

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/232534076>

Literacy Learning of At-Risk First-Grade Students in the Reading Recovery Early Intervention.

Article in *Journal of Educational Psychology* · May 2005

DOI: 10.1037/0022-0663.97.2.257

CITATIONS

101

READS

4,061

1 author:



[Robert M. Schwartz](#)

Oakland University

36 PUBLICATIONS 636 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Literacy Intervention: Learning to Improve [View project](#)

Literacy Learning of At-Risk First-Grade Students in the Reading Recovery Early Intervention

Robert M. Schwartz
Oakland University

This study investigated the effectiveness and efficiency of the Reading Recovery early intervention. At-risk 1st-grade students were randomly assigned to receive the intervention during the 1st or 2nd half of the school year. High-average and low-average students from the same classrooms provided additional comparisons. Thirty-seven teachers from across the United States used a Web-based system to register participants ($n = 148$), received random assignment of the at-risk students from this system, and submitted complete data sets. Performance levels were measured at 3 points across the year on M. M. Clay's (1993a) observation survey tasks, 2 standardized reading measures, and 2 phonemic awareness measures. The intervention group showed significantly higher performance compared with the random control group and no differences compared with average groups. Further analyses explored the efficiency of Reading Recovery to identify children for early intervention service and subsequent long-term literacy support.

Keywords: early intervention, at-risk readers, clinical trials, literacy research, achievement gap

Early intervention is based on the premise that low-performing students can be identified and provided supplemental support after a relatively short exposure to classroom literacy instruction. This approach differs from remedial programs that often require a 2-year discrepancy between the child's reading level and either his or her grade level or reading potential (Stanovich, 1991). Early intervention has potential costs and benefits. A promising benefit is that the instruction helps many children develop a processing system for reading and writing so they can continue to learn within the ongoing classroom program. Another potential benefit is that an intervention program can serve as a prereferral service, reducing the number of students who might otherwise need long-term literacy support. A possible disadvantage is that valuable and costly resources may be devoted to intervention programs for children who might have made adequate progress in the classroom context without the intervention.

The current study investigated four interrelated questions central to judging the effectiveness and efficiency of an early intervention program. First, does the intervention increase the literacy achievement of at-risk students compared with similar students participating in classroom-based instruction? Second, does the intervention help at-risk students to close the achievement gap with their average peers in first-grade classrooms? Third, what percentage of students identified for interventions at the start of the school year

make adequate literacy gains without an intervention program? Finally, what percentage of students need long-term literacy support after receiving an intervention program? The first two questions raise issues of effectiveness; the latter two relate to efficiency. The study examined these aspects of effectiveness and efficiency for at-risk first-grade students who had been identified to participate in the Reading Recovery (RR) early intervention program.

The often-observed relation between end-of-first-grade reading performance and subsequent achievement supports one argument in favor of early intervention. Juel (1988) provided longitudinal evidence on the reading and writing development of 54 children from first through fourth grades, 24 of whom were identified as poor readers at the end of first grade. Juel reported that the

probability that a child would remain a poor reader at the end of fourth grade, if the child was a poor reader at the end of first grade was .88; the probability that a child would become a poor reader in fourth grade if he or she had at least average reading skills in first grade was .12. (p. 440)

The stability of these achievement patterns over time is part of the reason that low-performing students are considered at risk or high risk for academic difficulty.

Early intervention programs attempt to close the gap between at-risk students and their average peers during initial literacy learning, before the gap widens. Demonstrations of the extent to which this goal can be accomplished have both theoretical and practical importance. Instructional interventions provide the means to differentiate between students whose reading difficulties derive from a lack of literacy-related experience or appropriate instruction and students with specific cognitive deficits related to the reading process (Clay, 1987; Stanovich, 1988; Vellutino & Scanlon, 2002; Vellutino et al., 1996). Large-scale intervention programs that can close this achievement gap for at-risk students would increase educational opportunities for many of these stu-

Editor's Note. This research was partially supported by a grant from the Reading Recovery Council of North America.—KRH

This research was conducted with grant support from the Reading Recovery Council of North America.

Correspondence concerning this article should be addressed to Robert M. Schwartz, Department of Reading and Language Arts, Oakland University, Rochester, MI 48309-4494. E-mail: rschwartz@oakland.edu

dents and reduce the number of students who need long-term literacy support, allowing the design of more effective services for this latter group.

RR is a widely disseminated, replicable, early intervention for the lowest performing first-grade students. It utilizes a uniform lesson framework and extensive professional development to help teachers make individual instructional decisions designed to accelerate the literacy learning of these children within one-to-one, 30-min daily lessons (Clay, 1993b, 2001; Clay & Cazden, 1990; Schwartz, 1997, 2005; Stahl, Stahl, & McKenna, 1999). The most recent program data indicated that RR is currently available in 10,584 schools across the United States (Gómez-Bellengé, 2002), or approximately one out of every five schools that have a first-grade program.

Despite the wide dissemination of this program and the evaluation data available for every student who participated in the program across the United States (Gómez-Bellengé, 2002), the research base for the program remains controversial. Elbaum, Vaughn, Hughes, and Moody (2000) presented a meta-analysis of one-to-one tutoring research. Of the 42 independent samples identified for this analysis, 16 came from research on RR. This was 38% of the entire sample and over 60% of the intervention research with first-grade students. Elbaum et al. reported that the “mean weighted effect size for the Reading Recovery interventions ($d = 0.66$) was significantly higher than that for the other matched interventions, ($d = 0.29$)” (p. 615). They concluded, however, that “the findings of this meta-analysis do not provide support for the superiority of Reading Recovery over other one-to-one reading interventions” (p. 617). They based their reservations on two methodological concerns: use of “measures that may bias results in favor of Reading Recovery students” and “selective attrition of students from some treatment groups” (p. 617). These are serious potential threats to the internal validity (Campbell & Stanley, 1966), but the claim is difficult to evaluate without a detailed analysis of the designs used in intervention research. (See *What Evidence Says About Reading Recovery*, 2002, for additional discussion of these issues.)

Three peer-reviewed publications are most relevant to issues of intervention effectiveness and efficiency: Center, Wheldall, Freeman, Outhred, and McNaught (1995); Iversen and Tunmer (1993); and Chapman, Tunmer, and Prochnow (2001). The first two studies demonstrated strong effects of the RR intervention across the intervention period and through the end of the first-grade year. The third study did not replicate these results, instead finding no intervention effect. All three studies highlighted issues related to the nature of early intervention, the needs of at-risk beginning readers, and the effectiveness of highly trained teachers in addressing these needs in the one-to-one context established by the RR intervention.

Center et al. (1995) examined the progress of three groups of first-grade students on a variety of reading-related measures across the beginning, middle, and end of first grade and again in the middle of second grade. This was a random assignment, time-series design. The lowest achieving students across 10 schools were randomly assigned to the RR intervention ($n = 31$) at the beginning of first grade or to a control group ($n = 39$). A comparison group ($n = 39$) of low-achieving students from five similar schools without the RR intervention was also assessed at each test period.

The treatments resulted in significant and large effect sizes in favor of the RR group on all measures at the middle and end of first grade. The effect sizes ranged from 0.42 on a cloze measure to 3.05 on Clay's (1993a) text reading measure. A year after the intervention period, medium-term maintenance, the RR group continued to score higher than both the control and comparison groups on all measures (see Center et al., 1995, Table 7, p. 254). At this point, the effect size relative to the control group was greatly reduced. One reason for this reduction was that 15 students from the control group had been identified within their schools as needing individual support and entered into the RR program. Removing many of the lowest performing students from the control group would inflate the mean scores for the remaining group.

To examine the efficiency of the intervention procedures, Center et al. (1995) conducted a single-case analysis of students in the RR, control, and comparison groups based on test results collected in the middle of the second-grade school year. Using independent criteria from their test battery, the authors concluded that 65% of the RR group appeared to be reading at near- or above-average levels. They contrasted this with 28% of the comparison group that met these criteria and concluded that about 30% of the RR group would have reached their criteria level without the intervention. They argued that the efficiency of the intervention was low because selection procedures identify a large number of students for service who would have made adequate progress without the intervention.

Iversen and Tunmer (1993) examined the effectiveness of early intervention by comparing two versions of the RR program against a small-group intervention. The two RR groups were referred to as the *standard RR program* and the *modified RR program*. Teachers in both groups used the standard RR lesson framework. The modified RR teachers added procedures to the letter identification component such that when children could identify 35 of the upper- and lowercase alphabet characters, the teachers began to use some of the time in this section of the lesson to manipulate letters in familiar words to make new words. This component might last 2 to 4 min in a 30-min lesson.

Students were assessed at three time periods: pretreatment, at discontinuation of the RR program, and at the end of the first-grade school year. Measures included the six components of Clay's (1985) diagnostic survey, the Dolch Word Recognition Test, and measures of phoneme segmentation, phoneme deletion, and phonological recoding. The three groups did not differ significantly on any of the measures on the pretreatment assessment. The standard and modified RR groups scored higher than children in the small-group intervention at discontinuation. These differences were both significant and large. On text reading level (Clay, 1985), the effect size was over eight standard deviations. Comparisons of the two RR groups with average students from their classrooms showed similar profiles, with the only significant differences in favor of the RR groups.

The only advantage for the modified RR group compared with the standard RR group resulted from an analysis of the number of lessons to successfully meet the criteria for program discontinuation. One unusual aspect of this study was that all of the RR students, in both groups, were discontinued. The high percentage of students discontinued and the reduction in the number of lessons to achieve this goal are measures of intervention efficiency. The RR national data (Gómez-Bellengé, 2002) indicated that even

when only students who have an opportunity for a full program are considered, the national discontinuation rate was 79%, with the other 21% recommended for additional support following the intervention. The national data are based on a lesson framework that incorporates procedures similar to those described as the modified RR program (see Clay, 1993b, Section 4.10, pp. 43–47). The Iversen and Tunmer (1993) treatment and the changes to the RR framework developed independently in response to basic literacy research and increased attention to phonological processes in beginning reading.

Chapman et al. (2001) used a longitudinal cohort analysis to examine the effectiveness of the RR intervention. They reported data collected at five points across the first 2.5 years of school for a cohort of 152 students from 16 New Zealand primary schools. Unlike the previous two studies, RR students showed no progress relative to a poor reader comparison group on any measures of phonological processing, word recognition, or reading comprehension and an increased gap in reading self-concept measures relative to a high-performing comparison group. The ineffectiveness of the RR intervention to close the achievement gap relative to average performance levels was supported by reading age norms that showed performance below norm-based expectations for the RR students and the poor reader comparison group.

There are several possible explanations for these results. Perhaps the intervention as implemented in this context was far less effective than the implementations resulting in the Center et al. (1995) and Iversen and Tunmer (1993) studies. It is also possible that the intervention had a significant and large effect on the performance of the RR group but that this effect was masked by the lack of a randomly assigned control group and by design issues related to the available comparison groups. The major threats to the internal validity of this study derive from the procedures used to form the comparison groups. If the poor reader comparison group was actually an average or low-average group of readers, then the equivalence of this group with the RR group after intervention would be the expected result.

There are two primary reasons to suspect this might be the case. First, Campbell and Kenny (1999) explained that in a matched control group design, regression toward the mean tends to mask the treatment effect. The control group is identified on the basis of extreme scores from the larger population and therefore will regress toward the mean of that group. This is exactly the opposite of what is expected in a simple pretest–posttest design where regression would lead to pseudotreatment effects for a compensatory program like RR (Shanahan & Barr, 1995). Second, the retrospective matching procedure excluded any low-performing students who entered the RR intervention after the start of the 2nd year. By excluding low-performing students judged to need an intervention program, the poor reader comparison group is limited to only students who made at least adequate progress in the classroom setting. These confounds result from the absence of random assignment to the intervention and poor reader comparison condition.

The three studies discussed above, Center et al. (1995), Iversen and Tunmer (1993), and Chapman et al. (2001), presented different pictures of the ability of highly trained teachers to address the needs of at-risk beginning readers in the context established by the RR intervention. Of these, only the Center et al. study provided the randomly assigned comparison group called for in previous re-

views of the intervention literature (Hiebert, 1994; Shanahan & Barr, 1995). The current study incorporates several design elements to evaluate intervention effectiveness and efficiency. Most important, low-performing students were randomly assigned to either first- or second-round service in the intervention program. This random assignment allowed for comparison of progress with and without an intervention program across the first half of first grade. The intervention was provided in addition to classroom literacy instruction and other forms of literacy support available within buildings. To control for these factors, I selected the at-risk pairs from the same classroom within each building. A high-average and a low-average reader from that classroom were also assessed to gauge the progress of the 2 at-risk students. The research design provided an experimental comparison of the early literacy progress of at-risk students in comparable instructional settings with and without an intervention program.

Method

Participants

Forty-seven RR teachers from different schools in 14 states obtained consent forms from the children's parents to participate in the study, submitted the names of 2 at-risk students to a Web-based program for random assignment to first- or second-round RR service, and submitted student data at the end of the school year. Because comparison of first- and second-round RR students at the transition period was critical to evaluating the intervention effect, only data from 37 teachers who included this information were considered in the data set for analysis. Incomplete sets indicated that either the first- or second-round student moved prior to the midyear transition testing. The teachers also submitted data on a low-average and a high-average student from the same classroom as the first- and second-round RR students for a total sample of 148 first graders.

The sample was 53% male and 47% female. Lunch subsidy figures were available for only 107 students because school district policies sometimes prevented release of this information. Of this group, 43% received free school lunches, 8% received reduced-price school lunches, and 49% did not receive lunch subsidies. The racial and ethnic breakdown of the student sample was 46% White, 40% African American–Black, 12% Hispanic–Latino, and 2% Asian. Demographic information for each of the four comparison groups and the RR national data (Gómez-Bellengé, 2001; Gómez-Bellengé & Thompson, 2000) are presented in Table 1. The RR national data provide some indication of the similarity of the current sample to the larger population involved in this large-scale intervention program. The teachers participating in this study were volunteers, and no attempt was made to obtain a representative sample of teachers or students from the national implementation.

For the first- and second-round RR students, the end-of-program-status data indicated that 65% of this group was considered successfully discontinued, 16% recommended for further services, and 16% had incomplete programs, with 1 second-round student who moved prior to completing the program and 1 first-round student who withdrew prior to 20 weeks of service (but who was assessed at the transition period and whose data are reported). This compares with 56% discontinued, 15% recommended, 19% incomplete, 5% moved, and 4% classified as “none of the above” in the 1998–1999 national RR data. In the national data, when only those students who had an opportunity to receive a full program (up to 20 weeks) were considered, 79% were successfully discontinued. The 65% of students discontinued in the current sample falls, as expected, between the 56% figure that includes third-round students with a high likelihood of incomplete programs and the 79% level reported for full program students (who received 20 weeks of lessons, if not successfully discontinued earlier). In the current sample, all of the incomplete program students came

Table 1
Age, Gender, Race, and Lunch Status by Group and National Reading Recovery (RR) Population

Demographic	RR		Classroom		RR population ^a
	1st round	2nd round	Low average	High average	
Age (months)					
<i>M</i>	77.4	76.4	77.8	77.5	—
<i>SD</i>	4.3	3.8	7.7	4.4	—
Gender (%)					
Male	61	41	45	66	58 ^b
Female	39	59	55	34	42
Race (%)					
White	38	47	50	48	58
Black	47	38	31	43	24
Hispanic	12	15	19	6	14
Asian	3	0	0	3	2
Lunch status (%)					
Free	46	50	38	36	53
Reduced price	14	7	12	0	8
Regular	40	43	50	64	39

Note. Dashes indicate that age was not reported for the population. ^a*n* = 142,291; from Gómez-Bellengé and Thompson (2000). ^bFrom Gómez-Bellengé (2001).

from the second-round group and all but 1 of the recommended students came from the first-round group. After the transition testing, 6 of the students initially identified for the low-average classroom group and 3 from the high-average group entered RR. Five of these 9 students successfully completed the program prior to the end-of-year assessment. Data from these students are included with their initial classification group.

Measures

Students were assessed at the beginning of the year, at the transition from first- to second-round service for the RR students, and at the end of the school year on the six measures from Clay's (1993a) *An Observation Survey of Early Literacy Achievement*. In addition, at the transition period and the end of the year, students were assessed on the Yopp-Singer Phoneme Segmentation Task (Phoneme Segmentation; Yopp, 1988), a sound deletion task, the Slosson Oral Reading Test—Revised (Nicholson, 1990), and the Degrees of Reading Power Test (Forms JO and KO; Touchstone Applied Science Associates, 2000). Teachers submitted a data summary for each child at each test period. They did not submit item information on each task, so reliability estimates for the research sample could not be calculated.

In the latest edition of the Observation Survey, Clay (2002) provided updated norms for these tasks as well as a summary of reliability, validity, and discrimination indices established in research on these tasks. The tasks included in the Observation Survey are designed to assess a variety of reading and writing knowledge related to literacy learning. Alternate forms were available for three of the tasks, with specified forms used at each test period. The set of tasks provides indications of strengths and needs that can guide instruction. Clay (2002) reported the intercorrelations between tasks for age groups ranging from 5.0 to 7.0 in half-year increments. For the total sample of 796 children, the correlations ranged from .554 to .894.

The text level task, as conducted in the United States, used a standard set of books that were leveled by difficulty and specific text characteristics (Peterson, 1991). The gradient of difficulty reflected in these texts was similar to instructional materials used in the RR program and many early literacy classroom programs (Pinnell & Fountas, 1999). The Ohio stanines for text level indicated an average of Level 2 for the fall of first grade, with

a range of Level 9 to Level 12 for average performance in the spring of first grade. These results were for an urban norm group (Clay, 2002). The National Data Evaluation Center random sample data indicated an average end of first-grade text level of 20 for a stratified national sample (Gómez-Bellengé & Thompson, 2000, 2004). Clay (2002) reported that the scoring of running records, on which the text level decisions were based, was reliable across two scorings by a trained recorder over a 2-year interval ($r = .98$).

The Letter Identification task (Clay, 2002) asked students to respond to 26 uppercase and 28 lowercase letter forms. The additional lowercase letters included two forms of *a* and *g*. The child could respond with a letter name, a sound, or a word beginning with that letter (maximum score = 54, Cronbach's $\alpha = .95$).

The Concepts About Print task was a research-based measure (Clay, 2002) of emergent readers' knowledge of conventions related to printed language. The task included standard procedures for administration and four specialized booklets to provide alternate forms (Clay, 2002). The adult read one of these booklets to the child. The child was asked to help by responding to questions or requests related to book handling, directional behavior, visual scanning, and specific concepts related to printed language, like punctuation, and the relationship of letters and words within sentences (maximum score = 24, Cronbach's $\alpha = .78$; split-half $r = .95$; Clay, 2002).

The Ohio Word Test (Clay, 2002) was a 20-item list of high-frequency words, available in three alternate forms. Scoring was based on the number of words read correctly (maximum score = 20, Cronbach's $\alpha = .92$). The Writing Vocabulary task (Clay, 2002) allowed 10 minutes for children to write as many words as they could on a blank sheet of paper. A standard set of prompts was used to encourage additional attempts if needed. Scoring was a count of the number of words correctly generated (test-retest $r = .62$ and $.97$).

The Hearing and Recording Sounds in Words (HRSW) task (Clay, 2002) was another type of writing assessment. The teacher read one of five short passages (alternate forms) aloud and asked the child to write each word as the passage was read again word by word. When a child did not know a word, the child was prompted to say the word slowly and think about what he or she heard and how to record it in print. The task was scored on the number of phonemes correctly recorded (maximum score = 37, Cronbach's $\alpha = .96$).

The Phoneme Segmentation Test (Yopp, 1988) required the separate articulation of the phonemes in a word. The task consisted of 4 practice items and 22 test items. Each item was scored as 1 point if all phonemes were separated and articulated (maximum score = 22, Cronbach's $\alpha = .95$).

The Deletion Task was a 10-item version of the Rosner (1975) task (as cited in Yopp, 1988). This task requires the child to repeat a word and then say it again but omit a given syllable or sound, for example, "Say *cowboy*. Now say *cowboy* but do not say /*cowl*." Two items ask for syllable deletion, with the remaining item requiring phoneme deletion from an initial, medial, or final position. The phoneme or syllable deletion resulted in a different word (maximum score = 10, for the Rosner task, Cronbach's $\alpha = .78$).

The Slosson Oral Reading Test—Revised (Nicholson, 1990) contained 200 words arranged in ascending order of difficulty with 20 words per list. The administration stopped after the child missed all the words on one list. This was a standardized, norm referenced measure (maximum score = 200, Kuder-Richardson 21 for ages 6 to 7 = .98).

The Degrees of Reading Power Test (Touchstone Applied Science Associates, 2000) provided two alternate forms of a primary reading comprehension measure. The JO form was used at the transition period, and the KO form was used at the end of the year. The task required children to read a passage with a word or set of words missing. A line with an item number indicated each missing word. Students selected the appropriate word to complete the sentence from a set of four or five alternatives listed

by the item number. All of the alternatives were semantically and syntactically consistent with the sentence in which the deleted word occurred, so students needed to consider information from the passage to make their selection. This was a standardized, norm referenced measure. The primary forms were recommended for the end of first grade and the beginning of second grade (maximum score = 28, Kuder-Richardson 20 at second-grade level = .92).

Procedures

A Web site was established to describe the purpose, design, and procedures required for RR teachers to participate in the study. Eight hundred teacher leaders from around the country were presented with a short orientation to the study at their June professional development institute and asked to seek district consent for participation the following fall. Teacher leaders are responsible for training RR teachers and supervising the program implementation for a district or consortium of districts that usually includes 20 to 50 RR teachers. This teacher leader group was contacted by mail in August and provided with district- and building-level consent forms as well as information on the study Web site. The site provided downloadable consent forms for districts, buildings, and parents of student participants. It also included timelines for testing, data submission, procedures for teachers to register for the study, and a process for submitting names of 2 at-risk students for programmed random assignment to first- or second-round RR service.

Following the normal selection procedure for the RR program (Askew, Fountas, Lyons, Pinnell, & Schmitt, 1998), the first-grade classroom teachers identified the lowest 20% to 30% of their students for assessment on six tasks from Clay's (1993a) Observation Survey. The tasks included Letter Identification, the Ohio Word Test, Concepts About Print, Writing Vocabulary, HRSW, and Text Reading Level.

Each RR teacher devoted one of his or her four 30-min teaching slots to this study. In line with program standards, the three lowest performing students from the group assessed were assigned to RR service in the teacher's other three teaching slots. Procedures for identifying the lowest performing students on the six assessments from Clay's (1993a) Observation Survey varied across sites. Typically, scores on each measure are converted to stanines, and the stanines are summed across tasks. Students with the lowest totals are considered for initial service. When students are equally low on these criteria, the pattern of results, observed behavior, and judgments by the first-grade teacher or kindergarten teacher may be used to select the lowest students.

On measurement criteria, the variation in student scores would suggest that selection of the lowest students was a somewhat arbitrary decision among a small group of low-performing students. A previous pilot study, however, indicated that many RR teachers considered the need to follow their selection process for serving the lowest students first an ethical issue. They would not participate in a study that did not attempt to ensure that principle was upheld. After the first three slots were assigned, 1 of the next 2 students would have to be assigned to second-round service. Participating teachers were willing to accept that identification of the next student for service was often arbitrary and that a random procedure would be reasonable, given that both students would receive service during the year.

The RR teachers identified the next child eligible for service and the next lowest child from the same classroom (even if there was a child in another class that might appear lower). These 2 students were randomly assigned to receive RR service either during the first half or second half (first or second round) of the school year in the remaining teaching slot. Two additional students from the same classroom were identified to participate in the assessments given at the beginning of the year, at the transition between service for the first- and second-round RR students, and the end of the year. These students were selected on the basis of the classroom teacher's ranking and available assessment information as a high-average and low-average reader. The high-average child was from the middle of the

teacher's rankings after the students expected to receive RR service were removed. The low-average child was the lowest student in the class who was not expected to receive RR service.

The transition test took place when the first-round RR student was judged to have met the criteria to terminate the intervention program or at the end of 20 weeks in the program if the child was judged not to be making accelerated progress that would lead to termination of the intervention. The criteria for program termination required that the child both reach the average level of literacy performance for his or her class and demonstrate a set of strategies sufficient to ensure continued progress given good classroom instruction. Strategy decisions were based on self-correction rates (Clay, 2002) and an analysis of error substitutions (Clay, 1993b, 2002; Schwartz, 1997). This usually required between 12 and 20 weeks of RR lessons. All 4 students, from each of the participating classrooms, were retested at the end of the first-grade school year, usually 2 weeks prior to the end of the school year. The RR teachers administered most of the assessments with the probable exception of the Observation Survey measures used to make program discontinuation decision. RR guidelines require that another trained teacher administer these measures (Askew et al., 1998).

Results

This section includes two sets of analyses designed to examine the effectiveness and efficiency of the RR intervention. First, analysis of variance (ANOVA), simple effects analyses, and comparisons were used to evaluate effectiveness based on changes in the relative performance of each group across the study on each measure. Second, intervention efficiency was evaluated using norms for midyear text reading level to assess the relative progress of each matched pair of at-risk students. The different patterns of progress were used to assess intervention efficiency in terms of reducing the number of at-risk students who needed long-term support and the selection of appropriate students for service.

Effectiveness Evaluation

Repeated measures analysis. For each of the Observation Survey measures, a 4 (group) \times 3 (test period) repeated measures ANOVA was conducted to examine intervention effectiveness. The remaining measures were analyzed using a 4 (group) \times 2 (test period) repeated measures ANOVA. Tables 2, 3, and 4 report the means and standard deviations for the pretreatment, transition, and end-of-year test periods, respectively, for the four groups—first-round RR, second-round RR, low-average classroom, and high-average classroom. A significant Group \times Test Period interaction for the Observation Survey variables was followed by a simple effects analysis among groups at each test period. For the four variables that were not measured prior to the intervention period, a significant main effect of group, or Group \times Test Period interaction, was followed by either main effect comparison or simple comparison, respectively. Because the random assignment to first- or second-round RR service was the most critical comparison, and there were 10 of these comparisons at the transition period, one for each dependent variable, a conservative alpha level for these comparisons was set at .005. Exact probabilities were reported for marginal *p* values between .05 and .005 (Keppel, 1982). Effect sizes (Cohen, 1988) were calculated only for significant simple comparisons between the first- and second-round RR students at the transition period because these were the only comparisons that reflect treatment effects for randomly assigned groups. These were

Table 2
Means and Standard Deviations for the Four Groups on Pretreatment Measures

Measure	Maximum score	Reading Recovery				Classroom			
		1st round ^a		2nd round ^b		Low average ^c		High average ^d	
		M	SD	M	SD	M	SD	M	SD
Text Level	30	0.61	0.80	0.64	0.90	2.13	3.10	4.55	5.97
Letter Identification	54	44.75	5.98	44.25	9.66	48.38	6.29	51.72	2.45
Ohio Word Test	20	0.81	1.74	0.83	1.60	3.47	4.44	6.69	5.56
Concepts About Print	24	10.92	2.61	10.81	3.27	13.03	3.02	15.90	3.35
Writing Vocabulary		5.44	2.85	6.08	4.14	12.44	8.58	20.10	12.23
HRSW	37	9.14	6.25	9.75	7.43	17.75	9.94	24.72	8.21

Note. HRSW = Hearing and Recording Sounds in Words.
^an = 36. ^bn = 36. ^cn = 29. ^dn = 32.

calculated as the mean difference between groups divided by the pooled standard deviation.

The analysis for each of the Observation Survey measures resulted in a significant Group × Test Period interaction: Text Level, $F(6, 216) = 6.52, p < .005$; Letter Identification, $F(6, 228) = 7.49, p < .005$; Ohio Word Test, $F(6, 228) = 11.37, p < .005$; Concepts About Print, $F(6, 224) = 11.54, p < .005$; Writing Vocabulary, $F(6, 228) = 2.91, p = .01$; HRSW, $F(6, 226) = 25.73, p < .005$. The analysis of two of the additional measures assessed at the transition and end-of-year test periods resulted in significant Group × Test Period interactions: Degrees of Reading Power, $F(3, 116) = 4.37, p = .006$, and Phoneme Segmentation, $F(3, 116) = 3.14, p = .03$. For the Slosson Oral Reading—Revised task, there was no interaction effect, but the main effect of group was significant, $F(3, 116) = 5.87, p < .005$. On the phonemic deletion measure, there was an increase from the transition period to the end-of-year testing, $F(1, 116) = 9.93, p < .005$, but there was no group effect or interaction.

Simple effects and comparisons at each test period. The simple effects for each of the Observation Survey measures were significant at the pretreatment period: Text Level, $F(3, 129) = 10.28, p < .005$; Letter Identification, $F(3, 129) = 8.45, p < .005$;

Ohio Word Test, $F(3, 129) = 18.91, p < .005$; Concepts About Print, $F(3, 129) = 19.08, p < .005$; Writing Vocabulary, $F(3, 129) = 25.82, p < .005$; and HRSW, $F(3, 129) = 27.14, p < .005$. Table 2 shows the means and standard deviations for each group on the pretreatment measures administered in the fall of first grade. As expected with random assignment, none of the simple comparisons between the first-round RR group and the second-round RR group approached significance ($p > .50$). The RR groups scored lower than the high-average classroom group on all measures ($p < .005$). The first- and second-round RR groups scored lower than the low-average group on all measures, with significant differences ($p < .005$) on the Ohio Word Test, Writing Vocabulary, and HRSW, with marginal or mixed differences on Letter Identification ($p = .03$ and $.01$, respectively, by group) and Concepts About Print ($p = .005$ and $.003$), and no significant difference on Text Level ($p > .05$).

The most critical test of the intervention effectiveness came at the transition test period where the randomly assigned first-round students have completed the intervention and the second-round RR students are now entering the intervention. The simple effects for each of the Observation Survey measures were significant at the transition period: Text Level, $F(3, 129) = 22.77, p < .005$; Letter

Table 3
Means and Standard Deviations for the Four Groups on Transition Measures

Measure	Maximum score	Reading Recovery				Classroom			
		1st round ^a		2nd round ^b		Low average ^c		High average ^d	
		M	SD	M	SD	M	SD	M	SD
Text Level	30	12.35	4.80	4.70	2.40	8.04	5.67	14.55	7.67
Letter Identification	54	52.68	1.27	51.68	2.78	53.23	0.90	53.52	0.87
Ohio Word Test	20	14.94	3.99	8.87	4.75	12.67	4.48	15.79	4.73
Concepts About Print	24	19.35	2.55	16.68	2.30	17.67	2.25	18.79	2.55
Writing Vocabulary		42.03	11.42	31.00	12.94	36.03	14.53	44.10	15.14
HRSW	37	34.97	2.70	29.08	7.37	33.32	4.11	33.86	3.32
Slosson Oral Reading Test-Revised	200	30.58	14.41	18.12	11.87	25.19	16.01	36.73	20.02
Degrees of Reading Power	28	4.82	3.88	4.27	3.88	5.29	4.66	5.37	4.97
Phoneme Segmentation	22	17.70	4.93	15.27	5.43	16.58	6.11	16.87	4.92
Deletion	10	6.64	2.56	5.58	2.50	5.84	3.14	7.23	2.19

Note. HRSW = Hearing and Recording Sounds in Words.
^an = 37. ^bn = 37. ^cn = 29. ^dn = 30.

Table 4
Means and Standard Deviations for the Four Groups on End-of-Year Measures

Measure	Maximum score	Reading Recovery				Classroom			
		1st round ^a		2nd round ^b		Low average ^c		High average ^d	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Text Level	30	17.07	7.91	14.20	6.30	14.82	7.68	21.18	7.43
Letter Identification	54	53.17	1.04	53.50	0.73	53.14	0.89	53.44	0.93
Ohio Word Test	20	17.48	3.22	17.20	3.39	17.86	3.25	18.93	2.38
Concepts About Print	24	19.66	2.79	20.37	2.57	19.18	2.35	20.48	2.04
Writing Vocabulary		48.86	14.25	48.40	13.97	42.89	17.75	56.07	20.15
HRSW	37	34.72	2.99	34.90	2.68	33.89	5.32	34.82	3.08
Slosson Oral Reading	200	49.38	26.95	39.30	17.82	44.89	21.36	58.56	26.00
Degrees of Reading Power	28	8.69	4.46	6.00	3.43	7.68	5.01	11.78	6.13
Phoneme Segmentation	22	18.28	5.15	18.20	4.07	17.35	5.12	16.81	6.03
Deletion	10	6.93	2.71	7.93	6.68	7.29	2.55	8.00	2.32

Note. HRSW = Hearing and Recording Sounds in Words.

^a *n* = 29. ^b *n* = 30. ^c *n* = 27. ^d *n* = 28.

Identification, $F(3, 129) = 7.54, p < .005$; Ohio Word Test, $F(3, 129) = 16.59, p < .005$; Concepts About Print, $F(3, 129) = 8.70, p < .005$; Writing Vocabulary, $F(3, 129) = 6.67, p < .005$; and HRSW, $F(3, 129) = 10.29, p < .005$. For the Degrees of Reading Power and the Phoneme Segmentation measures, there were no significant differences among groups at the transition period, $F(3, 123) = 0.43, p > .05$, and $F(3, 123) = 1.16, p > .05$, respectively.

As shown in Table 3, the first-round RR group scored higher than the second-round RR group on all the Observation Survey measures at the transition period. These differences were statistically significant ($p < .005$) and the effect size was large ($d > 0.80$ was considered large; Cohen, 1988) for Text Level ($d = 2.02$), the Ohio Word Test ($d = 1.38$), Concepts About Print ($d = 1.10$), Writing Vocabulary ($d = 0.90$), and HRSW ($d = 1.06$). On the basis of the significant group effect in the main analysis, comparisons were warranted for the Slosson Oral Reading Test—Revised. The comparison between the two RR groups indicates a significant advantage for the first-round group ($p < .005, d = 0.94$).

As shown in Table 3, the means for the high-average classroom group at the transition period were still slightly higher than those for the first-round RR group on all measures except Concepts About Print, HRSW, and Phoneme Segmentation. Statistical comparisons between these groups indicated no significant differences. The first-round RR group scored slightly higher than the low-average group on all measures except Letter Identification and Degrees of Reading Power. The only significant difference between these groups was on the Text Level measure ($p < .005$). Marginal differences between these groups were found on the Ohio Word Test ($p = .04$) and Concepts About Print ($p = .005$).

Table 4 shows the means and standard deviations for the four groups on the end-of-year measures. At the end-of-year test period, simple effects were calculated for each significant interaction and main effect from the overall analysis. The simple effects were significant at the end-of-year period for Text Level, $F(3, 110) = 5.15, p < .005$; Writing Vocabulary, $F(3, 110) = 2.90, p = .04$; the Slosson Oral Reading Test—Revised, $F(3, 110) = 3.45, p = .02$; and Degrees of Reading Power, $F(3, 110) = 7.14, p < .005$. Simple comparisons on these measures showed no significant difference between the first-round RR group and the other three

groups. Differences between the first-round RR group and the high-average group were marginally significant on Text Level ($p = .04$) and Degrees of Reading Power ($p = .02$). By this point in the school year, 6 of the students initially identified as low average and 3 students from the high-average group had entered and received some RR service. The largest differences at this test period were between the high-average group and the second-round RR group or the low-average group (Text Level, $p < .005$, for both comparisons; Writing Vocabulary, $p < .005$, between the two average groups; Slosson Oral Reading Test—Revised, $p < .005$ and $p = .03$, respectively; Degrees of Reading Power, $p < .005$, for both comparisons).

Efficiency Evaluation

To investigate the efficiency of identification of students for intervention and long-term literacy support, I conducted a matched-case analysis following the logic presented by Center et al. (1995). This analysis identified patterns in the rate of progress displayed by pairs of students from the same classrooms. The transition test scores mark the end of the intervention program for a first-round student and the beginning of the program for the classmate identified for second-round intervention. A low rate of progress by at-risk students prior to the transition testing indicates an increasing gap between the at-risk and average classroom group and a possible need for long-term support services. High rates of progress by second-round students would indicate difficulty in early identification of the students most in need of intervention support. Transition text reading levels were used to identify these patterns.

A pattern where the first-round child made accelerated progress toward average levels of performance and the second-round child made slow progress would confirm the efficacy of intervention as part of a comprehensive program to support at-risk students. Criteria of text reading levels of 12 or above for the first-round child and 6 or below for the second-round child at the transition period would confirm this pattern. On the basis of a recent stratified national random sample of first-grade students (Gómez-Bellengé & Thompson, 2004), Text Levels 12 to 14 were the range for

Stanine Group 5, indicating average levels of performance at midyear. Text Level 6 was the high end for Stanine Group 3, corresponding to the 27th percentile for the national sample in the middle of first grade. Disconfirmation could be indicated by the first-round child achieving a text reading level less than 12 or by the second-round child scoring above Text Level 6. The first pattern would indicate a lack of accelerated progress due to the intervention; the second pattern would indicate selection of a student for intervention who might make adequate progress without the intervention. Both forms of disconfirmation could appear in a single matched set of students. The number and percentage of students fitting these patterns are shown in Table 5.

Sixty-two percent of the matched pairs confirmed the expected pattern for an effective intervention with at-risk students. Twenty-four percent of the cases showed lower than expected progress by the intervention students. In 11% of the pairs, the second-round students disconfirmed expectations by making reasonable progress without an intervention. This is a conservative estimate of second-round students who may have been incorrectly identified for the intervention and might be able to make adequate progress in the normal classroom context. Only 2 of these students achieved a text reading level of 12 or above. One additional pair disconfirmed the expected pattern on both criteria, a first-round student with text level of below 12 and a second-round student reading above Text Level 6.

Discussion

Intervention Effectiveness

The results demonstrate the effect of the RR intervention on the literacy progress of low-performing students. The at-risk students who received an intensive, one-to-one early intervention during the first half of the school year performed considerably better than similar students from the same classrooms randomly assigned to receive the intervention in the second half of the year. This is most apparent on measures taken at the transition between first- and second-round intervention service, with large effect sizes for Text Reading Level, the Ohio Word Test, Concepts About Print, Writing Vocabulary, HRSW, and the Slosson Oral Reading Test—Revised.

Comparisons with the high-average and low-average classroom groups at the transition period further confirm that the intervention goals were met. The at-risk students who received the intervention, the first-round RR group, scored between these two groups on all measures. There were no significant differences between the intervention group and the high-average group. The intervention group scored higher than the students identified for the low-

average group who were not anticipated to need intervention support. Many of these low-average students made progress in the classroom setting, although 6 of these students entered RR during the second half of the year. The scores of these students may contribute to the significantly lower performance of the low-average group compared with the first-round RR group on the Text Level measure at the transition test period.

The overall pattern shown across Tables 2 and 3 is that at-risk students who received the intervention closed the performance gap with their average peers. This is particularly clear for the Text Level measure where the two at-risk groups scored approximately four text levels below the high-average group in the fall but by the transition period the intervention group had reduced this gap to two text levels, whereas the average for the other at-risk group is now 10 text levels below that for the high-average group. This pattern is also clear for the Writing Vocabulary measure but less apparent for measures with a closed set of items that all students might be expected to learn across first grade (i.e., Letter Identification, Concepts About Print, and the Ohio Word Test). When the pretreatment scores on the Ohio Word Test are viewed relative to the more open-ended word task from the Slosson Oral Reading Test—Revised at the transition testing, it is again clear that the intervention helped the at-risk children attain average performance levels. This closing-the-gap pattern of at-risk students relative to their average peers is similar to results reported by Center et al. (1995) and Iversen and Tunmer (1993) for change across the intervention period.

Similar patterns might be expected for the comprehension and phonemic awareness measures, but there were no significant differences among groups in this study on these measures at the transition test period. Meaningful measures of comprehension are difficult to obtain at early reading level (Paris & Paris, 2003; Stallman & Pearson, 1990). Performance on the Degrees of Reading Power measure of comprehension at the transition-testing period appears to be at near-chance levels. The task is quite demanding. On the simplest items, students need to read three sentences and select one word from a set of four to complete the middle sentence. All four choices fit the sentence, so an appropriate choice must combine meanings across sentences. The task is recommended for the end of first grade, and reliabilities are reported only at the beginning in second grade.

Only indirect evidence of comprehension gains was available in this study. At early reading levels, measures of word recognition and measures of reading comprehension tend to be highly related. Many comprehension measures for beginning readers assess low-level skills involving recognition or recall at the word or sentence level (Paris & Paris, 2003). Even at higher reading levels, word

Table 5
Number and Percentage of At-Risk Matched Pairs Achieving Different Patterns on Text Reading Level (TRL) at the Transition Test Period

Pattern	Rate of progress	Number	Percentage
Confirmation	1st-round TRL \geq 12 and 2nd-round TRL \leq 6	23	62
Disconfirmation			
Pattern 1	1st-round TRL $<$ 12 and 2nd-round TRL \leq 6	9	24
Pattern 2	1st-round TRL \geq 12 and 2nd-round TRL $>$ 6	4	11
Pattern 3	1st-round TRL $<$ 12 and 2nd-round TRL $>$ 6	1	3

recognition and comprehension remain highly correlated. For example, the developers of the Slosson Oral Reading Test—Revised measure report a correlation of .83 with the reading comprehension section of the Peabody Individual Achievement Test (Nicholson, 1990). The significant gains on this standardized measure and the Text Level measure are indications that RR students have increased reading comprehension if only through increased access to readable texts.

The phonemic measures present a different picture. Here the level of performance was relatively high at both the transition and end-of-year test periods. The second-round RR students showed the largest increase of any group on the phonemic measures from the transition period to the end of the year. Differences among groups were not large and other than the second-round RR group only the low-average group showed more than a one-item mean increase from the transition testing to end-of-year testing on either measure. The limited range of items and tasks used to assess phonemic awareness made it difficult to detect patterns across groups or time.

Intervention studies that included a broader range of phonemic awareness tasks have demonstrated increased performance on these measures. Both Iversen and Tunmer (1993) and Center et al. (1995) indicated that students who were successful in learning to read through intervention programs develop phonemic awareness skills. RR groups in their studies often performed significantly higher than control groups or average comparison groups on these measures. When Center et al. subdivided their RR group on the basis of outcome measures, the successful students scored higher on phonemic measures than the unsuccessful students both on entry and exit measures of phonemic awareness.

Good beginning readers score higher than struggling beginning readers on phonemic awareness measures, and these measures taken at the end of kindergarten or the beginning of first grade can predict progress across first grade (Center et al., 1995; Chapman et al., 2001; Iversen & Tunmer, 1993). This result is consistent across the literature on phonemic awareness (Blachman, 2000). These findings suggest that the efficiency of intervention programs might be improved by focusing greater attention on phonemic awareness instruction within an intervention program or by using phonemic awareness measures to select students with a higher probability of benefiting from the intervention. The latter strategy is not a viable option for an intervention program designed to serve the lowest performing first-grade students. It would require excluding many of the lowest progress readers or assigning them to an alternative intervention until the phonemic awareness criteria were achieved. Phonemic awareness instruction is incorporated across many components of the RR lesson framework (Adams, 1990; Pinnell, 2000). The matched case analysis was designed to address questions of intervention efficiency by examining the relative percentage of at-risk students who make adequate progress with or without an intervention program.

Intervention Efficiency

One aspect of early intervention efficiency is the number of children identified for intervention services who would have made adequate progress without the intervention. The matched case analysis indicated that 86% (62% + 24%) of the second-round students made very little progress in text reading across the first

half of the school year, with only 14% achieving text reading levels greater than 6. This compared with 73% (62% + 11%) of the first-round students whose text reading appears to be reasonably under way by this point—Text Level 12 or above. This analysis is similar to that reported by Center et al. (1995). In their single case analysis of the students from the comparison schools, they indicated that 28% of these students, identified as at risk at the beginning of first grade, achieved near- or above-average literacy levels by midsecond grade without intervention support. Identifying a group of at-risk children at the beginning of first grade to receive early intervention services is likely to include some children who might have made adequate progress in the classroom context. Serving these children increases the cost of early intervention programs. Estimates for the size of this group range from 28% in Center et al. to 14% in the current study.

Several factors influence the percentage of children in this category. Fewer children are likely to be misidentified for intervention service in the second or third round of service. These children have had a half year or more exposure to the classroom literacy program. If they are still performing at low levels, the likelihood of accelerated progress without intervention is small. Kindergarten programs that focus on literacy can also improve initial selection decisions. Early intervention decisions can be made more effectively if children have had many opportunities to learn. With good classroom instruction, interventions can be limited to only those children who have not benefited from a rich set of classroom literacy experiences.

The second aspect of intervention efficiency is the potential to reduce the number of children who need long-term literacy support. Center et al.'s (1995) single-case data showed that in the comparison schools, with no RR service, 66% of the at-risk students identified at the beginning of first grade were still reading at Text Level 4 or below by the middle of second grade. In contrast, there were only 2 students (9%) in their RR group reading below Text Level 10 by the middle of second grade. Over the shorter period of time involved in the current study, 27% (24% + 3%) of the first-round students appear to need long-term support following the intervention compared with 86% of the second-round students (prior to their intervention treatment). This is a conservative estimate using criteria of below Text Level 12 for the intervention students versus Text Level 6 or less for the second-round students. If these patterns were applied to the bottom 20% of the grade cohort, then long-term support would be needed for 5% ($.27 \times .20$) versus 17% ($.86 \times .20$) of the cohort with, and without, an intervention program, respectively.

Clay (1987) and Vellutino et al. (1996) argued that a labor-intensive early intervention is a necessary screening for the identification of children who may need long-term support. As a prereferral service, the intervention provides a dynamic assessment (Brown & Campione, 1985) of a child's ability to benefit from instruction. Vellutino et al. implemented a daily one-to-one tutoring intervention similar to RR in order to classify children in a reading disability study. They found that 67% of the poor-reader group could be brought to average or above-average levels in one semester of tutoring. They indicated that after the intervention, 15% of the tutored students fell in the severely impaired range (below the 15th percentile) with another 18% still below average (less than the 30th percentile). They estimated that use of the normal exclusionary criteria for identification of reading-disabled

students would result in a 9% referral rate. Combining the exclusionary criteria with an intensive intervention program would reduce the referral rate to 1.5% of the population. The estimated reduction in students needing long-term support varies across these studies on the basis of the criteria used to measure success and need for continued support. Still, it is clear that effective early intervention can greatly reduce the number of children requiring long-term support.

Limitations

One limitation of the current design is the lack of double-blind protection. In a medical setting, double-blind procedures would ensure that neither the patient nor the doctor knows whether a particular patient receives an experimental drug or a placebo. In the current design, the RR teacher knows the treatment status of all 4 participants. They conducted most of the testing with the probable exception of transition Observation Survey testing for the first-round RR students that would have been conducted by another trained teacher according to RR guidelines (Clay, 1993b). These teachers have been trained in the administration and scoring of the Observation Survey. Still, knowledge of treatment condition can introduce sources of bias.

A further limitation is that neither the RR teachers nor the sample of students can be considered representative of the national implementation of this intervention program. RR teachers volunteered to participate in this project. This field-based random experiment is only suggestive of the treatment effects represented in the national evaluation data on the over 140,000 first-grade students served annually in this program (Gómez-Bellengé, 2002). The national data reports show the gains for RR students across the intervention period and across first grade. These gains are compared with a national random sample that establishes average performance levels for these schools. What is not available in the national data, but is provided in the current study, is a random comparison group of at-risk students from the same schools and classrooms as the students who received the intervention.

In summary, this study indicated that the RR intervention was effective in reducing the gap between the first-round at-risk children and their average peers by raising at-risk students' literacy levels to a point where they can benefit from classroom instruction and other literacy experiences. Without intervention, the at-risk students identified for second-round service made slow progress in their classroom instructional settings. The measures of intervention efficiency indicated that 14% of the at-risk students served might have been able to make adequate progress in the classroom without an intervention. The cost of serving these children is balanced against the reduction in students who needed long-term literacy support, 5% versus 17% of the first-grade cohort, respectively, with and without early intervention.

Slow rates of literacy learning across first grade can have a cumulative impact, increasing the gap between the lowest achieving students and their average- or high-achieving peers. This pattern can negatively influence a child's entire school experience (Juel, 1988). An effective early intervention can close this achievement gap and substantially reduce the number of students who need long-term literacy support.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Askew, B. J., Fountas, I. C., Lyons, C. A., Pinnell, G. S., & Schmitt, M. C. (1998). *Reading Recovery review: Understanding outcomes & implications*. Columbus, OH: Reading Recovery Council of North America.
- Blachman, B. A. (2000). Phonological awareness. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 483–502). Mahwah, NJ: Erlbaum.
- Brown, A. L., & Campione, J. C. (1985). *Psychological theory and the study of learning disabilities* (Tech. Rep. No. 360). Urbana: University of Illinois, Center for the Study of Reading.
- Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford Press.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental design for research*. Boston: Houghton Mifflin.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly, 30*, 240–263.
- Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2001). Does success in the Reading Recovery program depend on developing proficiency in phonological-processing skills? A longitudinal study in a whole language instructional context. *Scientific Studies of Reading, 5*, 141–176.
- Clay, M. M. (1985). *The early detection of reading difficulties* (3rd ed.). Auckland, New Zealand: Heinemann.
- Clay, M. M. (1987). Learning to be learning disabled. *New Zealand Journal of Educational Studies, 22*, 155–173.
- Clay, M. M. (1993a). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.
- Clay, M. M. (1993b). *Reading Recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.
- Clay, M. M. (2001). *Change over time in children's literacy development*. Portsmouth, NH: Heinemann.
- Clay, M. M. (2002). *An observation survey of early literacy achievement* (2nd ed.). Portsmouth, NH: Heinemann.
- Clay, M. M., & Cazden, C. B. (1990). A Vygotskian interpretation of Reading Recovery. In L. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 206–222). Cambridge, England: Cambridge University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*, 605–619.
- Gómez-Bellengé, F. X. (2001). *Reading Recovery and Decubriendo la Lectura National Report 1999–2000*. Columbus: The Ohio State University, National Data Evaluation Center.
- Gómez-Bellengé, F. X. (2002). *Reading Recovery and Decubriendo la Lectura National Report 2000–2001*. Columbus: The Ohio State University, National Data Evaluation Center.
- Gómez-Bellengé, F. X., & Thompson, J. (2000). *Reading Recovery and Descubriendo La Lectura National Report 1998–1999*. Columbus: The Ohio State University, National Data Evaluation Center.
- Gómez-Bellengé, F. X., & Thompson, J. (2004). *Summary statistics for an observation survey of early literacy achievement tasks in U.S. schools* (Tech. Rep. No. 04). Columbus: The Ohio State University, National Data Evaluation Center.
- Hiebert, E. H. (1994). Reading Recovery in the United States: What difference does it make to an age cohort? *Educational Researcher, 23*(9), 15–25.
- Iversen, S., & Tunmer, W. E. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology, 85*, 112–126.

- 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.
- Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Nicholson, C. L. (1990). *Slosson Oral Reading Test: Revised manual*. East Aurora, NY: Slosson Educational Publications.
- Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly, 38*, 36–76.
- Peterson, B. (1991). Selecting books for beginning readers. In D. E. DeFord, C. A. Lyons, & G. S. Pinnell (Eds.), *Bridges to literacy: Learning from Reading Recovery* (pp. 119–147). Portsmouth, NH: Heinemann.
- Pinnell, G. S. (2000). *Reading Recovery: An analysis of a research-based reading intervention*. Columbus, OH: Reading Recovery Council of North America.
- Pinnell, G. S., & Fountas, I. C. (1999). *Matching books to readers: Using leveled books in guided reading, K–3*. Portsmouth, NH: Heinemann.
- Schwartz, R. M. (1997). Self-monitoring in beginning reading. *The Reading Teacher, 51*, 40–48.
- Schwartz, R. M. (2005). Decisions, decisions: Responding to primary students during guided reading. *The Reading Teacher, 58*, 436–443.
- Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly, 30*, 958–996.
- Stahl, K. A. D., Stahl, S. A., & McKenna, M. C. (1999). The development of phonological awareness and orthographic processing in Reading Recovery. *Literacy Teaching and Learning: An International Journal of Early Literacy, 4*, 27–40.
- Stallman, A. C., & Pearson, P. D. (1990). Formal measures of early literacy. In L. Morrow & J. Smith (Eds.), *Assessment for instruction in early literacy* (pp. 7–44). Englewood Cliffs, NJ: Prentice Hall.
- Stanovich, K. E. (1988). Explaining the difference between the dyslexic and garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21*, 590–604.
- Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly, 26*, 7–29.
- Touchstone Applied Science Associates. (2000). *DRP handbook: J & K test forms*. Brewster, NY: Author.
- Vellutino, F. R., & Scanlon, D. M. (2002). The interactive strategies approach to reading intervention. *Contemporary Educational Psychology, 27*, 573–635.
- Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (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.
- What evidence says about Reading Recovery*. (2002). Columbus, OH: Reading Recovery Council of North America.
- Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*, 159–177.

Received September 25, 2002

Revision received October 20, 2004

Accepted November 17, 2004 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.