Exceptional Children

Vol. 75, No. 3, pp. 263-281.

Repeated Reading Interventions for Students With Learning Disabilities: Status of the Evidence

DAVID J. CHARD
Southern Methodist University

LEANNE R. KETTERLIN-GELLER
University of Oregon

SCOTT K. BAKER
Pacific Institutes for Research/University of Oregon

CHRISTIAN DOABLER

CHANISA APICHATABUTRA
University of Oregon

ABSTRACT: *For students with or at risk for learning disabilities, developing fluency with reading connected texts remains a formidable challenge. In response, teachers often use repeated reading practices designed to provide students with multiple exposures to the same words. This study examined research focused on determining the efficacy of repeated reading approaches for improving reading fluency for students with or at risk for learning disabilities. Studies employed experimental/quasi-experimental and single-subject research designs. Results suggest that repeated reading is not supported by rigorous research as defined by the quality indicators used and, therefore, is not an evidence-based practice based on those criteria for students with and at risk for learning disabilities. Implications for future research and for practice are discussed.*

Learning to read remains a hallmark skill that essentially defines the degree of success students can achieve academically throughout their school career. For students who learn to read early and well, the occasion is set for them to flourish in vocabulary and language development, comprehension, and content area learning (Stanovich, 1986). Students with or at risk for learning disabilities (LD) are most often identified for special education services due to their measurable diffi-

culties with early reading acquisition, specifically difficulties with word identification. Basic deficits in alphabetic coding are the underlying cause of these difficulties, with deficits most often attributed to deficiencies in reading-related cognitive abilities, namely phonological skill deficiencies (Stanovich; Torgesen et al., 2001). Below-average readers frequently experience difficulties with metaphonological and metalinguistic tasks, supporting the notion that they experience a phonological core deficit (e.g., Snowling, 2000; Stanovich & Siegel, 1994; Vellutino et al., 1996).

Students who experience a phonological core deficit are characterized by poor phonological awareness and verbal short-term memory as well as below-average speed of access to phonological information in long-term memory (Adams, 1990; Lipka, Lesaux, & Siegel, 2006). Decoding difficulties limit students' opportunities to read texts, decrease students' exposure to words, limit vocabulary learning, and hamper the development of content-area expertise through reading comprehension. Because of the effect this core deficit can have on long-term reading achievement, early reading interventions generally focus on improving students' phonological awareness, decoding skills, sight word identification, and fluency development. Evidence suggests that this focus can be particularly fruitful for many students (Mathes, Howard, Allen, & Fuchs, 1998; McMaster, Fuchs, Fuchs, & Compton, 2005; Simmons et al., 2008; Vellutino, Scanlon, & Tanzman, 1998). Intervention efforts for students with LD in reading tend to focus early on the development of decoding skills followed by fluency building in connected texts. More specifically and important for the present review, fluency-building interventions have often featured repeated reading activities aimed at improving the speed and accuracy of students' text reading (Chard, Vaughn, & Tyler, 2002; Therrien, 2004). Repeated reading offers students an opportunity to read and reread the same text multiple times and is implemented in a variety of formats including partner reading, reading to an older peer or family member, or reading with an audiotape.

FLUENCY AND ITS ROLE IN READING PROFICIENCY

Proficient reading can be characterized as a multifaceted process including at least two activities: word identification or decoding and comprehension. The latter, understanding an author's message, involves making inferences, responding critically, and so on, and it always requires attention. In order for a reader to understand what is being read, she cannot focus attention on both word identification and comprehension. A reader who is not fluent can alternate attention between the two processes; however, this makes reading laborious. If attention is consumed by decoding words, little or no capacity is available for the attention-demanding process of comprehending. Therefore, automaticity of decoding—a critical component of fluency—is essential for high levels of reading achievement (Ehri, 1995; LaBerge & Samuels, 1974).

If attention is consumed by decoding words, little or no capacity is available for the attention-demanding process of comprehending. Therefore, automaticity of decoding—a critical component of fluency—is essential for high levels of reading achievement.

Several theoretical frameworks have been posited to explain the role of fluency in reading proficiency. Perfetti's (1985) proposed "verbal efficiency theory" highlights the importance of lower level lexical skills in reading and explains the impact of fluent processing of information to reading comprehension. The verbal efficiency theory suggests that lower level processes (e.g., word identification) must reach a particular threshold level before higher level processes (e.g., comprehension) can be performed simultaneously during reading. When lower level processes are inefficiently performed, higher order processes are compromised in an attempt to compensate. Perfetti's theory assumes that resource demands can be reduced through learning and practice.

Logan (1988) proposed an alternative, memory-based theory of fluency, the "instance theory of automatization," suggesting that automaticity and fluency are based on memory retrieval. The instance theory of automatization is based on three important assumptions: (a) obligatory encoding, (b) obligatory retrieval, and (c) instance representation (Logan, 1997). *Obligatory encoding* refers to focusing attention on a stimulus (e.g., a word) and storing details of that stimulus in memory. *Obligatory retrieval* suggests that merely attending to the stimulus is sufficient to retrieve previous exposures of the stimulus or similar stimuli from memory. *Instance representation* refers to the coding and storage of each memory trace of experiences with a stimulus in memory. Information recall is automatic when it relies on retrieval of "stored instances," the theoretical memory traces laid down in the brain each time a task is executed. Therefore, the strength of the memory trace is increased with the number of times the task is performed.

Logan (1997) suggested that the level of automaticity is dependent on the amount of practice, the level of consistency in the task environment, and the number of relevant instances of the task recorded in memory. As the reader's knowledge expands and becomes increasingly accurate, performance becomes more reliant on memory retrieval and less on problem solving (Logan, 1997). Applied to reading fluency, if a word is read frequently, the cumulative practice results in an increased likelihood that the word will be recognized when encountered later and the speed will increase. We find the combination of Perfetti's (1985) verbal efficiency theory and Logan's (1988) instance theory to provide intuitive support for the notion of repeated reading as an intervention for fluency building. As students repeatedly read the same content, it is likely that they will practice the same words multiple times, increasing the likelihood they'll be able to automatically retrieve those same words in future exposures. Simultaneously, they reduce the attention required to read the words and can focus more intently on the meaning of what they are reading.

Empirical Research on Fluency Interventions

A report of the National Reading Panel (NRP; 2000) significantly elevated attention to fluency. NRP's review reflected the position that "fluency develops from reading practice" (p. 3-1), and devoted much of this meta-analysis to research support for two major approaches to providing students with reading practice: (a) repeated oral reading practice or guided repeated oral reading practice, and (b) independent or recreational reading. NRP identified 98 articles that met the criteria for inclusion. The panel determined that the mean weighted effect size for guided oral repeated reading was 0.41, indicating that this procedure "has a moderate impact on the reading achievement of the types of students participating in these studies" (p. 3-17). It concluded that there is sufficient experimental evidence supporting the use of repeated reading procedures, but insufficient experimental research to suggest that increasing independent reading will increase fluency or reading achievement.

Research on Fluency Interventions and Students With Learning Disabilities

Fluency is an essential skill for all students; students with reading/learning disabilities are most at risk for presenting difficulties in fluency (Meyer & Felton, 1999). Based on its finding that repeated oral reading was associated with improved fluency for most students, NRP (2000) recommended teachers begin including repeated reading activities in their reading instruction. However encouraging this recommendation appeared, NRP's findings may not generalize to students with LD—those with the most significant reading problems. For this reason, researchers have focused on synthesizing the findings from fluency interventions specifically for students with identified LD. Many of the approaches to improving fluency for students with or at risk for LD could be categorized as focusing on repeated reading (Meyer & Felton); partner reading (e.g., Arreaga-Mayer, Terry, & Greenwood, 1998; Fuchs, Fuchs, Mathes, & Simmons, 1997); or other procedures which support repeated reading.

WHY FOCUS ON REPEATED READING?

The purpose of our review was to apply the standards of evidence recommended by Horner et al. (2005) and Gersten et al. (2005) to a body of research examining the effects of repeated reading on the reading achievement of students with and at risk for LD. We define *repeated reading* as any intervention procedure that requires students to read passages in connected text or word lists more than once. We selected repeated reading as our focus for several reasons. First, research on the effects of repeated reading has been well documented in the literature since the 1970s. Researchers have conducted numerous studies on the effects of this intervention for students in a variety of different contexts and with differing characteristics, such as LD, emotional and/or behavioral disorders, and average achieving students. In addition, several literature syntheses have been published documenting the research support for this practice (see Chard et al., 2002; Meyer & Felton, 1999) and meta-analyses have found positive outcomes on students' reading achievement as a result of implementing this intervention (Therrien, 2004). Therrien reported effect sizes from four studies of $d = 0.75$ on measures of fluency and 0.73 on measures of comprehension for students with LD. Second, in the field, repeated reading is widely held as an evidence-based practice with documented effectiveness. To many practitioners, there is little doubt as to whether or not repeated reading qualifies as a research-based practice that is justifiably implemented in support of student learning. Repeated reading deviates from "craft-based" practices that support reading development in that it is based on strong theories of learning and cognition (Ehri, 1995; Logan, 1988, 1997; Perfetti, 1985; Stanovich, 1986). Finally, because of the body of research available, we anticipated finding multiple studies in the literature over the past 30 years documenting the effects of repeated reading. Because the focus of this article is limited to students with LD or at risk for LD, we wanted to select a practice backed by research conducted with students with this disability.

Although these reasons would suggest that repeated reading has already been proven to be an "evidence-based" practice, as of yet the research on repeated reading has not been evaluated against the rigorous quality standards needed to justify the title of "evidence-based." Therefore, the following research question guided our investigation: Is the research base supporting the effectiveness of repeated reading based on high-quality standards of single-subject and experimental/quasi-experimental research that would lead to the determination that repeated reading is an evidence-based practice?

METHOD

To address this research question, we focused on four activities: (a) identification of intervention studies that examined the impact of repeated reading for students with or at risk for LD, (b) application of the quality indicators to existing literature on repeated reading, (c) identification of studies that met acceptable quality standards, and (d) determination of whether or not repeated reading could be considered "evidence-based" given the criteria established by Horner et al. (2005) and Gersten et al. (2005).

Identification of Studies

Prior to conducting the search of studies, we established a priori criteria for what would constitute a repeated reading intervention. The intervention was considered eligible if students were required to read connected texts or word lists more than once with the intention of improving rate and accuracy, and the intervention did not include other instructional components that were focused on other aspects of reading (e.g., comprehension, vocabulary development).

The selection of studies for review was a multistep process that began with an exhaustive search of the literature. First, we searched ArticleFirst, Educational Resources Information Center (ERIC), Google Scholar, OVID PsycINFO, and WorldCat electronic databases to locate relevant studies conducted between January 1975 and December 2006. Our search began with studies conducted in 1975 because it corresponded with the passage of the Education for All Handicapped Children Act of 1975 (Public Law 94-142). We included the following literature search terms: *re-*

peated reading, reading automaticity, fluency, reading speed, reading fluency, reading rate, reading disability, reading strategies, reading improvement, reading aloud, reading practice, reading instruction, assisted reading, oral reading, paired reading, single-subject design, group design, reading difficulties, students with learning disabilities, elementary education, secondary education, and *primary education.* Electronic searches used multiple combinations and sequences of the literature search terms.

Second, we conducted an ancestral search using the reference lists from pertinent studies conducted by Chard et al. (2002); Kuhn and Stahl (2003); Meyer and Felton (1999); Swanson, Hoskyn, and Lee (1999); and Therrien (2004). Third, we conducted a manual search of recent literature in several major journals (from 2004 to December 2006) of special, remedial, elementary, and secondary education: *American Journal of Education; Assessment for Effective Intervention; Behavioral Disorders; Cognition and Instruction; Educational Psychology; Educational Researcher; Elementary School Journal; Exceptional Children; Journal of Applied Behavior Analysis; Journal of Education and Behavior Statistics; Journal of Educational Psychology; Journal of Educational Research; Journal of Experiential Education; Journal of Experimental Education; Journal of Learning Disabilities; Journal of Literacy Research; Journal of Negro Education; The Journal of Special Education; Journal of Special Education Technology; Journal of Speech, Language, and Hearing Research; Learning Disabilities Research & Practice; Learning Disability Quarterly; Memory and Cognition; Mental Retardation; Peabody Journal of Education; Reading Horizons; Reading Improvement; Reading Research and Instruction; Reading Research Quarterly; Reading Teacher; Remedial and Special Education; School Psychology Review;* and *Scientific Studies of Reading.*

This process identified 92 articles pertaining to repeated reading. We applied additional criteria as a secondary screener to these studies: The studies should focus on the effects of repeated reading for students with or at risk for LD in reading in K to 12 settings. Studies with subjects who were performing below average in reading were not included unless the participants could be identified as having LD or being at risk for LD. We refrained from making any judgments about risk factors based on the achievement data reported and relied on the authors' documentation of status. Additional criteria included the isolation of the independent variable. We omitted studies if repeated reading was studied in conjunction with other practices and was not isolated as a level of the independent variable because it was not possible to examine the effects of repeated reading on student performance. This screening process resulted in 11 single-subject and experimental or quasi-experimental research studies that met all of our criteria (see Tables 1 and 2).

Application of Quality Indicators

We limited the examination of repeated reading to single-subject and experimental and quasi-experimental research studies. Focusing on suggestions provided by Horner et al. (2005) and Gersten et al. (2005), we adapted the quality indicators of rigorous research to create rubrics for evaluating the reported methodological rigor of the studies. The rubric encompasses the essential indicators of high-quality research proposed by each group of authors but provides a continuum for evaluation. Because most research studies implement rigorous methodological standards to varying degrees, we decided to represent this variance on a Likert scale. As such, instead of considering the indicators as dichotomous qualities of a study, we created a 4-point scale to reflect the variable nature of conducting and documenting research. We created separate rubrics for single-subject and experimental/quasi-experimental research (see Figures 1 and 2).

Interpreting the Quality Indicators. We created ratings for the levels of rigor for each indicator by evaluating the study for each component associated with the quality indicators proposed by Horner et al. (2005) and Gersten et al. (2005). A 1-point rating generally reflected a quality indicator component that was not documented in the published article. For example, an article reporting on an experimental or quasi-experimental study received a component score of 1 for the effect size calculation component (included in the data analysis quality indicator) if no effect size was reported. In contrast, a 4-point component rating for a particular quality indicator indicated that the article met the full criteria of the component as

TABLE 1

Average Ratings for Single-Subject Research Articles on Repeated Reading

	Quality Indicators						
Study	*Participants/ Setting*	*Dependent Variable(s)*	*Independent Variable*	*Baseline*	*EC/Internal Validity*	*External Validity*	*Social Validity*
Begeny, Daly, & Valleley (2006)	1.67[a]	3.50	2.33	2.50	2.67	1.00[a]	3.00[a]
Chafouleas, Martens, Dobson, Weinstein, & Gardner (2004)	1.83[a]	2.90[a]	2.50[a]	2.75	3.00	1.00[a]	3.13[a]
Daly & Martens (1994)	2.33[a]	3.10	2.67	3.50	2.00[a]	2.00	3.38
Freeland, Skinner, Jackson, McDaniel, & Smith (2000)	3.00	3.30[a]	3.17	2.75[a]	3.00	1.00[a]	3.13
Strong, Wehby, Falk, & Lane (2004)	2.67	3.40	3.33	3.50	3.33	1.00[a]	3.13
Weinstein & Cooke (1992)	2.33[a]	3.10[a]	2.67	3.00	3.17	1.00[a]	3.00

Note. Selection based on criteria proposed by Horner et al. (2005). EC = experimental control.
[a]Article received a 1-point score for at least one subcomponent in this methodological category.

established by Horner et al. or Gersten et al. For example, an article that documented 80% or greater interobserver agreement for data collection using single-subject research methods received 4 points for the data collected on the reliability or interobserver agreement component of the dependent variable quality indicator.

We created 2- and 3-point ratings to reflect the range of reported research findings. A 2-point rating of a component typically reflected a limited description of the methodological procedures used in the study. For example, if the description of the participants in an experimental or quasi-experimental study provided some information about the interventionists or teachers but no information about their comparability across treatment groups, the article received a 2-point rating on the information about interventionists or teachers component included in the description of participants quality indicator. Because some

TABLE 2

Average Ratings for Experimental and Quasi-Experimental Research Articles on Repeated Reading

	Essential Quality Indicators			
Study	*Description of Participants*	*Intervention/Comparison Conditions*	*Outcome Measures*	*Data Analysis*
Mathes & Fuchs (1993)	3.67	3.00	3.75	3.75
O'Shea Sindelar, & O'Shea (1987)	2.50[a]	2.33[a]	2.75	2.25[a]
Rashotte & Torgesen (1985)	2.50[a]	2.33[a]	2.50	1.50[a]
Sindelar, Monda, & O'Shea (1990)	2.33	1.83[a]	2.75	1.75[a]
Young, Bowers, & MacKinnon, (1996)	2.33[a]	2.00[a]	3.00	2.25[a]

Note. Selection based on criteria proposed by Horner et al. (2005).
[a]Article received a 1-point score for at least one subcomponent in this methodological category.

FIGURE 1

Quality Indicators of Single-Subject Research Articles and Reports

Participants and Setting	1	2	3	4	Score
Sample characteristics (e.g., age, gender, disability, diagnosis)	No detail provided	Limited detail provided	Some detail provided	Ample detail provided	
Process for selecting participants	No description of selection process	Procedures described but not appropriate and/or with limited detail	Procedures described are appropriate but minimally described	Procedures were appropriate and adequately described	
Critical features of the physical setting	No description provided	Limited description provided	Some description provided	Detailed description provided to allow replication	
Dependent Variable	1	2	3	4	Score
Description of dependent variable	No description provided	Limited description provided	Some description provided but not operational	Operational description provided	
Measurement procedure	No procedure provided or no quantifiable variables	Procedure provided but no quantifiable variables	Procedure provided but only some variables quantifiable	Procedure provided and all variables quantifiable	
Measurement validity and description	No valid measures and description not replicable	No valid measures or description not replicable	Some measures valid; description is replicable	Measures are valid and description is replicable	
Measurement frequency	No repeated measures	Measurement repeated but very infrequently	Measurement repeated but infrequently	Measurement repeated frequently	
Data collected on reliability (minimal standards: IOA = 80%; Kappa = 60%)	No reliability data reported	Reliability data incorrectly collected or analyzed	Reliability data reported but minimal standards not met	Reliability data reported and minimal standards met	
Independent Variable	1	2	3	4	Score
Description of independent variable	Only name or vague description of IV provided	IV is described with little detail	Major components of IV provided with some detail (e.g., scripts provided)	All components of IV described in detail with efforts to communicate precision	
IV manipulation	IV is provided with no control	Little control exercised (e.g., monitor, scripts)	Condition assignment is planned	Random assignment to condition	
Fidelity of implementation	No measure of fidelity	Fidelity is monitored but not directly	Fidelity is monitored directly, but at large component level	Fidelity is monitored in detail with corrections provided when necessary	

continues

FIGURE 1 *(Continued)*

Baseline	1	2	3	4	**Score**
DV measurement	DV not measured objectively	DV measured infrequently; data is missing or not stable	DV measured frequently but not stable	DV measured frequently and is stable before intervention	
Description of baseline condition	No description of baseline	Vague description of baseline	Baseline description detailed but limited	Baseline description detailed and extensive	
Experimental Control/ Internal Validity	1	2	3	4	**Score**
Design demonstrates experimental effect	No demonstration of experimental effect	Only one demonstration of experimental effect	More than one demonstration of experimental effect	Three or more demonstrations of experimental effect	
Design controls for common threats to internal validity (e.g., elimination of rival hypotheses)	No control for threats to validity	Few threats controlled	Most threats controlled	All threats controlled	
Pattern of results	Results do not suggest experimental control	Results suggest a change in trend, level, or variability	Results document a change in trend, level, or variability	Results document a pattern of experimental control	
External Validity	1	2	3	4	**Score**
Replication of effects (e.g., across participants, settings, or materials to establish external validity)	No effors to replicate efforts	Few replications attempted	Some replication attempted	Multiple replications across variables	
Social Validity	1	2	3	4	**Score**
Importance of DV	No importance			Important	
Importance of magnitude of change in DV	No importance	Somewhat important	Important	Very important	
Practicality and cost effectiveness of implementation of IV	Impractical and not cost effective	Either practical or cost effective, but not both	Some evidence of practicality and cost effectiveness	Practical and cost effective	
Typical nature of implementation of IV	IV implemented in atypical manner	IV implemented either in typical context or by typical agent, not both	Implementation extended in somewhat typical contexts and with a somewhat typical agent (e.g., certified teacher)	Implementation extended in typical contexts with typical agents (e.g., the certified teacher)	

Note. Based on quality indicators proposed by Horner et al. (2005) and Gersten et al. (2005). IOA = interobserver agreement; IV = independent variable; DV = dependent variable.

FIGURE 2

Quality Indicators of Experimental and Quasi- Experimental Research Articles and Reports

Description of Participants	1	2	3	4	Score
Information about diagnosis of disability or difficulty	No evidence and/or description	Little evidence and/or description	Some evidence and/or description	Ample evidence and/or description	
Samples are comparable across conditions on relevant characteristics	No procedures for comparability described	Procedures described but not appropriate	Procedures appropriate but minimally described	Procedures appropriate and adequately described	
Information about interventionists or teachers; comparability across conditions	No information or description provided; no information about comparability across groups	Some information or description provided; no information about comparability across groups	Some information or description provided; some information about comparability across groups	Sufficient information or description about interventionists provided; comparable across groups	
Description and Implementation of Intervention and Comparison Conditions	1	2	3	4	Score
Description of intervention and implementation procedures	Minimal description provided; no details	Some description provided; limited details	Some description provided; general details lacking specificity for replication	Description clear and specific for replication	
Description of fidelity of implementation procedures	No information provided	Some information provided; evaluation and effects on intervention impact not described	Some information provided; evaluation and effects on intervention impact minimally described	Sufficient information provided; evaluation and effects on intervention impact described	
Description of comparison condition activities	Minimal description provided; no details	Some description provided; limited details	Some description provided; general details lacking specificity for replication	Description clear and specific for replication	

continues

FIGURE 2 *(Continued)*

Outcome Measures	1	2	3	4	Score
Multiple measures or measures of generalized performance (Were multiple measures used to measure the DV? Were measures of generalized performance used?)	Only used measure tightly aligned to intervention	Used measure tightly aligned to intervention along with one other measure	Used measure of generalized performance	Used multiple measures and measure of generalized performance	
Appropriateness of data collection times	No information about timing provided	Timing of administration of outcome measures not appropriate	Timing of administration of outcome measures somewhat appropriate	Timing of administration of outcome measures appropriate	
Data Analysis	**1**	**2**	**3**	**4**	**Score**
Data analysis linked to research questions/ hypotheses; considered the unit of analysis	No information about data analysis provided	Data analysis techniques not appropriate given the research questions/ hypotheses; unit of analysis may or may not have been appropriate	Data analysis techniques mostly appropriate given research questions/ hypotheses (alternate methods could be used that were more elegant); used appropriate unit analysis	Data analysis techniques appropriate given the research questions/ hypotheses; used appropriate unit of analysis	
Effect size calculation	Effect sizes not reported	Effect sizes reported but not interpreted	Effect sizes reported but not accurately interpreted	Effect sizes reported and accurately interpreted	

Note. Based on quality indicators proposed by Horner et al. (2005) and Gersten et al. (2005). IOA = interobserver agreement; DV = dependent variable.

information was provided, the component should score higher than 1 point, but limited details still make this description unacceptable. Conversely, if more details were provided but lacked specificity for replication or generalization, the article received 3 points for that component. For example, a single-subject research study that reported monitoring the fidelity of implementation, a component of the independent variable quality indicator, might receive 3 points because although detailed information was provided about monitoring, the authors may not have documented correction procedures. The study should not receive 4 points for the independent variable quality indicator because fidelity of implementation information was not provided at the level of precision proposed by Horner et al. (2005). However, because much of the necessary information was present, we considered this acceptable.

Members of the author team not involved in creating the rubrics evaluated them to ensure alignment with the quality indicators proposed by Horner et al. (2005) and Gersten et al. (2005). In addition, we applied the rubric to a sample article

to calibrate the raters' interpretations of the scale. After independently reviewing the rubric and applying it to a sample article, we made several adjustments to clarify the expectations of the levels of rigor associated with each score.

Identification of Studies That Met Acceptable Quality Standards

Two raters using the corresponding rubric independently evaluated single-subject and experimental/quasi-experimental research articles. The raters included a researcher with 4 years experience teaching and researching in the field of applied measurement and assessment and a second-year doctoral student studying special education. Initial ratings were entered into a spreadsheet and evaluated for correspondence. In instances where raters disagreed by 2 or more points and/or one rater assigned a component score of 1, an independent third rater evaluated the quality of the component and provided an arbitration score. As such, the third rater evaluated any indicator that received a component score of 1. The arbitration score supplanted the initial rating. The independent rater was a third-year doctoral student in special education who was familiar with the rubric but not the literature in repeated reading.

Scores on the quality indicators were derived from the totals of each component associated with each quality indicator. We aggregated the scores for the quality indicators across raters based on the methodological category (e.g., description of participants, dependent variable). Combining scores at the broader level accounted for the minor variance across raters present at the component level. Any category that did not receive an average acceptable score (3 points or greater) across indicators was considered unacceptable. Moreover, any indicator that earned a component score of 1 by two or more raters (the two initial raters and/or the independent rater) was considered unacceptable.

Interrater reliability was calculated by dividing the number of exact matches on ratings at the component level by the total number of exact matches and disagreements. This resulted in reliabilities of 0.36 for single-subject studies and 0.53 for experimental/quasi-experimental studies. When calculating interrater reliability by considering agreements to include exact matches and 1-point discrepancies, reliabilities were 0.96 for single-subject studies and 0.91 for experimental/quasi-experimental studies.

Determination of Evidence-Based Practice

We used the criteria proposed by Horner et al. (2005) and Gersten et al. (2005) to make a final evaluation of repeated reading as an evidence-based practice. For single-subject research, Horner et al. noted that a practice can be labeled "evidence-based" if five studies that meet minimally acceptable methodological criteria document positive effects from implementing the practice. Of these studies, the research needs to be conducted by three different researchers from three regions of the country. In addition, the number of subjects across the five studies needs to total at least 20. According to Gersten et al., research conducted using experimental or quasi-experimental designs needs to meet all but one of the essential quality indicators to be considered high quality. These studies must also have one to four desirable indicators to be acceptable. Four acceptable-quality studies or two high-quality studies documenting the effectiveness of the practice along with a weighted effect size significantly greater than zero are needed in order to label the practice as "evidence-based." A "promising practice" requires the same numbers of studies but can have a 20% confidence interval for the weighted effect sizes that is greater than zero. We applied these criteria to the repeated reading studies we examined to determine if repeated reading could be labeled as "evidence-based" for students with or at risk for LD.

RESULTS

Quality of the Single-Subject Research Studies

We analyzed six single-subject research studies for correspondence with the quality indicators identified by Horner et al. (2005), organizing 21 components into seven methodological categories or quality indicators: description of participants and setting, dependent variable, independent variable,

baseline, experimental control/internal validity, external validity, and social validity (see Table 1).

We evaluated the description of the participants and setting by reviewing the articles' discussion of the sample characteristics, process for selecting participants, and description of critical features of the physical setting. Only one study met the minimum requirements of acceptability for this category (i.e., mean rating of 3 or greater for each quality indicator and no component rated as 1 by two or more raters). Most studies provided adequate information about the sample participants; however, the process for selecting the participants was either absent or superficially discussed. Moreover, the description of the setting lacked sufficient details to interpret the quality of the research design.

We considered documentation of the dependent variable, measurement procedures including validity and technical consideration, and frequency and reliability of data collection for each manuscript. Of the six single-subject research studies reviewed, all but one of the studies received an average score of 3 on this quality indicator. However, two studies received a score of 1 based on the lack of description of measurement validity and one study received a 1-point score because it did not discuss reliability information for data collection. Because the minimum criteria established for our review also required each study to have adequate representation of the subcomponents in the category, these three studies presented unacceptable evidence about the quality of the research for this category.

Reports of the overall description, manipulation, and fidelity of implementation of the independent variable determined if the research studies presented acceptable evidence to document the rigor of the research with regard to the quality indicator on the independent variable. Four of the six research studies did not meet the minimum criteria for acceptability in this category. Two of the studies that did not provide adequate documentation failed to provide details about the fidelity of implementation of the intervention. Other studies provided limited information about the independent variable including precise implementation techniques or scripts.

Three studies did not meet the minimum criteria needed to establish rigorous research methodology for documentation of baseline procedures primarily due to a limited description of the baseline condition. Two studies did not meet the minimum criteria for their documentation of the experimental control/internal validity. Most manuscripts reported on studies that were well designed to demonstrate experimental effects of repeated reading including observing a pattern of results due to the implementation of the intervention that documents variability, trend changes, or differences in level of performance. However, several studies provided limited information on the experimental controls employed by the researchers to address common threats to internal validity.

Only one subcomponent evaluated the adequacy of the external validity of the documented research. No study we reviewed met the minimum requirements needed to document the potential generalizability of results by replicating the study across participants, settings, or materials. It should be noted that we only considered participants with LD in this evaluation. As such, some studies may have included measures to ensure external validity across the range of participants, but did not replicate with students with or at risk for LD.

Four of the six research articles reviewed provided adequate documentation to justify the social validity of the research. We evaluated research studies on four dimensions: importance of the dependent variable, importance of the magnitude of change in the dependent variable, practicality and cost effectiveness of the independent variable, and nature of implementation of the independent variable. Two studies received 1-point scores because they used procedures and/or resources not typically available to classroom teachers (such as specially trained interventionists).

Quality of the Experimental and Quasi-Experimental Research Studies

We evaluated five research studies that used experimental or quasi-experimental research methods, categorizing nine components into the four methodological categories of quality indicator defined by Gersten et al. (2005): description of participants, description and implementation of the intervention and comparison groups, outcome

measures used, and data analytic techniques (see Table 2).

We evaluated each research study for its description of the participants—the quality of the documentation of students' demographic information including disability diagnoses procedures and comparability of students across conditions. In addition, we evaluated information about the interventionists or teachers implementing the intervention along with the comparability of their characteristics and/or credentials across conditions. One study met the minimum criteria of earning an average score of 3 across all components with no 1-point scores. The remaining studies did not adequately report information about the interventionists and/or teachers. In addition, several studies provided limited information about the diagnoses procedures used to document the participants' disabilities.

To evaluate the description and implementation of the intervention and comparison conditions, we reviewed the research studies for their overall description of the intervention and comparison conditions as well as the fidelity of implementation procedures. Four of the five studies reviewed did not meet the minimum requirements for rigorous research in this methodological category primarily due to the lack of information about the fidelity of implementation of the intervention. Most studies reviewed did not provide sufficient information to determine if the intervention was instituted as proposed by the researchers.

Three studies did not meet the minimum criteria needed to establish rigorous research methodology for documentation of the outcome measures. In most cases, the articles documented appropriate data collection times given the research design used. However, several studies did not use multiple dependent measures or measures of generalized performance to document the effects of the independent variable. Similarly, four studies did not meet the minimum requirements for data analysis. Although most studies used data analysis techniques that aligned with the research question and were applied at the appropriate unit of analysis, only one study reported effect sizes. As such, all other studies received a 1-point score because this information was missing.

DISCUSSION

Evaluation of Evidence Base

To determine if repeated reading could be classified as an evidence-based practice, we applied the criteria established by Horner et al. (2005) and Gersten et al. (2005) to the single-subject and experimental/quasi-experimental research studies, respectively.

Single-Subject Research Studies. We examined six single-subject research studies for correspondence with the expectations of rigorous research in seven methodological categories. No studies met the minimum requirements for rigorous research in all seven categories. Strong, Wehby, Falk, and Lane (2004) provided acceptable evidence in five of the seven categories and Freeland, Skinner, Jackson, McDaniel, and Smith (2000) provided adequate documentation in four of the seven categories. The remaining four studies provided satisfactory evidence in three or fewer methodological categories. Applying the criteria proposed by Horner et al. (2005), these studies do not provide ample evidence that the research was conducted following rigorous methodological standards. To be considered an evidence-based practice, at least five high-quality research studies are needed that document the effects of the intervention. Because no studies reviewed in the investigation qualified as high-quality single-subject research, repeated reading does not meet the requirements for being an evidence-based practice for students with or at risk for LD.

Experimental and Quasi-Experimental Research Studies. We evaluated five experimental or quasi-experimental research studies across four methodological categories to determine if the articles documented high-quality reports of research. Four of the five studies provided acceptable documentation in only one or fewer categories. However, Mathes and Fuchs (1993) provided ample evidence of rigorous research across the four methodological categories. Follow-up analysis documenting the presence of desirable indicators of quality research was necessary to determine if this article reports acceptable-quality or high-quality research as described by Gersten et al. (2005). Mathes and Fuchs provided adequate evidence for four desirable indicators by documenting the concurrent related evidence for validity of

the reading achievement measures, monitoring fidelity of implementation beyond "surface" features, documenting instructional practices for the comparison groups, and presenting the results in a clear and coherent manner. As such, the research reported in this article would be classified as high quality.

To be classified as an evidence-based practice, four acceptable-quality or two high-quality experimental or quasi-experimental research studies are needed as well as a weighted effect size that is significantly greater than zero. Because the article written by Mathes and Fuchs (1993) was classified as high quality, one additional high-quality study with a weighted effect size significantly greater than zero would be needed to classify repeated reading as an evidence-based practice. Without this documentation, our results indicate that repeated reading does not qualify as an evidence-based or promising practice for students with or at risk for learning disabilities according to the requirements proposed by Gersten et al. (2005). It is possible to interpret the criteria for an evidence-based practice as stating that the effect size across high-quality and acceptable studies had to be significantly different from zero. Our interpretation, however, was that the effect size in each experimental/quasi-experimental study had to be significantly greater than zero. We chose this more conservative interpretation as we could envision the possibility that in some high-quality studies there could be a finding of no difference between treatment and comparison but, if aggregated with the effect sizes from other studies with significantly large effects, the result could mislead the field.

Implications for Research

Most special education researchers would likely have identified repeated reading as an evidence-based practice for use in building reading fluency for students with learning disabilities. However, the results of our review—based on the standards for rigorous research established by Horner et al. (2005) and Gersten et al. (2005)—suggest otherwise. We believe these results have implications for the design and implementation of specific methodologies as well as for research and research funding in general.

Implications for Single-Subject Research. Reflecting on the findings from those studies employing single-subject research designs, we found several consistent problems. Many of the single-subject studies either did not describe or only superficially discussed the process for selecting participants. Additionally, the description of the setting often lacked sufficient detail for replicability. The single-subject studies also often omitted descriptions of measurement validity, and in some cases reliability. Moreover, many studies provided little detail about how fidelity of implementation was monitored, if at all. In lieu of measuring fidelity of implementation, we expected that the interventions would have been sufficiently detailed to allow replicability. However, this was often not the case. Frequently, the studies failed in their designs to address common threats to validity either by failing to describe the baseline condition thoroughly, not achieving stability in the baseline before implementing the treatment, or not having a control or comparison condition (e.g., some studies examined growth from pretest to posttest but did not include a second group that did not receive the experimental treatment). Finally, a common difficulty with the studies was the lack of replication across participants, settings, or materials. In fact, the set of studies included for criteria did not include any studies that replicated research on a specific treatment. In all cases, the treatments were so sufficiently different from treatments in all other studies that it was impossible to generalize the findings to a particular approach to repeated reading.

Our results indicate that repeated reading does not qualify as an evidence-based or promising practice for students with or at risk for learning disabilities according to the requirements proposed by Gersten et al.

These common problems suggest that the single-subject research on repeated reading needs to be more methodologically rigorous. Horner et al. (2005) required that in order for a study to be considered of acceptable rigor, it needed to meet all the established criteria. Although this standard is set high, it was not the case that many studies

failed to meet the standard by only one or two criteria; several studies failed to demonstrate high-quality research techniques in multiple methodological categories. As such, the research is not clearly presented to allow the authors of these studies to rule out alternative hypotheses for their findings.

Implications for Experimental and Quasi-Experimental Research. Several problems were persistent across the experimental or quasi-experimental research studies examined, including providing very limited information about the interventionists or teachers implementing the treatment. This is particularly problematic as it makes it difficult to attribute the outcomes of the intervention to the intervention itself rather than possibly to the teachers' specific qualifications or characteristics. In addition, studies frequently lacked detail regarding the diagnosis of students' disabilities. This is mostly a problem in terms of generalizing the effects of the interventions to populations of students with learning or reading disabilities. Moreover, several studies lacked sufficient detail about the fidelity of implementation both in terms of how it was measured, if at all, and how it was ensured. Many early studies failed to include any information on fidelity of implementation; more contemporary studies have included more detailed information in this area as it has become standard to expect fidelity data as well as information regarding how fidelity was maintained. In a similar fashion, many of the older studies did not report effect sizes. Until the American Psychological Association (2001) made it an expectation to report the magnitude of effect for treatments, they were not commonly included in studies. Finally, it was common for experimental and quasi-experimental studies to use a single measure of performance and in many cases they did not include a generalized measure of performance. Under these circumstances it is difficult to determine the efficacy of the repeated reading interventions on general measures of reading fluency.

In recent years, expectations for rigor have increased in experimental and quasi-experimental research designs. One could argue that these heightened expectations are both a result of the evolution of educational research in general as well as increased rigor related to funded research competitions. The fact that only one experimental or quasi-experimental study of repeated reading conducted with students with or at risk for learning disabilities met the criteria for high-quality research is disconcerting. On its face, this finding warrants attention and should encourage researchers in this area to more carefully consider the criteria for high-quality research. Additionally, these findings should alert funding agencies to consider supporting studies that focus on replication of research aimed at answering fundamental research questions about interventions that are often assumed effective. Unfortunately, there is little evidence in the studies we reviewed of efforts to replicate findings, a basic tenet of scientific research.

Beyond encouraging researchers to strive to meet the criteria, we believe it is important to consider different interpretations for the rather dismal results of our reviews of both sets of studies. One possible reason for the poor outcome of both our reviews may be that we applied the criteria inappropriately. As noted in the Method section, we created a process for our review that allowed reviewers to provide a score for each study across all indicators that reflected more than just a dichotomous "present" or "absent" score. However, we believe this approach actually relaxed the criteria and more likely overestimated the quality of the research rather than underestimated it. In addition, we implemented an approach that we felt resulted in relative consensus among our reviewers while allowing for varying views. Our approach was designed to reflect a process that would more closely represent the way these criteria will be applied generally.

A second possible reason for our results may be that the criteria established by Horner et al. (2005) and Gersten et al. (2005) are simply too rigorous. This explanation seems untenable to us. Although the criteria established by these two groups of researchers are indeed rigorous, the criteria represent fundamental features of research studies that we convey to doctoral students in introductory courses every year. Moreover, these criteria serve as clear targets that researchers should consider both in designing their research as well as in describing the results of their research for dissemination. In fact, we would suggest that several additional criteria should be considered. For example, a "high-quality" research study should

include a strong theoretical or conceptual framework for the intervention under review. Although these frameworks are often assumed, authors should be encouraged to ensure that the research questions they propose are logically linked to existing theories so that their answers contribute meaningfully to the literature. We also feel that several of the desirable indicators proposed by Gersten et al. should be included as essential quality indicators. For example, we would prefer to see the addition of construct and criterion-related evidence for validity and quality of implementation as essential quality indicators. In brief, we do not believe that the criteria were too rigorous; we felt the rigorous nature of the criteria should inspire research designs that more definitively address the focal research question(s) and encourage researchers to strive to accurately and precisely report their research findings.

These criteria serve as clear targets that researchers should consider both in designing their research as well as in describing the results of their research for dissemination.

Although the criteria as established by Horner et al. (2005) and Gersten et al. (2005) resulted in disappointing findings, our findings illustrate how difficult it is to conduct high-quality research. Rather than reducing the criteria for quality, our research community should carefully consider how to reinforce efforts to improve the quality of research. These considerations may include such steps as encouraging replication studies in dissertations, providing replication competitions through funding agencies, and encouraging research collaborations that would result in replicating studies in multiple regions of the country. To achieve the goal of increasing the quality of our research, we may need to consider criteria for tenure and promotion that are based on quality of research as much as quantity of research published. Perhaps, if application of these criteria across numerous intervention areas results in similarly poor outcomes, we should rethink the way we support the implementation of efficacy studies in the future.

Implications for Practice

We are reluctant to draw too many conclusions about the implementation of repeated reading practices with students with LD based on the findings of our review. Clearly, we are unable to describe repeated reading as an "evidence-based" practice according to the criteria set forth by Horner et al. (2005) and Gersten et al. (2005). Despite this finding, we are loath to suggest that teachers stop implementing repeated reading practices for two reasons. First, as described in some detail in the introduction, repeated reading is a logical extension of multiple theoretical frameworks that suggest its use in supporting students who need fluency development. Second, though few of the studies we reviewed met the criteria for high-quality research, the effect sizes published in meta-analyses suggest that repeated reading is likely to positively affect fluency outcomes for students who are building fluency. The combination of theoretical support and positive effect sizes suggest to us that in the absence of innovations with documented effectiveness in improving fluency, repeated reading practices should be continued.

Challenges in Reviewing the Research

As a team of researchers, we found the process of this review to be both humbling and daunting. Humbling because we found ourselves surprised by our own assumptions of what would be considered high-quality or acceptable quality research. We recognize that in most cases, our own research would not have met the criteria for "high quality." Daunting as we considered the prospect of evaluating critically the research published by our respected colleagues. We want to acknowledge, therefore, that we are acutely aware of how complex it is to conceive, conduct, and report research that will meet the criteria described here and contribute meaningfully to the research literature.

Our greatest challenge in this review was creating a process that took advantage of the range of knowledge of our reviewers about interventions and intervention research while accurately reflecting the quality of each study as it related to the components and indicators of quality. For us,

overcoming this challenge required us to pursue the rubric-based review we offered in this report. This was clearly an extension of the initial criteria for which we are responsible. We felt our rubric approach allowed us to inject variance in the review of studies that resulted in a more accurate reflection of the quality of each study. The rubric was difficult to apply consistently across reviewers as is illustrated in our interrater reliabilities. We believe this inconsistency was due to the subjectivity of language used in the rubric. For example, reviewers often did not agree about whether there was "little" information provided about participants or "some" information provided. However, we were encouraged when we relaxed the reliability criteria to include ratings for each component that were within one point of each other. In this case, the reliability increased significantly. Without the range of ratings, we feared that more studies would have failed to meet the standards. In the case of repeated reading, we anticipated that this approach would result in our finding that the practice was promising, if not evidence-based. This was not the case. However, we feel that our process warrants further use and critique and we welcome both. In addition, we feel it is premature to suggest changes to the quality indicators and standards for evidence-based practices based on our review alone.

Horner et al. (2005) and Gersten et al. (2005) should be applauded for their efforts to propel special education research forward through the establishment of rigorous standards for our work. We hope that in our effort to fairly evaluate the quality of the research, we haven't overlooked any particular feature that would have resulted in a different outcome. Research on repeated reading is very valuable and we implore special education researchers to consider further work in this area that will result in enhancing our understanding of repeated reading's efficacy for building reading fluency in students with or at risk for learning disabilities.

REFERENCES

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

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Arreaga-Mayer, C., Terry, B., & Greenwood, C. (1998). Classwide peer tutoring. In K. Topping & S. Ehly (Eds.), *Peer-assisted learning* (pp. 105–120). Mahwah, NJ: Erlbaum.

Begeny, J. C., Daly E. J., III, & Valleley, R. J. (2006). Improving oral reading fluency through response opportunities: A comparison of phrase drill error correction with repeated readings. *Journal of Behavior Education, 15,* 229–235.

Chafouleas, S. M., Martens, B. K., Dobson, R. L., Weinstein, K. S., & Gardner, K. B. (2004). Fluent reading as the improvement of stimulus control: Additive effects of performance-based interventions to repeated reading on students' reading and error rates. *Journal of Behavioral Education, 13,* 67–81.

Chard, D. J., Vaughn, S., & Tyler, B. J. (2002). A synthesis of research on effective interventions for building fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35,* 386–406.

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.

Education of All Handicapped Children Act, 20 U.S.C. § 1400 et seq. (1975).

Ehri, L. C. (1995). Stages of development in learning to read words by sight. *Journal of Research in Reading, 18,* 116–125.

Freeland, J. T., Skinner, C. H., Jackson, B., McDaniel, C. E., & Smith, S. (2000). Measuring and increasing silent reading comprehension rates: Empirically validating a repeated reading intervention. *Psychology in the Schools, 37,* 415–429.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34,* 174–206.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti. M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71,* 149–164.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165–179.

Kuhn, M. R., & Stahl, S. (2003). Fluency: A review of developmental and remedial strategies. *Journal of Educational Psychology, 95*, 3–21.

LaBerge, D., & Samuels, S. J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

Lipka, O., Lesaux, N. K., & Siegel, L. S. (2006). Retrospective analyses of the reading development of grade 4 students with reading disabilities: Risk status and profiles over 5 years. *Journal of Learning Disabilities, 39*, 364–378.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading and Writing Quarterly, 13*, 123–146.

Mathes, P. G., & Fuchs, L. S. (1993). Peer-mediated reading instruction in special education resource rooms. *Learning Disabilities Research & Practice, 8*, 233–243.

Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 33*, 62–94.

McMaster, K. N., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*, 445–463.

Meyer, M. S., & Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49*, 283–306.

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

O'Shea, L. J., Sindelar, P. T., & O'Shea, D. J. (1987). The effects of repeated readings and attentional cues on the reading fluency and comprehension of learning disabled readers. *Learning Disabilities Research, 2*, 103–109.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

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

Simmons, D. C., Coyne, M. D., Kwok, O., McDonagh, S., Harn, B., & Kame'enui, E. J. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41*, 158–173.

Sindelar, P. T., Monda, L. E., & O'Shea, L. J. (1990). Effects of repeated readings on instructional- and mastery-level readers. *Journal of Educational Research, 83*, 220–226.

Snowling, M. J. (2000). From language and reading to dyslexia. *Dyslexia, 7*, 37–46.

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

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based tests of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24–53.

Strong, A. C., Wehby, J. H., Falk, K. B., & Lane, K. L. (2004). The impact of a structured reading curriculum and repeated reading on the performance of junior high students with emotional and behavioral disorders. *School Psychology Review, 33*, 561–581.

Swanson, H. L., Hoskyn, M. W., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes.* New York: Guilford.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25*, 252–261.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K., Conway, T., et al. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*, 33–58.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*, 601–638.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology, 36*, 367–397.

Weinstein, G., & Cooke, N. L. (1992). The effects of two repeated reading interventions on generalization of fluency. *Learning Disability Quarterly, 15*, 21–28.

Young, A. R., Bowers, P. G., & MacKinnon, G. E. (1996). Effects of prosodic modeling and repeated reading on poor readers' fluency and comprehension. *Applied Psycholinguistics, 17*, 59–84.

ABOUT THE AUTHORS

DAVID J. CHARD (CEC TX Federation), Leon Simmons Endowed Dean, Annette Caldwell Simmons School of Education and Human Development, Southern Methodist University, Dallas, Texas. **LEANNE R. KETTERLIN-GELLER** (CEC OR Federation), Assistant Professor, Educational Leadership, College of Education, University of Oregon, Eugene. **SCOTT K. BAKER** (CEC OR Federation), Director, Pacific Institutes for Research, and Associate Director of the Center on Teaching and Learning at the University of Oregon, Eugene. **CHRISTIAN DOABLER** (CEC OR Federation), Doctoral Candidate; and **CHANISA APICHATABUTRA**, Doctoral Candidate, Department of Special Education and Clinical Sciences, College of Education, University of Oregon, Eugene.

Address correspondence to David Chard, Annette Caldwell Simmons School of Education and Human Development, Southern Methodist University, P.O. Box 750382, Dallas, TX 75275-0382 (e-mail: dchard@smu.edu).

Manuscript submitted September 2007; accepted April 2008.