

The Short- and Long-Term Effectiveness of Reading Recovery in a Rural School Setting

Hee-sook Choi and Gregg Bieber
University of South Dakota

Abstract

Five years of emergent literacy and literacy data from 2002 to 2007 were reviewed for first through third graders in a small, rural school in the Midwest. Forty first graders had received Reading Recovery services over that time span. Their scores on DIBELS were compared to 41 low average to average students. Subsequent placement in special education and Title I Reading was also tracked. Results indicate that the Reading Recovery students were just as proficient in emergent literacy skills such as phonemic awareness, but they scored significantly lower on a reading fluency measure. Most of them (78%) required subsequent reading services in second and/or third grades. Implications of the findings are discussed.

The passage of the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004) has led to an increased emphasis on the provision of scientific, research-based interventions prior to eligibility determination in special education. IDEA 2004 also allows local education agencies (LEAs) to use as much as 15% of special education monies for intervention services within the regular education curriculum. It is no surprise then that Reading Recovery, one of the widely implemented early intervention programs, is promoted by the Reading Recovery Council of America as “a compelling option for schools that are designing response to intervention (RTI) models to meet the needs of struggling readers and writers” (Lose et al., 2007, p. 1).

Reading Recovery is a broad-based early literacy program founded by Marie Clay in New Zealand in the 1970s (Clay, 1985) that attempts to target the lowest achieving 20% of students in first grade and utilizes one-on-one daily 30 minute tutoring sessions for 12-20 weeks. The goal is to increase students’ reading skills to the class average, allowing them to become relatively independent readers (Baenen, Bernholc, Dulaney, & Banks, 1997; Clay 1991), which, in turn, reduces the number of referrals and placements for special education. To that end, Reading Recovery is considered to be a “cost-effective investment” in preventing reading failure among first graders and reducing long-term costs of special education services in later years (Askew et al., 2003; Lose et al., 2007).

Given the scarcity of resources, particularly in rural schools, the issue of cost effectiveness takes precedence in deciding which intervention program should be implemented. Contrary to the developers’ claims, some scholars suggest that Reading Recovery may be costly due to the one-on-one nature of service delivery and the cost of teacher training and professional development (Hiebert, 1994; Reynolds & Wheldall, 2007). It is therefore important to ask if, in fact, Reading Recovery students do catch up to their average peers and if the gains made during early

intervention are sustainable, reducing the need for additional intensive and expensive reading instruction such as in special education and in Title I Reading services.

Brief Overview of the Findings of Efficacy Studies

Various researchers find that Reading Recovery is an effective short-term intervention for increasing reading skills of low performing students to the local school district's class average (Baenen et al., 1997; Center, Wheldall, Freeman, Outhred, & McNaught, 1995; Hurry & Sylva, 2007; Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994; Plewis, 2000; Reynolds & Wheldall, 2007). However, results from the same above cited studies indicate that long-term effects diminish and wash out over time. For instance, Baenen et al. found that Reading Recovery reduced retention and the need for Chapter 1 services in second grade, but that effect was not present in third grade. By that time, Reading Recovery students were just as likely to be identified for special education services, placed in Chapter 1 services, or retained as students in the control group. Also, on a high-stakes reading assessment, there was no difference between Reading Recovery students and control group students at third grade. On the other hand, other researchers find that Reading Recovery does have long-lasting effects on reading skills (Moore & Wade, 1998; Pinnell, 1989). Moore and Wade, for example, found in their study of 10-12 year olds that former Reading Recovery students scored significantly higher on a standardized test of reading (i.e., Neale Analysis of Reading) than a comparison group chosen from the same class that had higher ability.

The success rates of Reading recovery vary considerably. Up-to-date percentages of successful discontinuation rates of all students, depending on the individual school district in New Hampshire, ranged from 33% to 92% (Schotanus, Chase, Fontaine, Phillips, & Mattson, 2004). Other regional success rates in the Midwest range from 51% to 68% (Banks & Jackson, 2007; Gitz, 2006; Zalud, 2005). The Reading Recovery Council of North America (2002) cited that after 17 years of data collection, 60% of all children served read at a level comparable to the average of their peers. However, Reynolds and Wheldall (2007) contend that the percentages of successful students reported through "in-house" data collections such as those cited above tend to be generally higher than those through independent research. Reading Recovery affiliates also typically cite successful discontinuation rates for only the students who have completed the entire program and these rates are usually higher.

There are also disparities in the rate of recommended students for subsequent reading services. According to Reading Recovery site reports, the average rate was 20% in New Hampshire (Schotanus et al., 2004). The reports from the Midwest (Banks & Jackson, 2007; Gitz, 2006; Zalud, 2005) cite rates from 18% to 26%. Conducted in New South Wales in Australia, Center et al. (1995) found in their study that 35% had benefited from Reading Recovery while 35% were not "recovered" and would need to be recommended for subsequent reading services. They went on to say that 30% of Reading Recovery students would have improved without such services because 30% of the control group improved without any intervention.

Purpose of the Study

Reading Recovery is a widespread reading program that provides interventions for students who are just learning to read. However, the findings from studies examining Reading Recovery are mixed in their support of its benefits given the cost of implementation (For detailed review, see Reynolds, Wheldall & Madelaine, 2009; Schwartz, Hobsbaum, Briggs, & Scull, 2009). In particular, it is unclear as to the long-term effects of the program. If Reading Recovery were to be effective only in the short term, the potential for cost savings would be significantly compromised. Thus, in the present study, the short- and long-term outcomes of Reading Recovery on reading skills are examined, using historical data gathered over a five year period of time.

It should be also noted that the present study employed the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) as reading measures. This responds to one of the major criticisms of Reading Recovery studies: Reading Recovery does not use independent measures of reading skills, as most of its evaluations are limited to the developers' own measures. Therefore, the results may be inflated (Reynolds & Wheldall, 2007; Tunmer & Chapman, 2003). The use of the DIBELS for the present study also reflects current practice: Most LEAs that have adopted an RTI model use curriculum-based measures such as DIBELS for universal screening and progress monitoring (Brown-Chidsey & Steege, 2005). Within an RTI approach, it is likely that outcome measures and/or measures for progress monitoring would be curriculum-based rather than those that are specific to Reading Recovery.

Within in this context were framed the following specific research questions: 1) Do Reading Recovery students catch up to their low average to average peers in reading skills by the end of first grade? 2) Are the gains, if they exist, maintained in second and third grades? 3) To what extent do Reading Recovery students require additional intensive reading instruction, such as special education or Title I Reading services in second or third grade?

Method

Participants

A total of eighty-one first graders from five different cohorts over five years in a small, Midwestern rural school district were included in this study. Data were gathered from each cohort over the course of first, second, and third grades, except from the last cohort, which contains data from only first and second grade. There were 40 students in the Reading Recovery (RR) group and 41 in the comparison group (for detailed descriptions of the groups, see *Procedures*). The RR group consisted of 73% (29/40) male and 27% (11/40) female. In terms of ethnicity, it was composed of 80% (32/40) white, 2% (1/40) African-American, and 18% (7/40) Hispanic students. Forty-five percent of the students (18/40) received free or reduced lunches. In the comparison group, there were 49 % (20/41) male and 51% (21/41) female, whose ethnic breakdown was 90% (37/41) white and 10% (4/41) Hispanic. Twenty-seven percent of students (11/41) received free or reduced lunches.

Measures

Letter Naming Fluency, Phoneme Segmentation Fluency, and Nonsense Word Fluency probes of the DIBELS were administered to first graders as part of universal screening for reading difficulties. For the purpose of this study, the Letter Naming Fluency scores were not used, as it was administered only once in the fall.

Phoneme Segmentation Fluency (PSF) is “a standardized, individually administered test of phonological awareness” (Good & Kaminski, 2002; p. 7). This provides a measure of phonemic awareness as well. It assesses the student’s ability to segment three- and four-phoneme words within a one minute time limit. As determined by the benchmark cutoffs, students may score in the “deficit”, “emerging”, or “established” categories. Students in the established category are at the lowest risk for developing reading difficulties while those in the deficit category are most at risk. Alternate-form reliability ranges between .60 - .70. Concurrent validity is demonstrated by correlations ranging between .19 - .51 with the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R) Readiness cluster and by correlations with the Stanford-Binet Verbal Reasoning, ranging between .20 - .33. Predictive validity is displayed by a correlation with the second grade WJ-R Total Reading cluster, ranging from .20 - .59 and by a correlation with first grade Nonsense Word Fluency, ranging from .28 - .55 (Assessment Committee, 2002).

Nonsense Word Fluency (NWF) is “a standardized, individually administered test of the alphabetical principle – including letter-sound correspondence and of the ability to blend letters into words in which letters represent their most common sounds” (Good & Kaminski, 2002; p. 7). The student is presented with a sheet of randomly ordered vowel-consonant and consonant-vowel-consonant words and then asked to pronounce each letter or the entire nonsense word within a one-minute time limit. Students may score in the “at risk”, “some risk”, or “low risk” categories for developing reading problems. Its alternate-form reliability ranges between .67 - .88. Concurrent validity is denoted by a correlation with the WJ-R Readiness cluster, ranging between .36 - .59, and with the Stanford-Binet Verbal Reasoning, ranging between .17 - .40. Predictive validity is demonstrated by a correlation with the second grade WJ-R Total Reading cluster, ranging between .52 - .77 and by a correlation with second grade CBM-R (i.e., curriculum-based measurement – reading probe), ranging between .60 - .85 (median = .77) (Assessment Committee, 2002).

DIBELS Oral Reading Fluency probes were also utilized, beginning in the winter of first grade through the third grade. DIBELS Oral Reading Fluency (DORF) is “a standardized, individually administered test of accuracy and fluency with connected text” (Good & Kaminski, 2002, p. 8). Students are asked to read a short passage and are timed for one minute to determine how many words correct per minute they can read. They may score in the “at risk” category or in the “some risk” category. If a student scores in the “low risk” category, s/he is much less likely to develop reading difficulties. Alternate-form reliability of second grade passages ranges between .89 - .95 and concurrent validity is demonstrated by correlations ranging between .93 - .96 with the Test of Reading Fluency (Assessment Committee, 2002). DORF passages have also shown moderate

to high correlations (.65 - .80) with high stakes tests and other standardized measures of reading (e.g., Barger, 2003).

Procedures

Phoneme Segmentation Fluency (PSF) and Nonsense Word Fluency (NWF) probes were administered in the fall, winter, and spring to all students attending first grade as part of universal screening for reading difficulties. DIBELS Oral Reading Fluency (DORF) probes were administered starting in the winter of first grade and continued through the third grade. For the aforementioned measures, the standard procedure for administering the DIBELS probes was followed. The students received one, one-minute, probe of PSF and of NWF at each time period. For DORF, three passages were administered at each time period and the median score was used for data analysis. The records of spring administrations of DORF probes for second and third grades were accessed to track intermediate and long-term progress.

Reading Recovery group. First grade students were chosen using a team approach of kindergarten and first grade teachers, a special education teacher, the elementary principal, and a speech-language pathologist. This team of school personnel reviewed the results of the fall administration of DIBELS for first graders and then discussed which students were most in need and those who would benefit the most from Reading Recovery services. Excluding the use of DIBELS, this approach of using a team of teachers is similar to the one used by Schwartz (2005). The kindergarten and first grade teachers gave the primary input during the team meeting.

Of thirty-four first graders, ten (29%) received RR services during the 2002-2003 school year (first cohort). RR typically began shortly after the universal screening that took place in September and continued through February of the following year. For the 2003-2004 school year (second cohort) there were nine of 29 first graders (31%) who received RR. Eight of 27 first graders (30%) of the third cohort and seven of 27 (27%) of the fourth cohort received RR during the 2004-2005 and 2005-2006 school years, respectively. Of twenty-nine first grade students during the 2006-2007 school year (fifth cohort), six (21%) received RR. This resulted in an average of eight students served per year with a total of 40 students in the RR group. In other words, an average of 28% of all students passing through first grade received RR services over the course of 5 years (2002-2007). Attrition occurred in the RR group, one student each in the second and third grade year from the fourth cohort. All 40 first grade students included in the RR group either were successfully discontinued or completed 20 weeks of the RR program. Students are discontinued from the program when they can read texts that their average class peers can read (Clay, 1993).

One RR teacher instructed all of the students receiving such services over the course of the five year span. She had received an entire year of intensive training from a RR Teacher Leader employed by the regional area education agency and was participating in ongoing professional development. RR students were pulled out of the general education classroom during varying times throughout the school day and did not miss regular reading instruction.

Comparison group. The comparison group was selected after the RR students had been chosen. The present study asked whether students with RR caught up to their average peers and, therefore, the comparison group was established to serve as a reference point. Comparison group students were selected using the first grade fall DIBELS data from Letter Naming Fluency, Phoneme Segmentation Fluency, and Nonsense Word Fluency. Students who were not proficient on one or more of the above measures (i.e., students in the “some risk” or “at risk” categories, or for PSF “deficit” or “emerging” categories) were selected to be part of the comparison group. Forty-one students were included in the comparison group and were considered low average to average in emergent literacy skills. There were 9 students in the 2002-03 cohort (26%), 8 in the 2003-04 cohort (28%), 4 in the 2004-05 cohort (15%), 8 in the 2005-06 cohort (30%) and 12 in the 2006-07 cohort (41%). Of forty-one students selected for inclusion in the comparison group, one student each from the first and third cohorts left school in 2nd grade. In addition, three more students left in their 3rd grade year (i.e., 1 from the 3rd cohort and 2 from the 2nd cohort). No students from 2nd grade in either the RR group or the comparison group left prior to the 2nd grade spring DIBELS administration which took place in March.

In addition to the aforementioned attrition in both groups, it should be reiterated that the fifth cohort had data from only first and second grade. Taken together, there were 2nd grade data from 38 students and 39 students in the RR and comparison groups, respectively. In third grade, there were 32 students in the RR group and 24 in the comparison group.

Data Analysis

DIBELS benchmark scores for each administration of probes were used as national norms in recognition of the criticism of RR’s inequitable use of the local classroom average. For 2 X 2 chi-square analyses, the DIBELS “low risk” category was used as the proficient category while the “some risk” and “at risk” categories were combined together to form a non-proficient category. When one cell had fewer than five subjects, Fisher’s Exact Test was used to determine the level of significance (Cohen, 2008).

To determine whether the Reading Recovery group improved to the level of the comparison group containing low average to average students, repeated-measures analysis of variance (ANOVA) was used to compare the group performance means on all DIBELS subtests administered during first and second grade. The Newman-Kuels’ method of comparing means was used to determine whether the differences were statistically significant. For third grade analyses of DORF scores, t-tests were used to compare means because of smaller sample sizes.

Results

Chi-square tests revealed that at the start of the school year there were similar percentages of students who were identified as proficient on PSF in both the RR and comparison groups (58% & 56%, respectively). Similarly, the difference in the percentages of proficient students on NWF in the RR and comparison groups (10% & 22%, respectively) was not statically significant, $p < .23$. In addition, there were no statistically significant differences during the spring

administration of these DIBELS subtests, $p < .25$ and $p < .75$, respectively, which took place after the completion of RR services. One hundred percent of RR students and 95% of students in the comparison group were identified as proficient on PSF. On NWF, 75% and 78% of students in the RR and comparison groups, respectively, reached proficiency.

Likewise, when mean PSF scores were compared via repeated-measures ANOVA, there were no marked differences between the groups during the fall, winter, and spring (see Table 1). When NWF means were compared via repeated-measures ANOVA, there was no statistically significant difference between the groups during the fall. However, there were significant differences between the RR and comparison groups on the winter and spring administrations. In other words, the comparison group attained significantly higher mean scores on the winter and spring administrations (see Table 2).

Table 1.

Phoneme Segmentation Fluency First Grade: Means & Standard Deviations

	Reading Recovery ^a		Comparison Group ^b		Newman-Keuls
	Mean	SD	Mean	SD	
Fall	34.68	13.98	35.65	13.49	ns
Winter	58.55	11.48	56.15	11.96	ns
Spring	61.25	10.70	61.50	11.19	ns

Note: $n^a = 40$ students, $n^b = 41$ students

*Total number of first graders ($N = 129$): Fall ($M = 40.55$, $SD = 14.18$), Winter ($M = 60.39$, $SD = 11.67$), Spring ($M = 62.30$, $SD = 11.27$)

Table 2.

Nonsense Word Fluency First Grade: Means & Standard Deviations

	Reading Recovery ^a		Comparison Group ^b		Newman-Keuls
	Mean	SD	Mean	SD	
Fall	13.30	8.32	20.03	10.38	ns
Winter	48.13	14.85	56.40	21.04	.05
Spring	63.45	22.08	78.75	32.36	.05

Note: $n^a = 40$ students, $n^b = 41$ students

*Total number of first graders ($N = 129$): Fall ($M = 25.95$, $SD = 18.99$), Winter ($M = 62.60$, $SD = 25.39$), Spring ($M = 84.55$, $SD = 34.05$)

Reading Recovery students were significantly less proficient in oral reading fluency than the comparison group during all the times sampled (see Table 3). In terms of performance means,

the comparison group scored significantly higher than the RR group, $F(5, 185) = 90.58, p < .001$, on DORF passages sampled (see Table 4 for Newman-Keuls comparisons). The mean score of the comparison group ($M = 70.92$) was more than one standard deviation greater than that of the RR group ($M = 44.44$) during the spring administration of first grade which took place after the RR students had completed the program. The RR group never did catch up and remained 20 – 30 words correct per minute (wcpm) behind the comparison group throughout the first, second, and third grades.

Table 3.
Oral Reading Fluency –Percent Proficient

	Reading Recovery	Comparison Group	χ^2	p
1 st Grade Winter ^a	50%	80%	8.32	.004
1 st Grade Spring ^a	50%	88%	Fisher's Exact Test ^d	.001
2 nd Grade Spring ^b	39%	85%	16.71	.001
3 rd Grade Spring ^c	34%	67%	5.73	.017

Note: n^a : Reading Recovery = 40, Comparison Group = 41, n^b : Reading Recovery = 38, Comparison Group = 39, n^c : Reading Recovery = 32, Comparison Group = 24
d – two-tailed level of significance

Table 4.
Oral Reading Fluency: Mean Words Correct Per Minute & Standard Deviations

	Reading Recovery		Comparison Group		p
	Mean	SD	Mean	SD	
1 st Grade Winter ^a	21.89	12.13	42.58	24.99	.05 ^d
1 st Grade Spring ^a	44.44	21.58	70.92	24.48	.05 ^d
2 nd Grade Spring ^b	82.50	26.66	113.32	35.27	.05 ^d
3 rd Grade Spring ^c	97.09	22.31	124.76	29.61	.001 ^e

Note: n^a : Reading Recovery = 40, Comparison Group = 41, n^b : Reading Recovery = 38, Comparison Group = 39, n^c = Reading Recovery = 32, Comparison Group = 24

d: Newman-Keuls Comparisons

e: t -test ($t(54) = -4.00, p < .001$)

*Total number of first graders ($N = 129$): 1st Grade Winter ($M = 48.19, SD = 33.49$), 1st Grade Spring ($M = 73.05, SD = 35.92$), 2nd Grade Spring ($M = 111.24, SD = 35.07$), 3rd Grade Spring ($M = 120.24, SD = 31.12$)

A significantly higher number of RR students required reading assistance in second and/or third grade than students in the comparison group. When students who were identified for special education services or Title I Reading services in the second or third grade were collapsed

together and a 2 X 2 chi-square analysis calculated, comparing RR students to the comparison group, the result was statistically significant $\chi^2(1, N = 81) = 22.85, p < .001$. There were 78% (31/40) RR students and 24% (10/41) of the comparison group who required the above services in subsequent grades. Twelve students (30%) in the RR group needed special education services while three (7%) in the comparison group required special education services. All 15 of the students were identified for reading disabilities.

Discussion

This study sought to explore the effectiveness of a rural school district's Reading Recovery program in a Response to Intervention framework. At the start of first grade, there were similar percentages of students in both the RR and comparison groups who were proficient on Phoneme Segmentation Fluency (PSF) and Nonsense Word Fluency (NWF) of the DIBELS. Over the course of the school year, students receiving RR services made similar gains as students in the comparison group on emergent literacy skills in that the increase in the number of "proficient" students on PSF and NWF in the RR group kept pace with those of the comparison group. By the spring of the first grade, the percentages of students who scored in the proficient range on PSF and NWF were essentially the same between the RR and comparison groups. Hence, when the benchmarks set by DIBELS were used as national norms for proficiency purposes, the two groups were fairly equivalent in the progress they made on these emergent literacy skills.

With respect to mean scores of PSF, there were no marked differences between the RR group and the comparison group, and both groups made significant gains throughout the first grade year. By spring, both groups' mean rates on PSF were about the same as the class average (note bottom of Table 1 for class averages). To that end, the RR group met the goal of the Reading Recovery program, which is to bring early literacy skills up to the class average. On the other hand, the RR students did not make as much gain on the mean rate on NWF during the winter and spring administrations in the first grade as the comparison group. By spring, the RR group's performance on NWF was significantly below that of the comparison group as well as the class average (note bottom of Table 2 for class averages).

The performance difference on PSF and NWF may be attributable to the instructional focus of beginning reading skill development in the RR program. Some scholars contend that RR emphasizes the use of context or initial letter cues to predict unknown words and does not focus extensively on reading phonics (Tunmer & Chapman, 2003; Reynolds & Wheldall, 2007). Significantly lower performance of the RR group, when presented with nonsense words on NWF, may be reflective of its "top-down" reading instruction, common to Whole Language (Groff, 2004). Given the instructional focus, RR students would be at a disadvantage especially when presented with nonsense words rather than true words. It is also plausible that the use of the DIBELS may have resulted in dissimilar findings, compared to the studies that relied substantially on portions of Clay's Observation Survey of Early Literacy Achievement to assess beginning reading skills in the alphabetic domain. With Clay's Survey, there would be better alignment between the measures and the constructs taught in RR.

On DIBELS Oral Reading Fluency (DORF), a measure of basic reading skills, the RR students began at a significantly lower rate, even after the completion of RR, and stayed approximately 20-30 wcpm below the comparison group and class average throughout the time periods monitored between the first, second, and third grades (note bottom of Table 4 for class averages). Proficiency dipped below 40% in the second and third grades. RR students lagged significantly behind the comparison group in the short-term (first grade), intermediate (second grade), and long-term (third grade). They never did catch up to the comparison group in mean rate or proficiency, despite additional services in subsequent grades provided to 78% of the RR students. On the other hand, the comparison group's mean rates paralleled with the class averages throughout the time periods and surpassed them in subsequent grades with 24 % of those receiving additional services. These results are inconsistent with What Works Clearinghouse's finding (WWC, 2008) that RR had a "potentially positive" effect on fluency. However, it should be noted that, of the five WWC studies that met the minimum requirement for analysis, only one study demonstrated significant effects on reading fluency for students who received RR by the end of the first grade.

For this particular rural school district, while students in the RR group were initially as proficient as peers in the comparison group in emergent literacy skills, their performance did not generalize to more complex reading tasks such as reading fluency. The RR students began at a significantly lower rate and never did catch up to the comparison group in mean rate or proficiency. This suggests that the effectiveness of RR may be limited in terms of increasing reading fluency skills and maintaining early reading gains into 2nd and 3rd grades.

Response to Intervention programs utilize curriculum based measures (CBM) because they are specific, reliable, and show strong treatment validity (Brown-Chidsey & Steege, 2005). CBM's such as DIBELS also show strong relationships to student performance in the general curriculum (Deno, 2003). Thus, this study utilized DIBELS to identify students needing RR and to monitor their progress. This decision by the authors might have had a potentially important impact on the study. That is, the group of students selected for this study may have had more severe reading problems than do students identified by RR's "in-house" instruments. Gómez-Bellengé, Rogers, & Schulz (2005) found that the students identified as at-risk through the use of DIBELS represented a different population than those students who were identified using Clay's Observation Survey. Of the first grade students identified as needing intervention by the Survey, only 49% were identified as needing intervention through the use of the DIBELS. Some have criticized RR screening measures for over-identifying students who may otherwise find success in the general curriculum (e.g. Center et al, 1995).

Not surprisingly, there were a markedly higher number of RR students who required subsequent assistance with reading skills than the comparison group. Well over three times as many RR students required special education or Title I Reading services in the second or third grade. As stated earlier, 78% of RR students received these services after the first grade, 30% of whom required special education services for reading difficulties. The present study's finding is more in line with Pollock's (1996) estimate of 81% compared to New Hampshire's finding of an average of 20% (Schotanus et al., 2004). Moreover, this finding stands in stark contrast to the 2005-2006 national data by the Reading Recovery Council of North America which showed only 1% of

students who completed Reading Recovery services being placed in special education for LD reading by the end of first grade (National Data Evaluation Center, 2006).

This lackluster performance on actual reading skills and required subsequent assistance with reading may be a representation of the “Matthew Effect” (Stanovich, 1986), which contends that students who are less skilled readers are likely to have more difficulties with reading throughout their school years. In other words, students with poorer reading skills will most likely continue to have poorer skills in the future. Hurry and Sylva (2007) state that early intervention programs (such as Reading Recovery) are commonly utilized to try to prevent the “Matthew Effect”. In the present study, RR students would have to increase their rate of progress in oral reading fluency by several times over to catch up with the comparison group or their average peers.

In summary, there are several unique features of this study. While other researchers have explored the effects of RR in a rural school district, this is the first to look at the longitudinal effects of RR for a rural school setting. Similarly, this study utilized CBM for benchmarking and progress monitoring in lieu of RR’s own observational survey. Despite these methodological changes, the findings were quite similar to those of other non-affiliated studies. Overall, the RR program may be effective in terms of short-term gains in the alphabetic domain such as phonemic awareness. The results of the present study demonstrated that RR fulfills its goal of bringing up the students’ phonemic awareness to the level of average classmates by the end of first grade. However, RR students’ performance in reading fluency continued to lag significantly behind into 2nd and 3rd grade despite additional reading services, attesting to the limited efficacy of RR in terms of long-term efficacy. Consumers of Reading Recovery should be made aware that the intervention does not appear to be sufficient to help struggling readers catch up with peers and stay caught up. Additional resources may need to be in place to help students generalize alphabetic skills into higher level reading abilities such as reading fluency.

Limitations and Future Research

The primary limitation of the present study is that it is not experimental in nature – it was retrospective. Thus, random assignment to either a Reading Recovery group or a comparison group was not feasible. As a result, the comparison group was drawn to serve as a reference point rather than as a control group for causal comparisons, precluding any cause and effect statements as to the effects of Reading Recovery. For instance, certain intervening variables such as the quality of the students’ exposure to print outside of RR including the general curriculum and the home environment may account for some of the discrepancy between the performance of those in the RR group and those in the comparison group. Secondly, this study only investigated the effects of the implementation of RR in one rural school district. Thus, the generalizability of the findings may be limited. Future research should include larger samples of individuals from multiple rural settings. Finally, exploring the connection between emergent literacy and literacy skills and what mechanism(s) operate to transfer one to the other may be of special interest. It is well-founded in the research that a firm grounding in emergent literacy skills is essential for the development of fluent reading, but how does the former generalize to the latter? The answer may be born out of future research.

References

- Askew, B. J., Kaye, E., Frasier, D., Mobasher, M., Anderson, N., & Rodriguez, y. (2003). Making a case for prevention in education. In S. Forbes & C. Briggs (Eds.), *Research in Reading Recovery* (Vol. 2, pp. 133-158). Portsmouth, NH: Heinemann.
- Assessment Committee (2002). *Analysis of reading assessment measures: Dynamic Indicators of Basic Early Literacy Skills*. Retrieved March 31, 2008, from http://dibels.uoregon.edu/techreports/dibels_5th_ed.pdf
- Baenen, N., Bernholc, A., Dulaney, C., & Banks, K. (1997). Reading Recovery: Long-term progress after three cohorts. *Journal of Education for Students Placed At Risk*, 2, 161-181.
- Banks, B., & Jackson, D. (2007, September). *Prairie Lakes AEA 8: Reading Recovery site report*, 8. Pocahontas, IA: Prairie Lakes AEA 8.
- Barger, J. (2003). *Comparing the DIBELS oral reading fluency indicator and the North Carolina end of grade reading assessment*. (Technical Report). Asheville, NCL North Carolina Teacher Academy.
- Brown-Chidsey, R. & Steege, M.W. (2005). *Response to intervention: Principles and strategies for effective practice*. New York: Guilford Press.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly*, 30, 240-263.
- Clay, M.M. (1985). *The early detection of reading difficulties*. Portsmouth, NH: Heinemann Education.
- Clay, M.M. (1991). *The early detection of reading difficulties* (3rd ed.). Auckland, New Zealand: Heinemann Education.
- Clay, M.M. (1993). *Reading Recovery: A guidebook for teachers in training*. Auckland, New Zealand: Heinemann Education.
- Cohen, B.H. (2008). *Explaining psychological statistics* (3rd ed.). Hoboken, NJ: John Wiley & Sons.
- Gitz, B. (2006). *Reading Recovery site report: Keystone Area Education Agency*. Retrieved February 22, 2008, from <http://www.aea1.k12.ia.us/readrecovery/siterep0506.pdf>
- Good, R.H., & Kaminski, R.A. (2002). *Dynamic Indicators of Basic Early Literacy Skills*

- (6th ed.). Eugene, OR: Institute for the Development of Education Achievement.
- Gomez-Bellenge, F.X., Rodgers, E.M., & Schulz, M. (2005). *Reading Recovery and Descubriendo la Lectura national report 2003-2004*. Columbus, OH: Ohio State University.
- Groff, P. (2004). A critical analysis of the sources of Reading Recovery: An empiricist perspective. *Interchange*, 35, 31-58.
- Hiebert, E. (1994). Reading Recovery in the United States: What difference does it make to an age cohort? *Educational Researcher*, 23, 15-25.
- Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading*, 30, 227-248.
- Lose, M.K., Schmitt, M.C., Gomez-Bellenge, F.X., Jones, N.K., Honchell, B.A., & Askew, B.J. (2007). *Reading Recovery and IDEA legislation: Early intervening service (EIS) and response to intervention (RTI)*. Retrieved May 29, 2008, from <http://www.readingrecovery.org>
- Moore, M., & Wade, B. (1998). Reading Recovery: Its effectiveness in the long term. *Support for Learning*, 13, 123-128.
- Pinnell, G.S. (1989). Reading Recovery: Helping at-risk children learn to read. *The Elementary School Journal*, 90, 160-183.
- Pinnell, G.S., Lyons, C.A., DeFord, D.E., Bryk, A.S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 29, 8-39.
- Plewis, I. (2000). Evaluating educational interventions using multilevel growth curves: The case of Reading Recovery. *Educational Research and Evaluation*, 6, 83-101.
- Pollock, J.S. (1996). *Final evaluation report: Reading Recovery program 1995-96*. Columbus, OH: Ohio Department of Program Evaluation.
- Reading Recovery Council of North America. (2002). *What evidence says about Reading Recovery*. Retrieved March 22, 2008, from http://www.rrcna.com/pdf/research/What_Evidence_Says_Full_Report.pdf
- Reynolds, M., & Wheldall, K. (2007). Reading Recovery 20 years down the track: Looking forward, looking back. *International Journal of Disability, Development and Education*, 54, 199-223.

- Reynolds, M., Wheldall, K., & Madelaine, A. (2009). The devil is in the detail regarding the efficacy of Reading Recovery: A rejoinder to Schwartz, Hobsbaum, Briggs, and Scull. *International Journal of Disability, Development and Education*, 56, 17-35.
- Schotanus, H., Chase, C., Fontaine, A., Phillips, E., & Mattson, W. (2004). *Reading Recovery site report for NH 2003-2004*. Retrieved March 22, 2008, from <http://www.ed.state.nh.us/Education/doc/organization/curriculum/reading-recovery-report.pdf>
- Schwartz, R.M. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, 97, 257-267.
- Schwartz, R.M., Hobsbaum, A., Briggs, C., & Scull, J. (2009). Reading Recovery and evidence-based practice: A response to Reynolds and Wheldall (2007). *International Journal of Disability, Development and Education*, 56, 5-15.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences on individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407.
- Tunmer, W., & Chapman, J.W. (2003). The Reading Recovery approach to preventative early intervention: As good as it gets? *Reading Psychology*, 24, 337-360.
- What Works Clearinghouse. (2008). *Reading Recovery*. U.S. Department of Education. Retrieved January 19th, 2009, from http://ies.ed.gov/ncee/wwc/reports/beginning_reading/reading_recovery/
- Zalud, G. (2005). *Reading Recovery 2004-2005 executive summary*. Vermillion, SD: University of South Dakota Regional Reading Recovery Training Center.