* received the highest ratings based on the explicitness of their approach to decoding instruction *and* the presence of clear and consistent instructional language. In contrast, *Laubach Way to Reading* was rated lower due to the predominantly implicit approach to phonics instruction. In that

program, pictures were used as a primary strategy for teaching letter–sound relationships.

Interestingly, most of these adolescent literacy programs contain few or no instructional strategies to improve fluency. *Corrective Reading* and *REWARDS* were the only two programs to include daily fluency checks along with specific recommendations to teachers for improving fluency. For example, *Corrective Reading* suggests that teachers conduct daily repeated and timed readings. That program provides specific rate and accuracy guidelines to help teachers monitor student improvement in fluency. (See Table 6 for program content ratings.)

Program coherence. Program coherence in the adolescent literacy programs was also rated by examining the number of text selections and the relationship between decoding instruction and those text selections. Practice in reading connected text is an important component of remedial reading programs. To get a sense of how much practice each program provided, evaluators counted the number of text selections available in the weeks examined and listed the selections as short (less than 100 words) or long (more than 100 words). Early in the programs evaluators found one program that contains 11 passages (*Laubach Way to Reading*) and four programs that delay instruction in reading text selections until later in their programs (*Corrective Reading*, *REWARDS*, *Sounder*, and *Wilson Reading System*). Evaluators found that during a single week of instruction occurring relatively late in a given program, programs provided anywhere from 0 to 20 different text selections for student practice. (See Table 7.)

Unlike beginning reading programs that contain numerous types of text selections, remedial programs usually contain only those selections that are specifically written for each lesson or unit. However, even with these specially written text materials, the relationship

Table 6

Adolescent Literacy Programs, 2-Week Analysis: Program Content

Content	Challenger	Corrective Reading (Decoding)	Language!	Laubach Way to Reading*	REWARDS	Sounder	Voyager	Wilson Reading System
Phonics elements count								
(early)	22	3	6	0	19	5	6	18
(late)	0	2	13	1	0	6	0	0
Explicit phonics instruction rating	2	3	2	1	3	3	2	2
Fluency building activities rating	1	3	1	1	3	2	1	1

*Program duration less than 1 year.

Note. **Phonics elements introduced.** Evaluators counted the number of phonics elements introduced in 2 weeks of instruction during the 1st year, 1 week occurring early in the program, and a week in the 1st year of instruction.

Explicit phonics instruction. Evaluators rated the presence of explicit phonics instruction on a 3-point scale using the following criteria:

   3 Program contains an explicit phonics approach and uses phonologically based decoding instruction; instructional language is clear and consistent.

   2 Instruction includes both explicit and implicit phonics approaches; instructional language is inconsistent.

   1 Program contains an implicit phonics approach to decoding; instructional language is inconsistent.

Fluency building activities. Evaluators rated the fluency building activities in programs on a 3-point scale according to the following criteria:

   3 Program provides specific instructional recommendations for developing fluency.

   2 Nonspecific recommendations for developing fluency are included in the program.

   1 No fluency building activities are included in the program.

between the phonics instruction and the text selections is not always coherent. In some programs, evaluators detected a strong relationship between the phonics taught and the text selections provided (*Challenger*, *Corrective Reading*, *Language!*, *Sounder*, and *Wilson Reading System*), whereas in other programs there did not appear to be any observable relationship. For example, in *Laubach Way to Reading*, students are given reading passages beginning in the first lesson that are not decodable based on the phonics instruction presented at that time. The passages include letter–sound correspondences that have not been introduced and sight words that are taught *after* the passage reading occurs.

Assessment. As in the evaluation of beginning reading programs, the analysis of assessment options in remedial programs addressed five features: (a) program placement tests, (b) per-

---

Table 7
Adolescent Literacy Programs: Program Coherence

	Challenger Adult Reading Series	Corrective Reading (*Decoding*)	Language!	Laubach Way to Reading*	REWARDS	Sounder	Voyager	Wilson Reading System
Number of reading passages								
(Early program)	5 (long)	0	3 (short)	4 (short) 7 (long)	0	0	5 (short)	0
(Late program)	5 (long)	5 (long)	3 (long)	4 (long)	5 (short) 5 (long)	16 (short) 4 (long)	0	1 (short) 3 (long)
Program coherence rating	3	3	3	1	2	3	2	3

*Program duration less than 1 year.

Note. **Reading passages.** Evaluators counted the number of passages included in the week's activities. Passages were identified as short (less than 100 words) or long (over 100 words).

Program coherence. Evaluators rated program coherence on a 3-point scale using the following criteria:

3 The text selections are designed to provide practice on specified phonics elements.

2 Practice is provided on difficult words found in the text selection.

1 The text selections are minimally related to the decoding instruction.

formance standards (proficiency levels), (c) recommendations for remediation of students who are not meeting those standards, (d) recommendations for acceleration through the program for those students who do well on assessments, and (e) clarity of the directions for using the assessments.

Table 8 summarizes findings from the analysis of the assessment options. All of the adolescent literacy programs included a placement test. However, the placement tests for three programs (*REWARDS*, *Sounder*, and *Wilson Reading System*) provide only information about the appropriateness of the program and not placement options within the program. Unlike the first-grade programs, the majority of the

adolescent literacy programs include performance standards for some or most assessment options. Although none of the programs provide recommendations for acceleration, in general, the adolescent literacy programs offer more guidance to teachers for how to use the assessment information to make instructional decisions than do the beginning reading programs. Two of the programs consistently provide recommendations for remediation (*Corrective Reading* and *Sounder*). The clarity of directions for assessment options was rated higher in adolescent literacy programs than in beginning reading programs. The majority of the assessment instruments in the programs for struggling readers were found to include explicit directions and to be easily accessible.

Table 8
Adolescent Reading Program: Assessment/Coherence

Program	Placement test rating	Performance standards specified rating	Recommendations for remediation rating	Recommendations for acceleration rating	Clarity of directions to teachers rating
Challenger	3	1	1	1	2
Corrective Reading (*Decoding*)	3	3	3	1	3
Language!	3	2	2	1	3
Laubach Way to Reading	3	2	2	1	3
REWARDS	2	2	1	1	3
Sounder	2	3	3	1	3
Voyager	3	1	1	1	3
Wilson Reading System	2	3	1	1	2

Summary Comments

For the beginning reading curriculum analysis, the primary goals were first to explore procedures for generating research-based evaluation criteria, and second to determine if those criteria could be used to differentiate the programs from one another. Curriculum evaluation criteria were generated from the available research on beginning reading. However, those criteria did not successfully distinguish the

---

Table 8 (continued)

Adolescent Reading Program: Assessment/Coherence

Note. **Placement test.** Evaluators rated programs on a 3-point scale using the following criteria:

   3 Program provides a placement test including multiple entry levels for student placement in the program according to the test results.

   2 Program provides a placement test, but does not have multiple entry levels for student placement.

   1 Program does not provide a placement test.

Performance standards specified. Evaluators examined assessment options to determine whether student performance standards are available. Evaluators rated programs on a 3-point scale using the following criteria:

   3 Performance standards are available for most assessment options.

   2 Performance standards are available for some assessment options.

   1 Performance standards are rarely available for assessment options.

Recommendations for remediation and acceleration. Evaluators examined assessment options to determine whether specific recommendations for remediation and/or acceleration were available. Evaluators rated programs on a 3-point scale using the following criteria:

   3 Recommendations are consistently available.

   2 Recommendations are sometimes available.

   1 Recommendations are rarely available.

Clarity of directions to teachers. Evaluators examined the clarity of directions to teachers to determine if they specify when and how to use the various assessment tools. Evaluators rated programs on a 3-point scale according to the following criteria:

   3 Directions to teachers for using assessment tools are explicit and easily accessible.

   2 Directions to teachers for using assessment tools are not consistently explicit and accessible.

   1 Directions to teachers for using assessment tools are vague and difficult to access.

---

major beginning reading programs from each other. One reason may be simply that recently published reading programs are *not* very different. Since the criteria for the recent English/Language Arts adoptions in both Texas (1999) and California (2001) require that beginning reading programs use explicit, systematic phonics instruction, most commercially published comprehensive reading programs now use explicit and systematic phonics instruction. Therefore, the decoding instruction in these programs is more alike than it is different. Another reason for the apparent similarity among beginning reading programs may be that the criteria were not adequate to detect subtle program differences. More research on specific instructional practices may be needed before these types of evaluation criteria can be more useful. For example, research on effective sequences for using decodable text and less-decodable anthology selections would provide a helpful basis for evaluating how a program organizes the various text selections students are asked to read.

Unlike beginning reading programs, the remedial programs not only differed markedly in their general organizational structure (e.g., number of lessons, entry skills required), but also dramatically in their program content. Some programs include specific fluency building activities while others do not; some programs include specific recommendations for remediation based on progress monitoring while others do not. As mentioned above, none of the remedial programs include specific recommendations for acceleration based on student performance. Differences in the program design suggest that some programs may be more appropriate than others in addressing different student needs.

In examining the voluminous teacher manuals for beginning reading programs, evaluators found that these programs contain many more instructional activities than do the remedial programs. During the beginning reading cur-

riculum analysis, evaluators counted the number of independent activities found in a single week of instruction. In this week, they found that programs provide teachers with approximately 150–225 different activities from which to choose. Having such a voluminous selection of activities from which to choose in the beginning reading programs without guidance in how to select the most important ones can be overwhelming for new or struggling teachers. Without such guidance, teaching from these programs resembles trying to find information on the World Wide Web without benefit of a search engine that locates, prioritizes, and organizes (e.g., Google).

Beginning reading programs and remedial programs also differ in the quality and quantity of their assessment recommendations. Remedial programs appear to provide more guidance to teachers in how to use the results from student assessments to make instructional decisions. The remedial programs also include more initial placement options than do comprehensive reading programs. One reason that these programs contain more and better assessment systems may be due to the fact that the programs were designed for those students who had difficulty learning to read in a basal reading program. Remedial reading programs offer more systematic monitoring and better placement options in order to provide teachers with additional resources for helping them resolve their students' reading problems.

Future Directions

The work represented here was generated and reported in much the same way as the analysis of decodable text in Stein et al. (1999). While there is no evidence to suggest an optimal percent of decodable text or adequate number of decodable texts for first-grade programs, the simple analysis of the presence or absence of decodable text in Stein et al. contributed to the larger discussion of the role that text selections play in effective beginning reading instruction. The projects reported in this

paper have contributed to the larger discussion of how programs can be systematically evaluated using research-based criteria.

Educators interested in developing systematic curriculum evaluation instruments need to generate evaluation criteria, validate those criteria, develop objective evaluation procedures derived from the criteria, and determine the reliability of their instruments. Moreover, educators need to design the evaluation process with practitioners in mind. That is, the evaluation instruments need to be user friendly and provide useful information.

The two projects presented here have also contributed to more extensive conversations on the role that reading curricula play for struggling students. In fact, the project focused on the evaluation of remedial reading curricula has resulted in the development of a research-based curriculum evaluation instrument for remedial reading programs that is currently undergoing reliability and validity evaluation in four states.

Finally, using research to drive the development of evaluation criteria for examining reading programs is an avenue for persuading publishers to improve their reading programs. By providing teachers and administrators with current information about scientifically based practices through the use of research-based curriculum evaluation instruments, practitioners have a means of communicating to publishers the importance of including those practices in commercially developed programs. An example of the influence curriculum evaluation can have on publishers is evident in the area of reading fluency. Once a strong relationship between reading fluency and reading comprehension was established through research, disseminated widely (National Reading Panel, 2000; Snow et al., 1998), and reflected in the evaluation criteria of several states, publishers began to include fluency building activities in their comprehensive reading programs where there were none pre-

viously. For example, reading fluency was not listed in the index of the 1991 edition of one major reading program; in the 2003 edition of that program, there are 18 listings under the heading Fluency. In addition, many educational publishers have begun to develop supplementary fluency building programs.

Clearly, effective reading instruction requires more than choosing a well designed commercially developed reading program. However, evidence from comprehensive school reform efforts (American Institutes for Research, 1999) and the school change literature (Gersten & Brengelman, 1996) suggests that when teacher preparation and professional development are anchored to the use of scientifically based and well designed reading programs, student achievement can be enhanced. Common sense suggests that the use of a reliable and valid research-based curriculum evaluation instrument in evaluating both beginning and remedial programs is likely to have benefits far beyond the selection of a reading program.

References

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

American Institutes for Research. (1999). *An educators' guide to schoolwide reform.* Arlington, VA: Educational Research Services.

Baker, S. K., Kame'enui, E. J., Simmons, D. C., & Stahl, S. A. (1994). Beginning reading: Educational instruments for diverse learners. *School Psychology Review, 23,* 372–391.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25*(9), 6–8.

Comas, J. C., & Farr, R. (1989). Training preservice teachers to analyze and evaluate textbooks. *Book Research Quarterly, 5,* 5–15.

Donahue, P. L., Finnegan, R. J., Lutkus, A. D., Allen, N. L., & Campbell, J. R. (2001). *The nation's report card: Fourth-grade reading 2000.* Retrieved May 14, 2004, from http://nces.ed.gov/nationsreportcard

Fuchs, L. S. (1988). Effects of computer-managed instruction on teachers' implementation of systematic monitoring programs, student achievement, and student awareness of learning. *Journal of Educational Research, 81,* 294–304.

Gersten, R., & Brengelman, S. U. (1996). The quest to translate research into classroom practice: The emerging knowledge base. *Remedial and Special Education, 17,* 67–75.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and 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.

Heibert, E. H., & Taylor, B. M. (2000). Beginning reading instruction: Research on early interventions. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 455–482). Mahwah, NJ: Erlbaum.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437–447.

Juel, C., & Roper-Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly, 20,* 134–152.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25,* 215–227.

Lyon, G. R. (1996, Spring). Learning disabilities. *The Future of Children, 6*(1), 54–76.

Mesmer, H. A. E. (in press). Text decodability and the first grade reader: Does decodability result in greater application of phonics instruction during text reading. *Reading and Writing Quarterly.*

National Assessment of Educational Progress. (2002). *NAEP 2002 reading report card: National and state highlights.* Retrieved May 14, 2004, from http://nces.ed.gov/nationsreportcard/reading/results2002/

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: Reports of the subgroups.* (NIH Publication No. 00-4754). Bethesda, MD: National Institute of Child Health and Human Development.

No Child Left Behind. (2001). Retrieved May 14, 2004, from http://www.ed.gov/about/offices/list/oese/legislation.html#leg

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407.

Stein, M., Johnson, B., Boutry, S., & Bortleson, C. (2000). *An analysis of the decoding instruction in six first grade reading programs.* Unpublished manuscript.

Stein, M., Johnson, B., & Gutlohn, L. (1999). Analyzing beginning reading programs: The relationship between decoding instruction and text. *Remedial and Special Education, 20,* 275–287.

Stein, M., Stuen, C., Carnine, D., & Long, R. M. (2001). Textbook evaluation and adoption practices. *Reading and Writing Quarterly, 17*(1), 5–24.

Texas Education Agency. (2000). *Report of decodable text in conforming first grade language arts-reading instructional materials.* Retrieved May 14, 2004, from http://www.tea.state.tx.us/shoe/schedule/2001/decodabletext.html.

Author Note:

Marcy Stein is currently an author of the *Open Court Reading Program.* However, she was not an author when she conducted the program analysis reported here.

Appendix
Reading Programs Analyzed

First-Grade Reading Programs

Collections. (2000). Orlando, FL: Harcourt, Inc.

Journeys. (2000). Worthington, OH: SRA

Literacy Place. (2000). New York, NY: Scholastic, Inc.

McGraw-Hill Reading. (2001). New York, NY: McGraw-Hill

Open Court Reading. (2000). Worthington, OH: SRA/McGraw-Hill

Scott Foresman Reading. (2000). Glenview, IL: Addison-Wesley

Adolescent Literacy Programs

Challenger. (1988). Syracuse, NY: New Readers Press

Corrective Reading—Decoding. (1999). Worthington, OH: SRA/McGraw-Hill

Language! (1995). Longmont, CO: Sopris West

Laubach Way to Reading. (1991). Syracuse, NY: New Readers Press

REWARDS. (1999). Longmont, CO: Sopris West

Sounder. (1995). Seaside, OR: Ronald F. Smith (Author)