



**Evaluation of Short- and Long-term Effects of
Reading Recovery in Montgomery County
Public Schools**

Office of Shared Accountability

June 2008

Nyambura Susan Maina, Ph.D.



OFFICE OF SHARED ACCOUNTABILITY

Dr. Stacy L. Scott, Associate Superintendent
850 Hungerford Drive
Rockville, Maryland 20850
301-279-3925

Dr. Jerry D. Weast
Superintendent of Schools

Dr. Frieda K. Lacey
*Deputy Superintendent
of Schools*

Table of Contents

Executive Summary	vi
Introduction	1
The Reading Recovery Intervention	1
Reading Recovery in Montgomery County Public Schools	1
Review of Literature	3
Evaluation Design	4
Methodology	5
Samples of Students	5
Student Achievement Measures	6
Analysis Procedures	7
Data Availability	7
Results	8
Evaluation Question 1. What are the Short-and Long-term Impacts of Reading Recovery on Students Who Successfully Completed Reading Recovery Compared with a Random Sample of Grade-level Peers?	8
<i>Short-term Impact</i>	8
<i>Long-term Impact</i>	11
Evaluation Question 2. How Do Students Who Successfully Complete Reading Recovery Perform on Local and State Assessments Compared With Grade-level Peers Who Were Similar but not Served in Reading Recovery?	18
<i>Short-term Impact</i>	18
<i>Long-term Impact</i>	19
Discussion	25
Conclusion	26
<i>Recommendations</i>	28
References	29
Appendix A: Elementary Schools Implementing Reading Recovery	32
Appendix B: Demographic Characteristics of Students by Cohort	33
Appendix C: Variables in the Analysis of Covariance Model	34

Appendix D: Calculation of Effect Sizes	35
Appendix E: Reading Outcomes for Random and Reading Recovery Students	36
Appendix F: Grade 2 Outcomes for Random and Reading Recovery Students	41
Appendix G: Outcomes for Reading Recovery and non-Reading Recovery Students.....	44
Appendix H: Assessment Data.....	47

List of Tables

Table 1	End-of-Program Status for Montgomery County Public Schools Reading Recovery Students.....	3
Table 2	Assessment Data Compiled by Cohort Year and Grade.....	6
Table 3	Percentage of Students Meeting Grade 1 End-of-Year Reading Benchmark by Subgroup.....	9
Table 4	Proportion of Students Achieving Grade 2 End-of-Year Benchmark by Subgroup.....	12
Table 5	Proportion of Students at or Above 50 th National Curve Equivalent for TerraNova Second Edition Assessments by Subgroup.....	14
Table 6	Measures of Effect Sizes for Association Between Meeting Grade 1 Benchmark and Participation in Reading Recovery by Cohort.....	19
Table 7	Adjusted Means and Mean Differences for Grade 1 and Measures of Effect Sizes for Association Between Grade 1 Spring Text Reading Comprehension Levels and Participation in Reading Recovery by Cohort.....	19
Table 8	Measures of Effect Sizes for Association Between Grade 2 End-of-year Benchmark Status and Participation in Reading Recovery by Cohort.....	20
Table 9	Adjusted Means and Mean Differences for Grade 2 Text Reading Comprehension Levels and Measures of Effect Sizes for Association Between Reading Achievement and Participation in Reading Recovery by Cohort.....	21
Table 10	Measures of Effect Sizes for Association Between Scoring at or Above the TerraNova Second Edition 50th National Curve Equivalent and Participation in Reading Recovery by Cohort.....	22
Table 11	Adjusted Means and Mean Differences for TerraNova Second Edition Mathematics Scale Scores and Measures of Effect Sizes for Association Between Reading Achievement and Participation in Reading Recovery.....	22
Table 12	Measures of Effect Sizes for Association Between Proficiency on the Maryland School Assessment and Participation in Reading Recovery by Cohort.....	23

Table 13 **Adjusted Means and Mean Differences for Grade 3 Reading Measures and Measures of Effect Sizes for Association Between Reading Achievement and Participation in Reading Recovery by Cohort** 24

List of Figures

Figure 1	Proportion of students at or above Grade 1 end-of-year benchmark by cohort year	8
Figure 2	Proportion of students achieving Grade 2 end-of-year benchmark	11
Figure 3	Proportion of random sample and reading recovery students at or above the TerraNova Second Edition 50 th National Curve Equivalent.....	13
Figure 4	Proportion of students performing at or above proficiency on Grade 3 reading Maryland School Assessment	15
Figure 5	Proportion of students performing at or above proficiency on reading Maryland School Assessment by race/ethnicity.....	16
Figure 6	Proportion of students performing at or above proficiency on the Grade 3 reading Maryland School Assessment by service receipt.....	16
Figure 7	Proportion of Reading Recovery and non-Reading Recovery students meeting the Grade 1 end-of-year benchmark.....	18
Figure 8	Proportion of Reading Recovery and non-Reading Recovery students meeting the Grade 2 end-of-year benchmark.....	20
Figure 9	Proportion of Reading Recovery and non-Reading Recovery students at or above the TerraNova Second Edition 50 th National Curve Equivalent.....	21
Figure 10	Percentage proficient and above on the Grade 3 reading Maryland School Assessment.....	23

Executive Summary

This report presents findings on the short- and long-term impact of the Reading Recovery (RR) intervention on student literacy achievement. RR, a research-based early intervention program is one of the targeted initiatives implemented in Montgomery County Public Schools (MCPS) since 1997 to improve student achievement in elementary schools. This initiative supports Goal 1 of the MCPS strategic plan *Our Call to Action: Pursuit of Excellence* (MCPS, 2007). Delivered in Grade 1, the RR intervention is for those students reading below grade level, always among the lowest 25% of the class. The intervention is designed to accelerate the literacy skills of the lowest performing Grade 1 students from the bottom of their class to the average, where they can profit from regular classroom instruction. Approximately 300–470 students participate in RR in a given year.

The following two questions guided the evaluation study:

1. What are the short- and long-term impacts of RR on students who successfully complete the program, as measured by local and state assessments compared with a random sample of grade-level peers?
2. How do students who successfully complete RR perform on local and state assessments compared with grade-level peers who were reading below grade level by the end of kindergarten or fall of Grade 1 (similar in initial reading ability to RR students) but not served in RR?

Achievement data for three successive cohorts of RR students—2003–2004 (cohort 1), 2004–2005 (cohort 2), and 2005–2006 (cohort 3) were used to address the evaluation questions. The performance of RR participants was compared with two groups: stratified randomly sampled grade-level peers and students similar in reading ability to RR participants but who did not participate in RR.

To assess the short-term impact, the study included analyses of the average performance of RR and comparison students within each cohort at the end of Grade 1. To assess long-term impact, the performance of RR students was compared with the performance of students in the comparison groups at the end of Grades 2 and 3.

Key Findings

Short-term impact

- **The short-term impact of participating in RR was positive.** Within cohorts, the performance of RR students on spring Grade 1 text reading comprehension levels and rates of meeting the MCPS Grade 1 end-of-year benchmark was comparable with or significantly exceeded the performance of their random sample grade-level peers. The extent of the effect of association between participation in RR and Grade 1 reading achievement varied depending on cohorts and measures examined and the effect was strongest for students in cohorts 2 and 3. Also, medium to large effect sizes for students in cohorts 1, 2, and 3 confirmed the positive effect of RR.

- **Within each cohort, RR students as a group showed less variation in their performance on the Grade 1 reading measures than students in the random sample comparison group.**
- **The spring text reading comprehension levels and the rates of meeting the Grade 1 end-of-year benchmark for RR students were significantly higher than levels for students reading below grade level by spring of kindergarten and fall of Grade 1 (similar peers) who did not receive RR.** The corresponding large and medium effect sizes for association between participation in RR and reading achievement for students in cohorts 1, 2, and 3 confirmed the marked effect of RR.

Long-term impact

- **Contrary to a number of longitudinal studies conducted in the United States, the study did not find sustained effects of RR in Grade 2.** The findings for Grade 2 reading outcomes vary by measure and cohort being examined. Overall, across cohorts, only one third of RR students met the Grade 2 end-of-year reading benchmark, and even fewer students placed at or above the 50th National Curve Equivalent (NCE) for TerraNova Second Edition (TN/2) assessments. Contrary to the expectation that RR students would perform at comparable levels with their grade-level peers, across the three cohorts statistically significant differences for Grade 2 spring text-reading levels and TN/2 scale scores were found in favor of randomly sampled peers. The small negative effect sizes for Grade 2 reading measures confirmed that the performance of the RR students lagged behind that of their random sample grade-level peers.
- At the same time, the performance of RR students on the Grade 2 text reading comprehension levels was comparable with or lower than the performance of similar students who did not receive RR. The negligible effect sizes signify that any slight differences were not meaningful in an educational setting. Furthermore, significant differences in the TN/2 scores between RR and similar students were found in favor of similar students in all three cohorts. The related small to medium negative effect sizes for association between participation in RR and TN/2 scores verify that the performance of students in RR lagged behind the performance of similar students not in RR.
- **The findings for Grade 3 reading outcomes vary by measure and cohort being examined.** In general, more one than half of the RR students were at or above proficiency on the Grade 3 Maryland State Assessment (MSA) in reading. The Grade 3 reading MSA proficiency rates for RR students in cohort 1 were significantly lower than the rates for their randomly sampled peers. Also, the difference in adjusted MSA reading scale scores for RR and random sample peers in cohorts 1 and 2 were statistically significant in favor of randomly sampled students. The small negative effect sizes confirm that the performance of RR students lagged behind that of their random sample grade-level peers. Moreover, the difference in adjusted MSA scale scores for RR and similar students in cohort 1 was not statistically significant. Negligible effect sizes confirm that the performance of cohort 1 RR students was comparable with similar students who did not participate in RR. On the other hand, the differences in adjusted mean MSA scores for RR and similar students in cohort 2 were significantly different in favor of similar students who did not participate in RR. The

small negative effect sizes confirm that the performance of RR students on the Grade 3 reading MSA lagged behind that of similar grade-level peers who did not receive RR.

- **Significant differences in Grade 3 spring Measures of Progress in Reading (RIT) scores were found in favor of random sample students in cohorts 1 and 2.** The medium effect sizes confirm that the performance of RR students lagged behind that of their random sample peers. Moreover, significant differences in Grade 3 spring RIT scores were found in favor of similar students in cohort 2 but not for cohort 1. The small negative effect sizes confirm that the performance of cohort 2 RR students lagged behind that of similar grade-level peers who did not receive RR.

Recommendations

- **Investigate factors that contribute to low reading proficiency levels of RR students and students receiving special services on the Grade 2 end-of-year benchmark and TN/2 assessments.** Reasons for RR students falling behind in Grade 2 are unclear, particularly since the Grade 2 findings counter findings reported by others (United States Department of Education (USDE, 2007). In general, districtwide trends show that less than 50% of students receiving special services and students from the African American and Hispanic subgroups meet the Grade 2 end-of-year reading benchmark or score above the 50th NCE on the TN/2. Furthermore, the greatest variability among schools and within subgroups in the district occurs for the Grade 2 end-of-year reading benchmark and on TN/2 (Chesney & Fink, 2007; Douglas, 2006; Zhao & Von Secker, 2007). Therefore, understanding and addressing issues related to reading performance in Grade 2 is critical for closing the achievement gap among student subgroups and schools.
- **Establish targeted training of Grade 2 teachers and reading specialists to support former RR students and students receiving special services.**
- **Consider the development of appropriate Grade 2 reading intervention programs for former RR students and other low-performing students.** In particular, appropriate interventions matched with specific needs of students for schools with less than 75% of students achieving the spring Grade 1 and 2 end-of-year reading benchmark are needed.
- **Add a marker to Online Administrative Students Information System (OASIS) to identify students who received RR each year.** This will enable schools and researchers to track and monitor performance of RR students over time.
- **Monitor progress of RR students in Grade 2 and provide additional support as needed.** Eliciting classroom teachers' observations of RR students before RR, changes in students after RR, and areas of further needs for RR students after the intervention will help teachers target their support to the specific needs of the students at each stage.

Evaluation of Short- and Long-term Effects of Reading Recovery in Montgomery County Public Schools

Nyambura Susan Maina, Ph.D.¹

Introduction

Reading Recovery (RR) is one of the targeted initiatives implemented in Montgomery County Public Schools (MCPS) since 1997 to improve student achievement in elementary schools. These initiatives support Goal 1, “Ensure success for all students,” of the MCPS strategic plan *Our Call to Action: Pursuit of Excellence* (MCPS, 2007). Delivered in Grade 1, the goal of RR is to provide intensive support to the lowest achieving first graders to enable them to attain grade-level literacy skills. During the 2006–2007 school year, RR was implemented in 33 elementary schools. Between 300–470 students receive RR in a given year.

The Reading Recovery Intervention

The RR intervention is designed to accelerate the reading and writing performance of the lowest performing Grade 1 students in a short time from the bottom of their class to the average, where they can profit from regular classroom instruction. The first expected outcome of RR is that most students accelerate their learning so that they are able to learn in the classroom and meet grade-level expectations. The second expected outcome of RR is that the children who need long-term learning support are identified and referred for further evaluation.

Prior to the intervention, an Observation Survey (OS) of Early Literacy Achievement is administered to the lowest achieving first graders. The OS includes six tasks: Letter Identification, Ohio Word Test, Concepts about Print, Writing Vocabulary, Hearing and Recording Sounds in Words, and Text Reading (Clay, 2002).

During the intervention, individual students receive a half-hour lesson with specially trained RR teachers daily for 12–20 weeks. As soon as the students accelerate their reading to be within the average range in their class and demonstrate that they can continue to achieve at or above grade level, their daily RR lessons are discontinued and new students begin individual instruction.

The six tasks of the OS also are administered to RR students at the conclusion of the program. Year-end OS scores serve as the posttest measures in judging the progress made by RR students. In addition to strong performance on the OS, the students who successfully complete RR are expected to demonstrate the ability to work well within their classroom settings.

Reading Recovery in Montgomery County Public Schools

Setting. Since 1999, all federally funded Title I schools (with the exception of Reading First schools¹) with Grade 1 students are staffed with at least one .5 full-time equivalent (FTE) RR teacher. In addition, non-Title I schools are allocated a .5 RR teacher based on need, as identified by district data. By the 2006–2007 school year, 17 Title I schools and 16 non-Title I schools implemented RR (See Appendix A). Also, some schools elect to use a .5 supplemental position (such as focus teacher or academic intervention teacher) for RR.

¹ Highland, Rosemont, Summit Hall, and Wheaton Woods elementary schools

In addition, MCPS is an authorized site for RR through Ohio State University. The MCPS RR teacher training site is housed at Broad Acres Elementary School. One full-time RR teacher leader provides the training for teachers and coordinates the MCPS RR program.

Eligibility for RR. Kindergarten teachers complete ranking sheets in June, listing their current students' literacy skills from high to low, based on the MCPS Assessment Program Reading (*mClass 3D*) and informal classroom literacy assessments. RR teachers first use the kindergarten ranking sheets to select the lowest children to begin testing for RR on the morning of the first day of school. Grade 1 teachers are asked to recommend any other students they are concerned about and any students new to the school. No child is excluded from screening, except those who have been retained in Grade 1 or those with Individual Educational Programs (IEPs) for reading and writing. Accordingly, RR teachers make sure they are administering the OS to approximately 25%–30% of the Grade 1 students they serve. The data collected are recorded on “selection sheets” for the “selection meeting for first-round students.” There is no preestablished cut-off score. The lowest students are selected first, looking across all six assessments of the OS (Randall, 2007).

The MCPS RR site teacher leader collects and evaluates all selection sheets from each school to verify that the students selected are indeed the lowest performing students, based on assessment data in each school. The RR school team makes a preliminary decision regarding who is served.

Level of coverage. Level of coverage refers to the percentage of Grade 1 students served by RR. Full implementation of the program requires enough RR teaching time to ensure that all Grade 1 students, in need of additional assistance (approximately 25%), have access to RR (Randall, 2007). Full implementation also means every child needing RR has the opportunity to be served.² The majority of the schools were partially implementing the intervention.

End-of-program status. At the end of the intervention, RR students are assigned to one of the following end-of-program status categories:

- **Discontinued:** A child who successfully met the criteria to be discontinued from the intervention during the school year or at the time of year-end testing.
- **Recommended Action After a Complete Intervention:** A child who was recommended by RR professionals for assessment/consideration of other instructional support at the point of departure from RR, after receiving a full series of lessons of at least 20 weeks.
- **Incomplete Program at Year-End:** A child who was still in RR at the end of the school year with insufficient time (less than 20 weeks) to complete the intervention.
- **Moved While Being Served:** A child who moved out of the school while being served before specific program status could be determined and who may or may not have had a full program of 20 weeks.
- **None of the Above:** A rare category used only for a child who was removed from RR under unusual circumstances, with fewer than 20 weeks of instruction (Randall, 2004).

Table 1 presents information regarding end-of-program status for MCPS students who received RR during the 2003–2004 through 2005–2006 school years. More than one half of students were

² 2003–2004: Broad Acres and New Hampshire Estates; 2004–2005: Broad Acres, New Hampshire Estates, and Rolling Terrace; 2005–2006: New Hampshire Estates and Rolling Terrace.

successfully discontinued from the program each year, ranging from 55%–66%. The national average is 58%–59% for the same period (Randall, 2005; 2006).

Table 1

End-of-Program Status for Montgomery County Public Schools Reading Recovery Students

Status	Cohort					
	2003–2004		2004–2005		2005–2006	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Discontinued	214	55.0	212	66.0	195	58.0
Recommended action after full program of 20 weeks or more	88	23.0	49	15.0	52	15.0
Incomplete program at year end	60	15.0	46	14.0	63	18.0
Moved while being served	14	4.0	10	3.0	19	6.0
None of the above	14	4.0	6	2.0	11	3.0
Total	390	100.0	323	100.0	340	100.0

Source: RR program data. Columns do not add up to 100.0 due to rounding.

Review of Literature

Evidence from literature points out that RR provides compensatory reading benefits, reduces the achievement gaps that exist along racial and socioeconomic lines, and provides accelerated reading benefits (Schmitt & Gregory, 2001; West, Denton & Reaney, 2000; Swartz & Hoffman, 1993; Pinnell, 1989; D’Agostino & Murphy, 2004).

Compensatory reading benefit. Studies by Schmitt and Gregory (2001; 2005) demonstrated that the majority of randomly selected RR children were performing as well as a random sample of their grade-level peers on the vocabulary and comprehension subtests of the Gates-MacGinitie Reading Test and running records of oral text reading on the OS. Similarly, Grade 3 students’ results on the Comprehensive Test of Basic Skills-5/TerraNova Form B reflected a normal distribution with a mean at the 45th percentile for those students who were the lowest achieving first graders before receiving the program.

Reduction in achievement gaps along racial and socioeconomic lines. Mounting statistical evidence continues to document that achievement gaps lie between the races and also between more or less economically disadvantaged students (Lee, 2002; West, Denton & Reaney, 2000). On the basis of what has been observed about the achievement gap and who is affected, the expectation is that not only will children fall behind their peers along racial and economic lines but also this gap may stay with them over time. One study evaluated whether students who receive RR make differential progress along racial and economic lines (Rodgers, Wang, & Gomez-Bellenge, 2004). The findings concur with those of Pinnell et al. (1994) and indicate that an early intervention such as RR is sufficient to make a difference in closing the literacy gap that exists along racial and economic lines. Because of an early intervention such as RR, these findings suggest that race and socioeconomic factors ceased to be significant predictors of the reading success of Grade 1 students. In addition, findings by Ashdown & Simic (2000) suggest that RR produces similar outcomes for students with different levels of English proficiency and offers an appropriate solution for Grade 1 students experiencing problems in reading and writing.

Accelerated reading benefit of RR. The success of RR in accelerating reading has been carefully documented (Swartz & Hoffman, 1993; Swartz, 2005; Pinnell, 1989; USDE, 2007). In these studies, students entered as nonreaders, discontinued at a level considered to be the end of Grade 1, and, by the end of Grade 1, had accelerated their reading to a level equivalent to Grade 2. Other studies indicate that students who participated in RR maintained progress in Grades 3, 4, and 5 (Tlawar & Hill, 1993). Pinnell (1989) also reported that Grade 4 RR students demonstrated that they could accurately read text at or above the Grade 6 level.

Skepticism about RR. Because it is widely studied and well known, RR has drawn criticism from several researchers and educators. Some critics assert that a) the RR data reporting system is flawed; b) the standard for successful completion of RR is not equitable across schools and participants; c) RR does not raise overall school achievement levels; d) some studies do not use strong causal design, and some researchers report misleading success rates; or e) some of the findings are derived from unorthodox research procedures (Elbaum, Vaugh, Huges, & Moddy, 2000; Shanahan & Barr, 1995; Pinnell & Lyons, 1994).

Overall, the evidence that RR is effective in a variety of contexts far outweighs the criticisms. A recent study reported that RR has outcomes of educational significance for the majority of the participating students (Timperley, Fung, Wilson & Barrar, 2006).

Evaluation Design

This study follows a quasi-experimental design and compares the achievement of students who did not participate in RR intervention with students who did. Specific evaluation questions were as follows:

1. What are the short- and long-term impacts of Reading Recovery on students who successfully completed Reading Recovery as measured by local and state assessments compared with a random sample of grade-level peers? Are there differential patterns of reading achievement among student subgroups?
2. How do students who successfully complete RR perform on local and state assessments compared with grade-level peers who were similar in reading ability/eligible but were not served by RR intervention?

For evaluation question 1, if RR is effective, the expectation is that, after the intervention, the performance of RR students will be comparable with that of their random sample peers. For evaluation question 2, the expectation is that the performance of the RR students will be better than the performance of similar students who did not participate in RR.

Short-term impact denotes the performance of students immediately after the intervention at the end of Grade 1. Long-term impact corresponds to the performance of the students at the end of Grades 2 and 3, which are 2 and 3 years after the RR intervention.

Methodology

Samples of Students

As noted earlier, students are not randomly assigned to RR. Consequently, experimental designs are not possible to employ. A partial solution is to use a quasi-experimental design, which includes an appropriate comparison group. This study also used statistical controls to balance the groups, due to lack of random assignment of students to the treatment or comparison groups.

To examine whether a) RR students catch up with their grade-level peers and b) whether RR students perform better than similar students who did not receive RR necessitated comparing the performance of RR students with two groups of students. The appropriate comparison groups are a random sample of peers and a sample of students who were reading below grade-level by the start of Grade 1 who did not participate in RR.

Treatment groups (RR). The treatment groups in this study were three successive cohorts of students who completed and were successfully discontinued from Reading Recovery during the 2003–2004, 2004–2005, and 2005–2006 school years.

Comparison groups. The following procedures were used to create appropriate comparison groups. First, enrollment data for all students who were in Grade 1 during the 2003–2004, 2004–2005, and 2005–2006 school years and student demographic information pertaining to race/ethnicity, gender, special services received (English for Speakers of Other Languages—ESOL, Free and Reduced-price Meals Services—FARMS, and special education services) and average attendance for a given year were compiled from a data source for yearly summary data for students enrolled in MCPS. Second, a sampling technique was used to draw the following stratified (by ESOL, FARMS, race/ethnicity, and gender) random samples of grade-level peers:

1. *Random groups.* This included randomly selected students who were not exposed to RR at any time. A sample was drawn for each cohort year. These stratified random samples were used as the comparison groups for evaluation question 1.
2. *Non-RR groups.* To form an appropriate comparison group for addressing question 2, a randomly selected sample of students was drawn from a group of students of similar demographic profile and reading ability to those enrolled in RR. These students were drawn from a data set of students who were reading below grade level by spring of kindergarten and fall of Grade 1—at risk of reading failure. These students are defined as those who were reading at Text Reading Comprehension (TRC) level of 0–2 at the end of kindergarten. These students did not receive RR because a) there was no RR intervention program in their schools or b) if their school implemented RR, the implementation was partial and these students were not ranked among the lowest 25% in their class. A sample was drawn for each cohort year. These samples were used for analyses related to evaluation question 2.

The stratified random and non-RR comparison groups closely matched and corresponded to the number and profile of students enrolled in RR in a given year on the basis of race/ethnicity, gender, FARMS, and ESOL status (See Appendix B).

Student Achievement Measures

Local assessments. The following local assessments were used in this evaluation:

1. Montgomery County Public Schools Assessment Program-Primary Reading (MCPSAP-PR). MCPSAP-PR is designed to measure foundational reading skills and reading proficiency in kindergarten, Grade 1, and Grade 2. For this evaluation, the outcome measures of Grades 1 and 2 end-of-year benchmark status and the students' TRC levels were used. The MCPS end-of-year benchmarks are specified as follows:
 - a. Grade 1—reading at TRC level 16 or higher with 90% accuracy and achieving a score of 4 out of 5 on oral comprehension.
 - b. Grade 2—reading at TRC level M (27) or higher with 90% accuracy, having an oral comprehension score of 4 out of 5, and scoring at least 2 out of 3 on two written comprehension items (Chesney & Fink, 2007).
2. TN/2.³ Grade 2 NCE and reading scale scores were analyzed. The national norm or average score for all TN/2 tests is the 50th NCE (Zhao & Von Secker, 2007).
3. Measures of Academic Progress Assessment in Reading (MAP-R). Grade 3 2006 and 2007 MAP-R Rasch Unit (RIT) scores were used for the analyses— a measurement scale developed to simplify the interpretation of test scores. RIT scores range from about 150 to 300. RIT scores make it possible to follow a student's educational growth from year to year.

State assessments. Grade 3 test scores on the 2006 and 2007 Maryland School Assessment (MSA) reading subtest and the corresponding reading proficiency levels were used in the evaluation. Cohort 3 was in Grade 2 in 2006–2007 and their MSA scores were not available because this is not given until spring semester Grade 3. Table 2 presents the data on the year each cohort participated in RR instruction and the information compiled for each cohort of students.

Table 2
Assessment Data Compiled by Cohort Year and Grade

Reading Recovery Participation Year	Grade as of 2006–2007	Available state and district assessment data
2003–2004 (cohort 1)	4	<ul style="list-style-type: none"> • 2003–2004 and 2004–2005 MCPSAP-PR, end-of-year benchmark status, TRC level • 2005 CTBS NCE and scale scores • 2006 MSA—Grade 3 scale score and proficiency level • 2006 MAP-R—Grade 3
2004–2005 (cohort 2)	3	<ul style="list-style-type: none"> • 2004–2005 and 2005–2006 MCPSAP-PR, end-of-year benchmark status, TRC level • 2006 TN/2 NCE and reading scale score • 2007 MSA Grade 3 scale score and proficiency level • 2007 MAP-R—Grade 3 scale score
2005–2006 (cohort 3)	2	<ul style="list-style-type: none"> • 2005–2006 and 2006–2007 mClass 3D- MCPSAP-PR, end-of-year benchmark status and TRC level • 2007 TN/2 reading NCE and reading scale score

³ TN/2 is not locally developed but is a locally administered norm-referenced test.

Analysis Procedures⁴

The analyses conducted for evaluation questions 1 and 2 are designed to a) describe the performance of RR students relative to the comparison group and b) measure the net short- and long-term effects of the RR program on student achievement. Three different analytical procedures were applied to address the questions: descriptive analyses of performance levels, analyses of covariance (ANCOVA) for continuous outcomes, and logistic regression models for categorical outcomes measures. The following paragraphs describe the analyses used to address evaluation questions 1 and 2.

1. To statistically control for the nonequivalence of the RR and comparison groups and to isolate the program effects of RR, a propensity score was computed. The propensity score is described as a conditional probability that a student will be “treated” (be placed in RR) based on an observed group of covariates (Lueleen, Shadish & Clark, 2006). Propensity scores created by ESOL, FARMS, special education status, and race/ethnicity, and a measure of initial reading ability were estimated using logistic regression procedures. Because two groups could be similar on demographic variables but dissimilar on academic measures, a measure of initial reading ability by fall of Grade 1 was included in the computation. A separate propensity score was estimated for each yearly cohort. Subsequently, the propensity scores were categorized into quintiles and included in the ANCOVA and logistic regression models.
2. ANCOVA and logistic regression analytic models were developed to isolate the effects of RR, while simultaneously controlling for other preexisting or intervening variables, such as participation in FARMS, ESOL, special education services, race/ethnicity, initial ability at each grade level, plus a propensity score. For each reading achievement measure, student data were analyzed separately for each cohort and at Grades 1, 2, and 3. The dependent variables were the spring reading measures at each grade level. The independent variable of RR status included two levels: participation or nonparticipation in RR (See Appendix C, Tables C1–C2).
3. Effect sizes were computed to assess the association between participation in RR and student achievement in reading or the practical significance of the RR intervention for each outcome measure and at each grade level (See Appendix D). In evaluating educational programs, effect sizes of 0.20 and above are considered important by educational researchers, i.e., the larger the effect sizes, the stronger the impact of the intervention (Slavin, 1990; Cohen, 1988).

Data Availability

Only students who were successfully discontinued from RR (Table 1) were included in the analyses. Also, to compare the proportion of students achieving a certain grade-level benchmark and/or the performance of students on other achievement measures, the scores of students who had valid scores for a specific measure were included. Therefore, the number of students varied at each step of the analyses and the same students could not be compared at each assessment.

⁴ The author would like to thank Dr. Scot McNary, associate professor, Towson University, for his assistance with reviewing the analytical procedures and analyses used in this study.

Results⁵

The results are organized by evaluation questions.

Evaluation Question 1. What are the short- and long-term impacts of reading recovery on students who successfully completed reading recovery compared with a random sample of grade-level peers? Are there differential patterns of reading achievement among student subgroups?

Short-term Impact

The achievement data on Grade 1 reading measures were analyzed in three ways. First, achievement was considered relative to grade-level MCPSAP-PR benchmark. More specifically, these analyses investigated what percentage of students from the random and RR samples met the spring TRC and comprehension criteria for Grade 1 end-of-year reading benchmark. Second, summary statistics of spring TRC levels were calculated. Third, while controlling for preexisting differences, the impact of RR on student performance was estimated.

Grade 1 end-of-year reading benchmark. After participation in RR, students are expected to be reading at or above grade level and also meet the reading benchmark at the same rate as their grade-level peers. Information on performance of students on the Grade 1 end-of-year benchmark is presented in Figure 1.

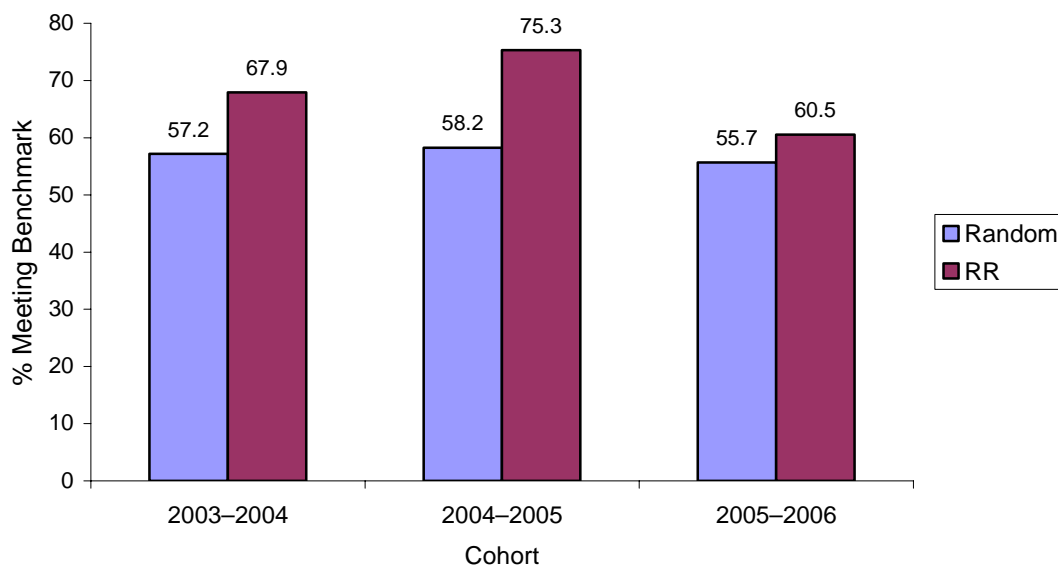


Figure 1. Proportion of students at or above Grade 1 end-of-year benchmark by cohort year.

For students in cohorts 1, 2, and 3, the passing rates for RR students on the Grade 1 end-of-year benchmark were significantly higher than rates for their random sample peers ($p < 0.05$).

⁵ This study assumes that the RR intervention was implemented as designed.

After controlling for preexisting differences, small effect size was found for cohort 1 students (0.27) whereby large and moderate effect sizes for cohort 2 (0.87) and cohort 3 (0.51) were found (Appendix E, Table E1a through E1d). The small to moderate to large effect sizes for cohorts 1, 2, and 3 indicate that the differences in the proportions for the two groups were educationally meaningful. That is, the average RR students in cohorts 1, 2, and 3 performed at the 62nd, 79th, and 69th percentile of the random sample students, respectively.

Grade 1 end-of-year benchmark by subgroup. Comparing the performance of students among groups within cohorts, RR students in cohorts 2 and 3 outperformed their randomly sampled grade-level peers and this pattern of performance was reflected for the racial/ethnic subgroups, except for Asian Americans. The lowest passing rate for RR students was among the Hispanic students in cohort 1 (Table 3).

Historical trends indicate that passing rates of students receiving special services (e.g., ESOL, FARMS, and special education) on the Grade 1 end-of-year benchmark are lower than passing rates for students not receiving these services (Chesney & Fink, 2007). The same trends were observed for students in this study, with a few exceptions: Cohort 2 RR students who received FARMS services outperformed their Non-FARMS counterparts. Additionally, cohort 3 RR FARMS students also outperformed the random sample Non-FARMS peers. Notably, for every cohort, RR students who received ESOL services outperformed their ESOL counterparts in the randomly selected sample (Table 3).

Table 3

Percentage of Students Meeting Grade 1 End-of-Year Reading Benchmark by Subgroup

Subgroups	Cohort 1 (2003–2004)				Cohort 2 (2004–2005)				Cohort 3 (2005–2006)			
	Random		RR		Random		RR		Random		RR	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Race/Ethnicity												
African American	72	64.9	60	54.6	55	60.4	36	83.7	68	59.6	51	61.4
Asian American	21	72.4	18	72.0	14	73.7	7	53.8	17	70.8	12	60.0
Hispanic	81	46.3	68	43.9	77	50.7	65	79.3	81	47.9	83	58.5
White	31	72.1	30	62.5	38	70.4	25	75.8	18	72.0	14	77.8
Gender												
Female	86	54.8	72	50.4	85	55.6	52	77.6	85	62.0	76	63.3
Male	120	59.1	104	53.3	99	60.7	81	79.4	101	51.3	85	58.2
Special Services												
Non-FARMS	80	71.4	74	64.9	85	66.4	51	75.0	66	62.3	59	74.7
FARMS	126	50.8	102	45.5	99	52.7	82	79.6	120	52.6	102	54.5
Not ESOL	137	72.9	109	58.3	123	64.1	79	84.0	126	72.4	83	70.9
ESOL	69	40.1	67	44.4	61	49.2	54	70.1	43	35.8	65	52.8
Not Special Ed.	198	60.0	165	53.6	172	61.6	108	78.8	163	64.7	131	63.3
Special Ed.	8	26.7	11	36.7	12	32.4	5	71.4	5	13.2	8	36.4

Note. RR= Reading Recovery. Results for American Indian and subgroups with fewer than five students are not reported.

Grade 1 spring TRC levels. The students in the RR sample started Grade 1 with fall TRC levels that were significantly lower than those of their grade-level peers. While comparing the performance of students among groups within each cohort, the data show that mean Grade 1 TRC levels for the students in RR and random samples increased substantially and became very similar by spring of Grade 1 (See Appendix E, Table E2). As graphically illustrated in Appendix

E, Figures E1a–1c, page 37, the distribution of RR students’ mean spring Grade 1 TRC levels approximated the spread of the spring TRC levels for students in the random sample. Clearly, the TRC levels of the RR students were no longer clustered at the low end of the distribution.

Grade 1 spring TRC levels by subgroup. Across cohorts, the students in the RR sample made average gains of 10–12.3 TRC levels compared with 9–10 TRC levels for their random sample counterparts. Each racial/ethnic group benefited from RR, with White and Asian American students making higher gains in mean spring Grade 1 TRC levels than other subgroups (See Appendix E, Tables E3–E5). Asian American students in the cohort 1 RR sample did slightly better than Asian Americans in the random sample, whereas those Asian American students in cohort 2 did not. The difference in performance of African American students who got RR and those who did not was less than one TRC level. White students who had RR performed as well (cohort 1 and 2) or better (cohort 3) than White students in the random sample. The performance of Hispanic students varied depending on the cohort. Additionally, the performance of cohort 1 RR Hispanic students was comparable with the performance of Hispanic students in the random sample; cohort 2 Hispanic students in the random sample did better than Hispanic students in the RR sample; cohort 3 Hispanic students in the RR sample outperformed their random sample Hispanic counterparts.

Grade 1 students are expected to be reading at TRC level 16 or above by spring of Grade 1. To be performing as well or better than the students in the random sample, the ANCOVA should show no significant differences between the spring Grade 1 TRC levels for RR and random students and/or show significant differences in favor of the RR students. Also, the effect sizes for association between participation in RR and performance on the reading measures to assess the practical significance of the intervention would be expected to be around or higher than zero.

While controlling for preexisting differences, the results of the ANCOVA indicated a significant RR effect. As expected, the mean spring TRC levels for students in the RR samples were comparable with or higher than spring TRC levels for students in the random sample across the three cohorts (See Appendix E, Table E6). The results show small to high effect sizes for association between participation in RR and the spring TRC levels for cohorts 1 (0.32), 2 (0.76), and 3 (0.68). That is, the average student in the RR sample would have a higher score than 63% of cohort 1 students, 79% of cohort 2 students, and 73% of cohort 3 students in the random sample (Springer, 2006; Becker, 2007).

Short-term impact summary

Across cohorts, the frequency distribution of spring Grade 1 TRC levels for students in RR and the random sample was comparable, indicating that RR students were no longer clustered at the low end of the spring Grade 1 TRC distribution. In addition, the majority of students in the RR sample met or exceeded the Grade 1 end-of-year benchmark at rates that were comparable with or significantly higher than rates for the random sample students.

As expected across the three cohorts: the means for the RR students were either comparable or exceeded the TRC levels for their randomly sampled peers. The size of the effect of the intervention varied by cohort but was in favor of RR students in all three cohorts. RR was effective in raising the Grade 1 spring TRC levels and passing rates on the Grade 1 end-of-year benchmark of struggling Grade 1 students to the level of their grade-level peers or better.

In addition, RR students as a group showed less variation in their performance on the spring Grade 1 TRC levels and rates of meeting the Grade 1 end-of-year benchmark than their randomly sampled peers. Also, for every cohort, RR students who also received ESOL services outperformed their ESOL counterparts in the randomly selected sample.

Long-term Impact

Long-term impact of the RR program for each cohort was investigated relative to reading performance in Grades 2 and 3. The stratified random sample of grade-level peers was used as the comparison group in the analyses of long-term impact for evaluation question 1.

Grade 2 end-of-year reading benchmark. The expectation is that RR students will continue to perform as well as their grade-level peers as they move on to higher grades. While comparing groups within each cohort, the trends observed in Grade 1 were reversed for the Grade 2 MCPSAP-PR spring benchmark. The performance of Grade 2 students varied by cohort and was mostly in favor of randomly sampled students. As graphically displayed in Figure 2, 35–40% of RR students met the Grade 2 spring end-of-year reading benchmark compared with 50–55% of students in the random sample. There were significant differences in the passing rates on the Grade 2 end-of-year benchmark between the random and RR students in each cohorts, in favor of random students ($p < 0.05$).

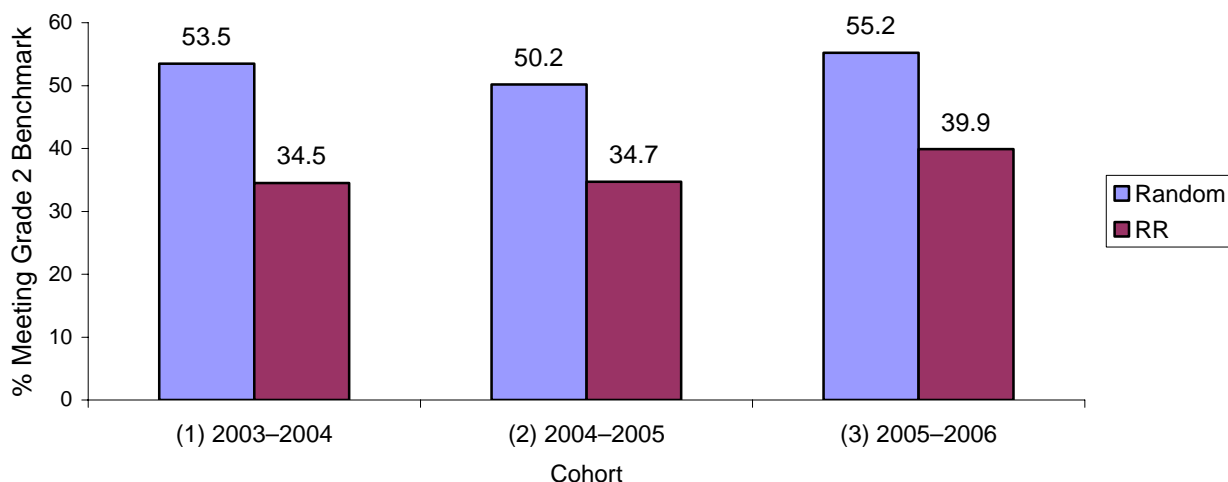


Figure 2. Proportion of students achieving Grade 2 end-of-year benchmark.

After adjusting for preexisting differences, the small negative effect sizes for the association between participation in RR and meeting the Grade 2 end-of-year reading benchmark were found for students in cohorts 1, 2, and 3 (-0.30, -0.37, -0.47) (See Appendix E, Table E1b). As such, the average student in the random sample would have higher chances of passing the Grade 2 end-of-year benchmark than more than 62% of the students in the RR sample (Springer, 2006; Becker, 2007).

Grade 2 end-of-year reading benchmark by subgroup. Table 4 presents the proportion of students meeting Grade 2 end-of-year benchmark by subgroup.

Table 4
Proportion of Students Achieving Grade 2 End-of-Year Benchmark by Subgroup

Subgroup	Grade 2 Benchmark					
	Cohort 1 (2003–2004)		Cohort 2 (2004–2005)		Cohort 3 (2005–2006)	
	% Random	% RR	% Random	% RR	% Random	% RR
Race/Ethnicity						
American Indian	--	--	--	--	--	--
African American	61.3	28.2	57.3	39.5	57.4	45.6
Asian American	60.0	58.3	--	--	76.0	55.6
Hispanic	42.9	30.9	39.0	32.3	49.7	33.6
White	66.7	36.2	76.2	41.4	54.5	52.9
Gender						
Female	53.0	36.5	53.5	33.3	59.3	41.7
Male	53.9	30.1	46.9	35.6	52.3	38.4
Special Service						
Non-FARMS	64.2	40.2	66.0	35.0	69.8	55.9
FARMS	48.5	29.3	40.1	34.5	48.3	33.1
Not ESOL	65.2	32.6	60.3	38.6	65.7	46.4
ESOL	39.4	33.3	35.8	29.5	40.8	33.6
Not Special Ed.	56.3	33.6	54.4	36.7	58.2	41.0
Special Ed.	22.2	27.6	19.4	14.3	0.0	0.0

Note. RR= Reading Recovery. Results for American Indian and subgroups with fewer than five students are not reported.

Overall, students in the random sample outperformed their RR counterparts in each subgroup, except for cohort 1 special education students. In addition, the performance of RR students in Grade 2 was lowest for cohort 2 Asian American students; Hispanic students; and students receiving special education, FARMS, and ESOL services across cohorts. For these groups, the passing rates on the Grade 2 end-of-year benchmark for RR students, which varied by cohort, were mostly less than 40%.

Grade 2 spring TRC levels. The students in the RR sample started Grade 2 with fall TRC levels that were similar to or higher than those of their grade-level peers (See Appendix F, Table F1). Similar to trends observed for Grade 1, graphical displays of the distribution of Grade 2 TRC-level data illustrate that, by spring of Grade 2, the distribution of Grade 2 spring TRC levels for RR students and random sample peers were similar (See Appendix F, Figures F1–F3). As such, the spring Grade 2 TRC levels for the RR students are not clustered at the lower end of the distribution.

While controlling for preexisting differences, ANCOVA analyses found significant differences in spring Grade 2 TRC levels between the two groups for students in cohorts 2 and 3 in favor of students in the random sample (See Appendix F, Table F2). The negligible effect size for cohort 1 (0.09) confirms that any differences in the TRC levels between the two groups are not meaningful in an educational setting. The negative small effect sizes for association between participation in RR and Grade 2 spring TRC (-0.42, -0.26) indicate that the average random sample students in cohorts 2 and 3 would have a higher spring Grade 2 TRC level than the majority of students in the RR sample.

Grade 2 TN/2 NCE scores. The expectation is that the majority of the RR students would be at or above the 50th NCE on TN/2 and that there would be no significant differences in performance

between RR students and students in the random sample across cohorts. Contrary to the expectation, about 20%–29% of the RR students across cohorts were at or above the 50th NCE on TN/2 assessments compared with 45%–53% of their random sample grade-level peers (Figure 3). Follow-up analyses were conducted to evaluate the differences in proportions. Within each cohort, significant differences in proportions of students at or above the 50th NCE were found in favor of the random sample students ($p < 0.05$). Across cohorts, the negative small to large effect sizes (-0.40, -1.07, -0.59) confirmed that the differences in the performance between the two groups was large enough to be meaningful in an educational setting (See Appendix E, Table E1c). That is, across cohorts, the average student in the random sample would have a higher NCE score than the majority of RR students (66%, 84%, and 58%, respectively).

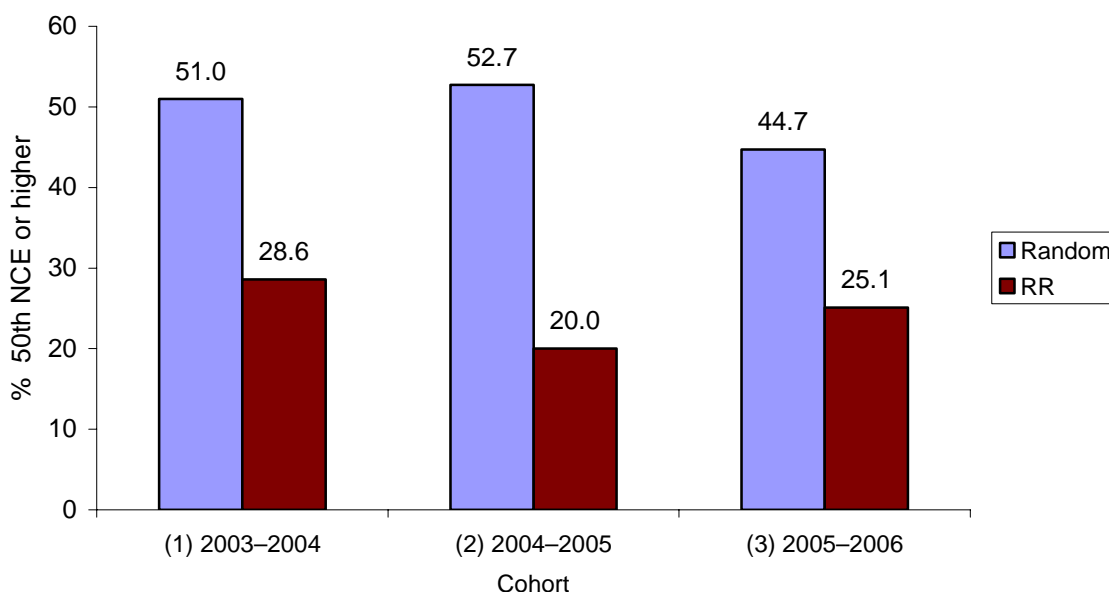


Figure 3. Proportion of random sample and Reading Recovery students at or above the TerraNova Second Edition 50th National Curve Equivalent.

Grade 2 TN/2 NCE scores by subgroup. Table 5 presents the proportion of students at or above the 50th NCE for TN/2 assessments by subgroup.

Table 5

Proportion of Students at or Above 50th National Curve Equivalent for TerraNova Second Edition Assessments by Subgroup

Subgroup	At or above 50 th TN/2 ^a NCE					
	Cohort 1(2003–2004)		Cohort 2(2004–2005)		Cohort 3(2005–2006)	
	% Random	% RR	% Random	% RR	% Random	% RR
Race/Ethnicity						
African American	50.5	20.5	52.0	18.4	49.1	26.5
Asian American	55.6	45.5	--	--	51.9	33.3
Hispanic	44.7	27.0	45.9	16.1	39.3	19.0
White	71.8	41.5	78.6	32.1	58.3	58.8
Gender						
Female	53.1	28.6	49.2	25.9	42.4	26.2
Male	49.4	28.6	56.2	16.3	46.3	24.2
Special Services						
Non- FARMS	69.9	39.1	68.0	20.3	59.4	36.8
FARMS	42.7	23.0	42.9	19.8	37.2	20.1
Not ESOL	59.9	32.7	59.5	18.5	54.9	27.0
ESOL	39.7	23.3	42.7	22.0	29.5	23.3
Not Special Ed.	52.6	28.6	55.6	18.1	46.1	24.9
Special Ed.	26.3	28.6	32.3	0.0	13.3	33.3

Note. RR=Reading Recovery. -- indicates that results for subgroups with less than five students are not reported.

^a A different edition of the TN/2 assessment was administered each year (from 2004 to 2007).

Depending on the cohort, the differences in performance between the RR and randomly sampled peers across the three cohorts were as follows: Asian Americans (10–25 percentage points); African Americans (22–33 percentage points); White students (30–47 percentage points); Hispanic students (17–29 percentage points). Groups with the lowest performance on the TN/2 assessment were the students receiving special services and Hispanic students (Table 5).

TN/2 scale scores. While the expectation is that the performance on the TN/2 scale scores would not differ between the two groups, the mean scale scores adjusted for preexisting differences indicated the TN/2 scale scores for the random samples students were significantly higher than scores for students in the RR sample (See Appendix F, Table F2). The medium negative effect sizes (0.46, -0.44, -0.37) for students in cohorts 1, 2, and 3 also confirm that the performance of RR students lagged behind that of their random sample grade-level peers. That is, for each cohort, the average student in the random sample would have a higher TN/2 scale score than the majority of students in the RR sample.

Grade 3 MSA proficiency levels. The expectation is that RR students will continue to perform as well as their grade-level peers in Grade 3. Figure 4 displays the proportions of students who were at or above proficiency on the 2006 and 2007 MSA.⁶

⁶ No official student-level MSA data are available as of this writing for cohort 3, who were in Grade 2 during the 2006–2007 school year.

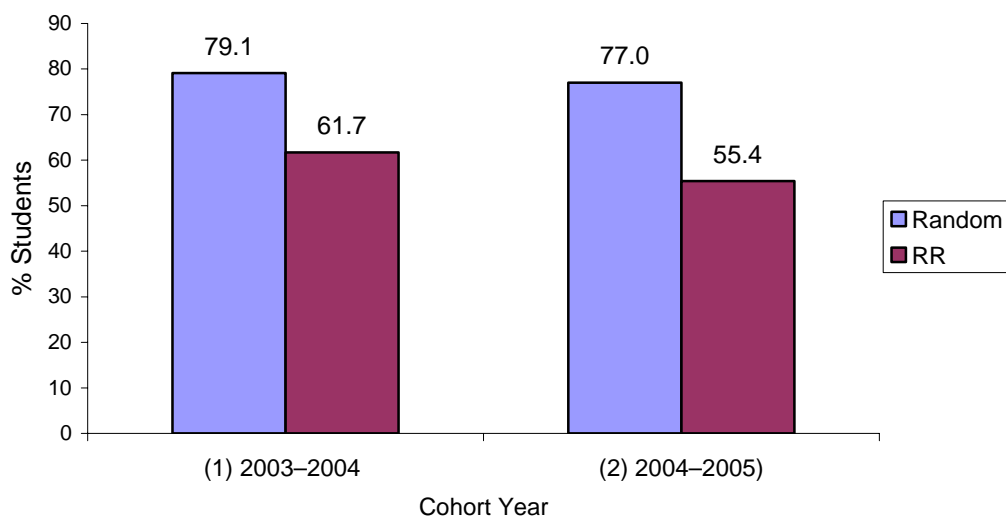


Figure 4. Proportion of students performing at or above proficiency on Grade 3 reading Maryland School Assessment.

Comparing among groups within cohorts, more than one half of RR students in cohort 1 (62% vs. 79%) and cohort 2 (55% vs. 77%) were at or above proficiency on the Grade 3 MSA reading subtest compared more than 70% of students in the stratified random samples. Follow-up analyses were conducted to evaluate the differences among these proportions. The results revealed that the proficiency rates for random students in the two cohorts were significantly higher than rates for the RR students ($p < 0.05$). Also, the small but negative effect sizes (-0.29 and -0.28) confirm that the differences in passing rates, which were in favor of students in the random samples, were noteworthy and meaningful in an educational setting (See Appendix E, Table E1d).

MSA proficiency levels by subgroup. Overall, higher proportions of White students were at or above proficiency on the MSA compared with African American and Hispanic students in their respective sample across groups within cohorts. Comparing groups within cohorts, about 62% of RR White students from cohort 1 were at or above proficient on the MSA compared with 50% and 55% of the African American and Hispanic students' peers, respectively. Contrary to historical trends in MCPS, cohort 2 Asian American students who received RR performed the same as the African American students in the group (Figure 5).

Overall, the random sample students outperformed their RR counterparts on the MSA in every subgroup, with the exception of cohort 1 ESOL students. Cohort 1 RR students who received ESOL services outperformed their non-ESOL counterparts (Figure 6). Also, RR students as a group showed less variation among their subgroups on their performance on the MSA than their randomly sampled peers. Notably, for both cohorts, the difference between the proficiency rates for FARMS and non-FARMS students in the random sample was higher than the differences in proficiency rates for FARMS and non-FARMS among RR students.

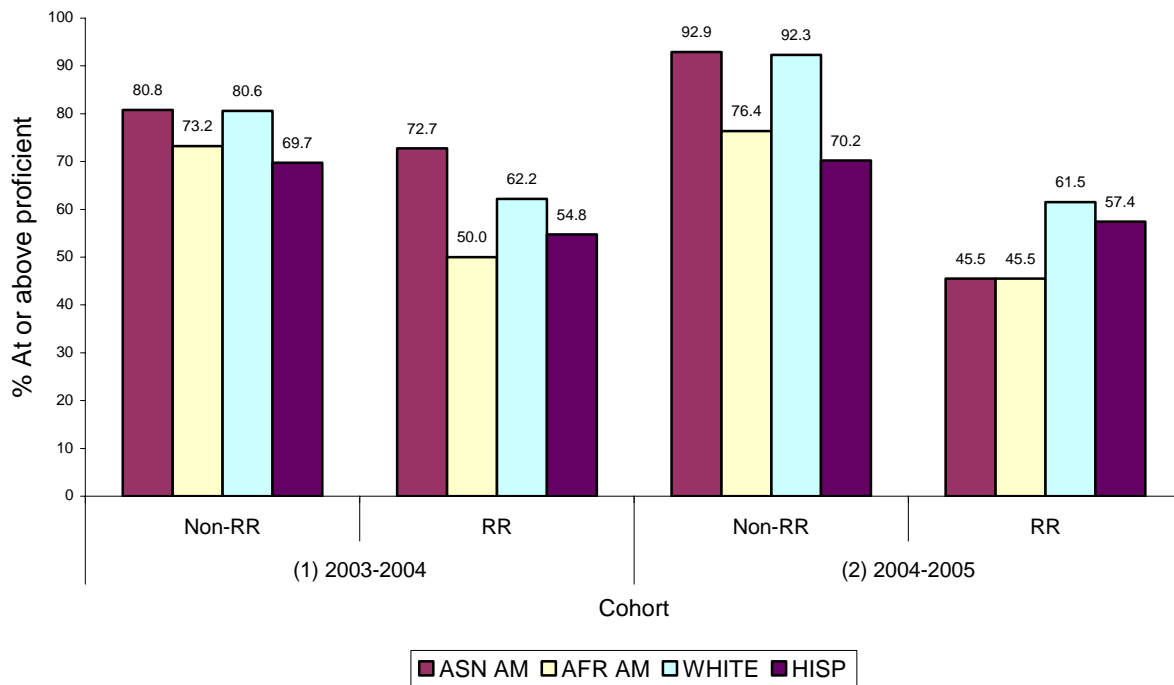


Figure 5. Proportion of students performing at or above proficiency on reading Maryland School Assessment by race/ethnicity.

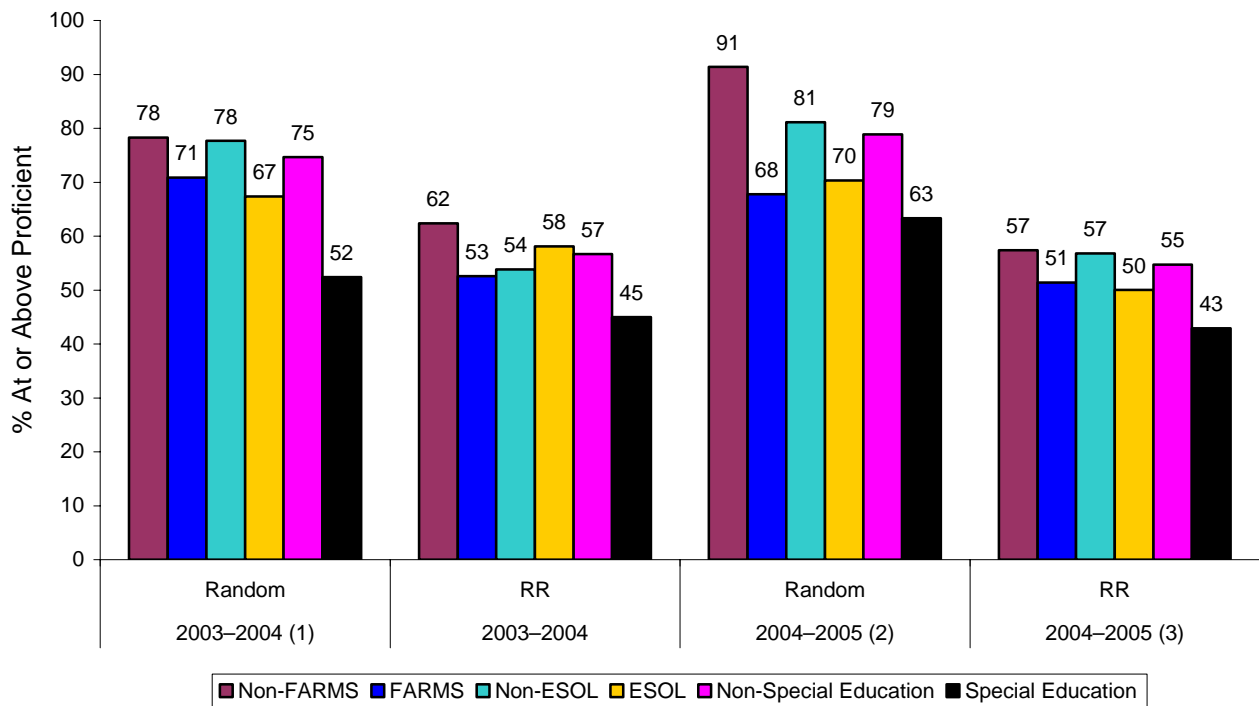


Figure 6. Proportion of students performing at or above proficiency on Grade 3 reading Maryland School Assessment by service receipt.

Grade 3 MSA scale scores. Means and standard deviations for MSA Reading scale scores are presented in Appendix F, Table F3. There were significant differences in the adjusted mean MSA scores between the random and RR students in cohorts 1 and 2 in favor of random students. The small effect sizes for comparison with random students for cohorts 1 and 2 (-0.25, -0.29) indicate that the average students in the random sample had a higher MSA score than the majority (62%–66%) of the students in the RR sample.

Grade 3 MAP-R RIT scores. Means and standard deviations for MAP-R RIT scores also are presented in Appendix F, Table F3. The ANCOVA for MAP-R RIT revealed there were significant differences in the adjusted mean RIT scores between the random and RR students in cohorts 1 and 2, in favor of students in the random sample (See Appendix F, Table F4). Also, the medium negative effect sizes (-0.34, -0.50) confirm that the performance of RR sample students on the MAP-R lagged behind the performance of their random sample peers across the two cohorts.

Long-term impact summary

Compared with their random sample peers, the majority of the students in the RR sample did not meet the Grade 2 end-of-year benchmark or place at or above the 50th NCE on the TN/2 assessment. For cohorts 2 and 3 students, statistically significant differences were found in favor of random sample students on the spring Grade 2 TRC levels and the passing rates on the Grade 2 end-of-year benchmark. Small negative effect sizes confirmed that these differences are meaningful in an educational setting. Across each cohort, statistically significant differences were found in favor of random sample students on the TN/2 scale scores and rates of being at or above the 50th NCE for TN/2 assessments. The medium effect sizes for TN/2 scores confirm that the performance of RR students lagged behind the performance of their random sample peers.

Also, for each subgroup, the students in the random sample outperformed their counterparts in the RR samples on the spring Grade 2 TRC levels, passing rates on the Grade 2 end-of-year benchmark, and TN/2 assessments. Further, compared with their random sample peers, there was less variation on the performance among students of various subgroups in the RR sample than among their counterparts in the random sample on the Grade 2 reading measures.

By the end of Grade 3, more than one half of the students in the RR sample were at or above proficiency on the MSA reading subtest. However, the students from the random sample attained proficiency at significantly higher rates than RR students. In addition, random sample students outperformed their counterparts in the RR sample in every subgroup.

There were significant differences in the adjusted mean MSA scores between the random and RR students in cohorts 1 and 2 in favor of random sample students. The small negative effect sizes for association between MSA and participation in Reading Recovery for students in both cohorts signify that the performance of students in the RR sample lagged behind the performance of their random sample peers.

Further, by spring of Grade 3, significant differences were found on the MAP-R RIT scores between random sample and RR students in favor of random sample students. At the same time, the small negative effect sizes indicate that the average student in the random sample would have higher RIT scores than the majority of students in the RR sample.

Evaluation Question 2. How do students who successfully complete reading recovery perform on local and state assessments compared with grade-level peers who were similar but not served in reading recovery?

Evaluation question 2 assesses the relative benefit of having RR by comparing the group that was eligible but did not receive RR (non-RR) with RR students. Three different analytical methods were applied to evaluation question 2: descriptive analyses of student performance levels, logistic regression models for categorical outcome reading measures, and ANCOVA of scale score measures.

Short-term Impact

Grade 1 end-of-year benchmark. The expectation is that a) after participating in RR, the majority of RR students would meet the Grade 1 end-of-year benchmark and b) the performance of RR students would exceed that of their non-RR peers. The findings from this study reveal that the majority of RR students met the Grade 1 benchmark (Figure 7). In addition, the proportions of cohorts 2 and 3 RR students meeting the end-of-year Grade 1 benchmark far exceeded those of the non-RR students and were significantly higher than rates for students in the non-RR sample ($p < 0.05$).

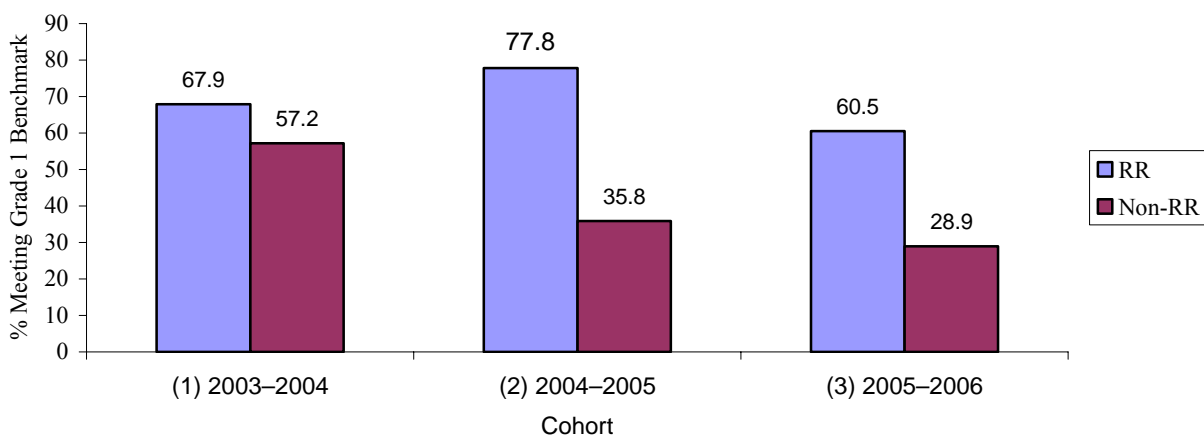


Figure 7. Proportion of Reading Recovery and non-Reading Recovery students meeting the Grade 1 end-of-year benchmark.

Effect sizes for association between participation in RR and meeting the benchmark are summarized in Table 6. The effect sizes varied by cohort: cohort 1 (0.55), cohort 2 (1.11), and cohort 3 (0.73), respectively. Specifically, across cohorts, RR students would outperform over the majority (69%, 86%, and 76%) of their non-RR peers (Springer, 2006; Becker 2000).

Table 6

Measures of Effect Sizes for Association Between Meeting Grade 1 Benchmark and Participation in Reading Recovery by Cohort

Achievement Benchmark	Cohort	Odds ratio	Effect Size (Logit <i>d</i>)
Grade 1 benchmark	Cohort 1 (2003–2004)	2.711	0.55
	Cohort 2 (2004–2005)	7.524	1.11
	Cohort 3 (2005–2006)	3.763	0.73

Grade 1 spring TRC levels. Summary statistics on Grade 1 reading measures are presented in Appendix G, Table G1. Figures G1a–G1c in Appendix H also present the frequency distribution of the spring Grade 1 distribution for RR and non-RR students. In general, descriptive information reveals that RR students outperformed their non-RR peers on measures of reading in Grade 1.

While controlling for initial reading ability and demographic characteristics, the ANCOVA results showed there were significant differences in the spring Grade 1 TRC levels between the non-RR and RR students in all three cohorts. Table 7 displays the adjusted mean differences from the ANCOVA analyses and effect sizes for the spring TRC level. The effect sizes for association between participating in RR and Grade 1 spring TRC level are large for cohorts 1, 2, and 3 students, indicating that RR students made greater gains in reading in Grade 1 compared with non-RR peers. For example, effects sizes of 0.99 indicate that the average RR student would have a higher TRC level than 84% of students in non-RR sample. The effect sizes for cohort 3 Grade 1 spring TRC levels are above 0.73, indicating that the TRC level of the average RR students would be higher than for 82% of students in the non-RR sample.

Table 7

Adjusted Means and Mean Differences for Grade 1 and Measures of Effect Sizes for Association Between Grade 1 Spring Text Reading Comprehension Levels and Participation in Reading Recovery by Cohort

Grade 1 Spring TRC	RR Students		Non-RR students		Mean Difference	<i>F</i> ^a	<i>p</i> ^b	Effect Size
	Mean	<i>N</i>	Mean	<i>N</i>				
Cohort 1 (2003–2004)	13.6	182	10.4	408	3.22	61.8	0.00	0.70
Cohort 2 (2004–2005)	15.7	124	11.8	261	3.90	82.34	0.00	0.99
Cohort 3 (2005–2006)	12.4	160	8.9	253	3.36	84.12	0.00	0.92

Note. RR=Reading Recovery

^aThe *F*-test is used to determine if the scores for the two groups are different at the 95% confidence level.

^b *p* value indicates the results of tests of significance.

Long-term Impact

Grade 2 end-of-year benchmark. Figure 8 presents information on performance rates for RR and non-RR students on the Grade 2 end-of-year benchmark. Across the cohorts, less than 50% of students from either the RR or the non-RR groups met or exceeded the benchmark. Also, the unadjusted percentages show that the non-RR students from cohorts 2 and 3 passed at higher rates than their RR counterparts, whereas, for cohort 1 students, RR students outperformed their non-RR counterparts (Table 8).

After adjusting for preexisting differences, the small negative effect sizes for association of participation in RR and meeting Grade 2 end-of-year reading benchmark (-0.25, -0.36) indicate that the differences in performance levels of the two groups of students for cohorts 2 and 3 are noteworthy (Table 8).

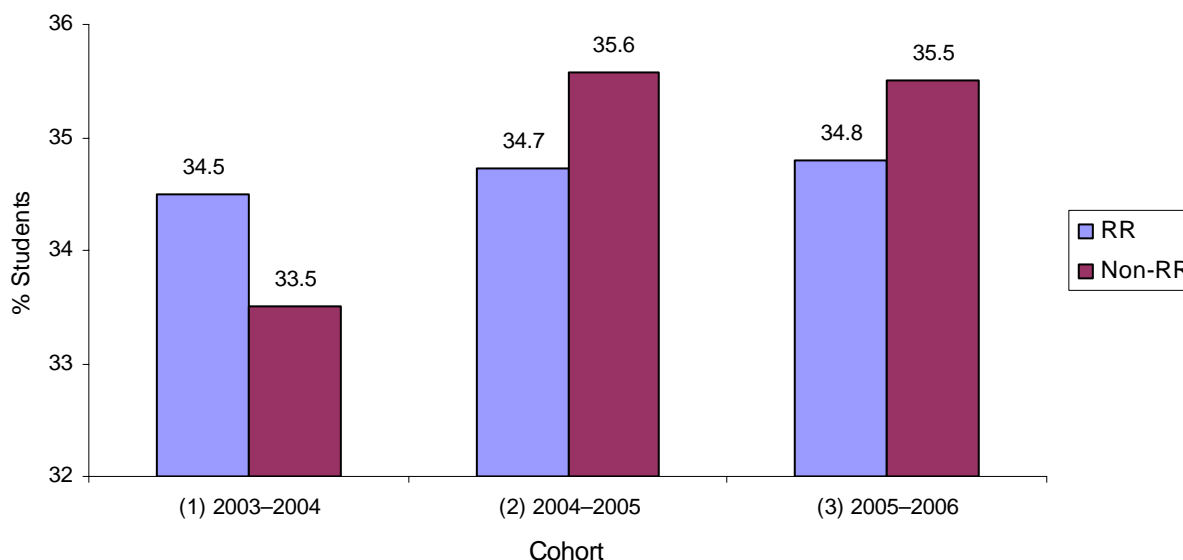


Figure 8. Proportion of Reading Recovery and non-Reading Recovery students meeting the Grade 2 end-of-year benchmark.

Table 8

Measures of Effect Sizes for Association Between Grade 2 End-of-year Benchmark Status and Participation in Ready Recovery by Cohort

Achievement Benchmark	Cohort	Odds ratio	Effect Size (Logit d)
Grade 2 benchmark	Cohort 1 (2003-2004)	1.031	0.55
	Cohort 2 (2004-2005)	0.632	-0.25
	Cohort 3 (2005-2006)	0.525	-0.36

Grade 2 spring TRC levels. On average, descriptive analyses reveal that RR students had higher Grade 2 spring TRC levels than non-RR students (See Appendix G, Table G1). Figures G2a-G2c in Appendix G also present the frequency distribution of the spring Grade 2 distribution by sample and cohort.

After controlling for initial reading ability and demographics, the ANCOVA results show that there were significant differences in the spring Grade 2 TRC levels between the non-RR and RR students in cohort 3 in favor of non-RR students, but no significant differences for students in cohorts 1 and 2 because of sample size. In addition, the effect sizes are negligible for cohort 1 (0.01), indicating no practical difference in the spring Grade 2 TRC levels exist between RR and non-RR students. The small negative effect sizes for cohorts 2 and 3 (-0.26, -0.20) signify that

the Grade 2 spring TRC level for the average student in the non-RR sample would be higher than the level for 58% of the students in the RR sample (Table 9).

Table 9
Adjusted Means and Mean Differences for Grade 2 Text Reading Comprehension Levels and Measures of Effect Sizes for Association Between Reading Achievement and Participation in Reading Recovery by Cohort

Spring Grade 2 TRC	RR Students		Non-RR students		Mean Difference	F ^a	p ^b	Effect Size
	Mean	N	Mean	N				
Cohort 1 (2003–2004)	22.5	119	21.1	118	1.39	1.09	0.30	0.01
Cohort 2 (2004–2005)	19.3	111	20.8	212	-1.54	4.74	0.03	-0.26
Cohort 3 (2005–2006)	19.0	159	20.1	249	-1.21	3.98	0.04	-0.20

Note: RR=Reading Recovery.

^a The F-test is used to determine if the scores for the two groups are different at the 95% confidence level.

^b p value indicates the results of tests of significance (Bordens & Abbott, 1999).

TN/2 NCE scores. Descriptive analyses reveal that, across the cohorts, about one third or less of the students from the RR and non-RR samples were at or above the 50th NCE on the TN/2 assessment (Figure 9). In general, with the exception of cohort 1 students, significantly higher proportions of non-RR students were at or above the 50th NCE than RR students ($p < 0.05$). As indicated by the effect sizes shown on Table 10, the adjusted differences in proportions of students at or above the 50th NCE are meaningful in an educational setting.

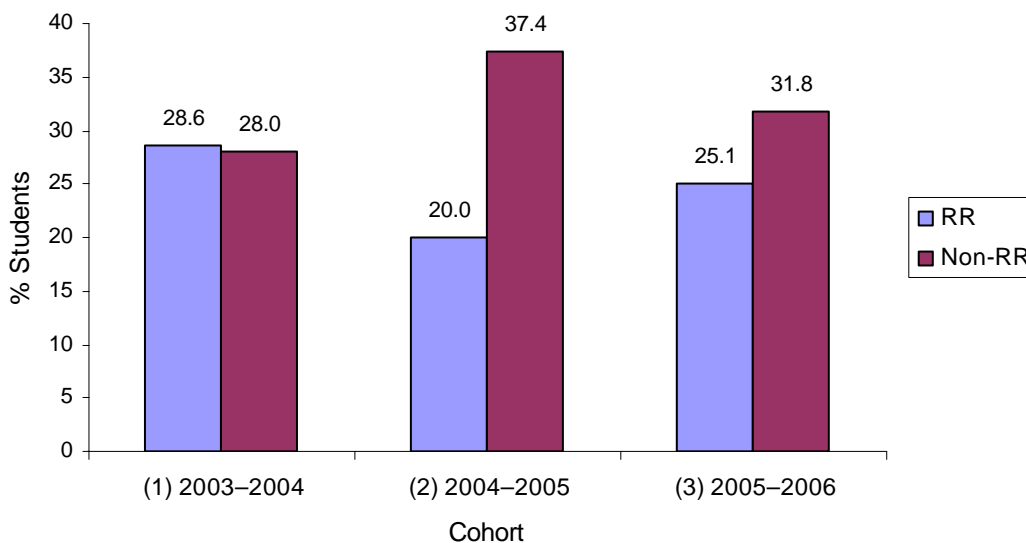


Figure 9. Proportion of Reading Recovery and non-Reading Recovery students at or above the TerraNova Second Edition 50th National Curve Equivalent.

Table 10

Measures of Effect Sizes for Association Between Scoring at or Above the TerraNova Second Edition 50th National Curve Equivalent and Participation in Reading Recovery by Cohort

TN/2	Cohort	Odds ratio	Logit <i>d</i> (Effect Size)
At or above the 50 th NCE	Cohort 1 (2003–2004)	0.785	-0.13
	Cohort 2 (2004–2005)	0.191	-0.91
	Cohort 3 (2005–2006)	0.329	-0.61

TN/2 scale scores. Summary statistics for TN/2 scale scores are presented in Appendix G, Table G3. After controlling for preexisting conditions, significant differences in adjusted TN/2 scores were found between the two groups for students in each cohort in favor of non-RR students. In addition, the effect size for cohorts 1 and 2 are small to medium, confirming that the differences in performance of the two groups are of practical significance (Table 11).

Table 11

Adjusted Means and Mean Differences for TerraNova Second Edition Scale Scores and Measures of Effect Sizes for Association Between Reading Achievement and Participation in Reading Recovery

TN/2 Scale Score	RR Students		Non-RR students		Mean Difference	<i>F</i> ^a	<i>p</i> ^b	Effect Size
	Mean	<i>N</i>	Mean	<i>N</i>				
Cohort 1 (2003–2004)	584.2	161	593.8	343	-9.5	7.7	0.00	-0.27
Cohort 2 (2004–2005)	569.3	102	596.1	218	-26.8	20.2	0.00	-0.54
Cohort 3 (2005–2006)	594.2	160	601.3	238	-7.1	4.4	0.04	-0.16

Note. RR=Reading Recovery.

^a The *F*-test is used to determine if the scores for the two groups are different at the 95% confidence level.

^b *p* value indicates the results of tests of significance (Bordens & Abbott, 1999).

Grade 3 MSA performance levels. Figure 10 presents information on the observed proportion of students at or above proficiency on the MSA reading subtest. Comparing between groups within each cohort, no significant differences were found in the proficiency rates between the two groups for cohort 1 students ($p > 0.05$).

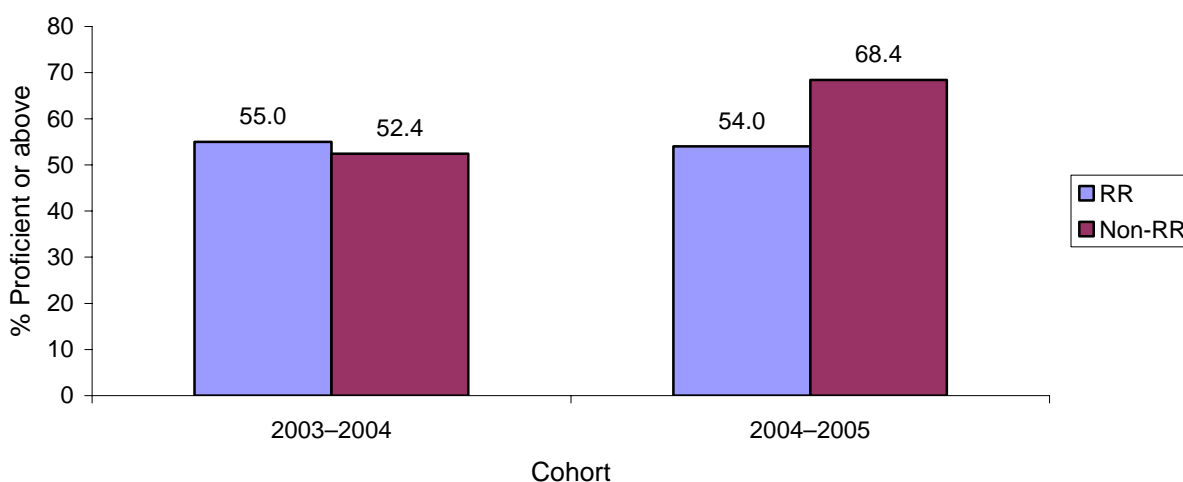


Figure 10. Percentage proficient and above on the Grade 3 reading Maryland School Assessment

After controlling for preexisting differences, the negligible effect sizes for cohort 1 (-0.01) confirm that the differences in MSA proficiency rates between the two groups are not practically meaningful in an educational setting. However, for cohort 2 students, a significant difference in proficiency rates was found in favor of the students in the non-RR sample ($p < 0.05$). On the other hand, the small negative effect sizes for association between participation in RR and proficiency on the MSA for cohort 2 students (-0.29) are noteworthy (Table 12).

Table 12

Measures of Effect Sizes for Association Between Proficiency on the Maryland School Assessment and Participation in Reading Recovery by Cohort

Grade 3 MSA	Cohort	<i>p</i>	Ords ratio	Logit d Effect Size
Proficient or above	Cohort 1 (2003–2004)	0.18	0.977	-0.01
	Cohort 2 (2004–2005)	0.00	0.593	-0.29

Note. Cohort 3 was in Grade 2 during the 2006–2007 school year.

Summary statistics for MSA scale scores are presented in Appendix G, Table G1. Comparing the performance between groups within cohorts, the ANCOVA results showed there were no significant differences in MSA reading scores between the RR and non-RR students in cohort 1 (Table 13). The effect size for association between participation in RR and MSA reading scores for cohort 1 students was negligible (-0.01). On the other hand, the non-RR students in cohort 2 had significantly higher MSA scores than their RR peers (Table 13). Further, the small negative effect sizes for association between participation in RR and MSA scale scores for cohort 2 students (-0.29) indicate that the performance of RR students lagged behind that of their non-RR peers.

MAP-R RIT scores. Summary statistics for MAP-R RIT scores are presented in Appendix G, Table G1. The results from ANCOVA showed no significant differences in RIT scores between the two groups for cohort 1 students (Table 13). These findings also are confirmed by the negligible effect sizes (-0.05). On the other hand, significant differences in MAP-R scores in

favor of non-RR students were found for students in cohort 2. The small effect size indicates that these differences are meaningful in an educational setting.

Table 13

Adjusted Means and Mean Differences for Grade 3 Reading Measures and Measures of Effect Sizes for Association Between Reading Achievement and Participation in Reading Recovery by Cohort

Outcome measure	RR Students		Non-RR students		Mean Difference	F^a	p^b	Effect Size
	Mean	N	Mean	N				
Cohort 1 (2003–2004)								
2006 MSA Scale Score	392.5	144	392.6	294	-0.09	0.01	0.97	-0.01
2006 MAP-R RIT Score	189.4	139	190.1	281	-0.72	0.28	0.50	-0.05
Cohort 2 (2004–2005)								
2007 MSA Scale Score	383.2	106	392.3	220	-9.13	6.21	0.01	-0.29
2007 MAP-R RIT Score	184.0	95	187.5	215	-3.46	4.08	0.04	-0.25

Note. RR=Reading Recovery. No MSA or MAP-R RIT scores are reported for Cohort 3 students who were in Grade 2 in 2006–2007.

^a The F -test is used to determine if the scores for the two groups are different at the 95% confidence level.

^b p value indicates the results of tests of significance.

Long-term impact summary

Overall, less than one half of the students in the RR and non-RR groups attained the Grade 2 end-of-year reading and cohorts. In addition, the mean differences in Grade 2 spring TRC scores between RR and non-RR students varied by cohort. Specifically, there were significant differences in the spring Grade 2 TRC levels between the non-RR and RR students in cohorts 2 and 3 in favor of non-RR students, but no significant differences for students in cohort 1.

Again, for the two groups, only 20%–38% of students scored at or above the 50th NCE on the TN/2 across cohorts. Subsequently, significant differences in the adjusted TN/2 scores between the non-RR and RR students in the three cohorts were found in favor of non-RR students. The associated small to medium negative effect sizes for association between participation in RR and TN/2 scores confirm that the performance of RR students lagged behind that of their non-RR peers.

By Grade 3, more than one half of the students from RR and non-RR students were at or above proficiency on the Grade 3 MSA. However, cohort 2 non-RR students attained MSA reading proficiency at significantly higher rates than their RR peers.

The performance of RR and non-RR students on the MSA reading and MAP-R RIT scores varied by cohort. Contrary to expectation, there was no significant difference in the adjusted MSA mean scores and MAP-R mean scores for non-RR students and RR students in cohort 1. Also, the negligible effect sizes indicated that the performance of the two groups was comparable.

For cohort 2 students, significant differences in the adjusted mean MSA and MAP-R scores were found in favor of the non-RR students. Also, small negative effect sizes of association between participation in RR and MSA and MAP-R RIT scores confirm that the performance of the RR

students lagged behind that of their non-RR peers on the MSA reading scores and MAP-R RIT scores.

Discussion

This study found a significant impact of RR on the short-term performance of students who had successfully completed the RR intervention. As such, these findings confirm that RR accomplished its primary goal of bringing the lowest achievers up to the average classroom reading level by the end of Grade 1. As expected, by the end of Grade 1, the spring TRC levels and the passing rates on the Grade 1 end-of-year benchmark for RR students were comparable with or significantly higher than the performance level for their random sample peers. The effect sizes for association between participation in RR and Grade 1 reading measures were close to or greater than zero. This finding demonstrates that, across the three cohorts, the RR students performed at the level of their peers or outperformed their peers in the random sample. At the same time, the passing rates for RR students far exceeded the rates for similar students—students who were also performing below grade level by the fall of Grade 1 who did not participate in RR. Further, the moderate-to-large effect sizes confirm the positive and educationally meaningful short-term effect of the RR intervention.

The pattern observed in the performance of RR students on Grade 1 end-of-year benchmark was evident for all the student subgroups. Compared with students in the random sample, there was minimal variation in the TRC levels and proportions of students meeting the grade-level benchmark among students from the racial and service receipt subgroups who received RR intervention.

The findings do not suggest an ongoing impact of RR on the reading achievement of the students in Grade 2. Two patterns are evident for RR students in Grade 2. First, the majority of RR students did not attain the MCPS Grade 2 end-of-year reading benchmarks. Second, the RR students were performing at levels lower than the random sample peers and had performance rates comparable with similar peers who did not participate in RR. Specifically, contrary to the expectation that the RR students would be performing at grade level and at rates similar to their random sample peers, the overall passing rates of RR students on the Grade 2 end-of-year reading benchmark were lower than 50%. For students in two of the cohorts, statistically significant differences were found in favor of random sample students on the spring Grade 2 TRC levels and the Grade 2 end-of-year benchmark. The corresponding small negative effect sizes confirmed that these differences are meaningful in an educational setting.

Fewer than 40% of RR students scored at or above the 50th NCE for TN/2. For two of the cohorts, statistically significant differences were found in favor of random sample students on the TN/2 scale scores. Also, the small negative effect sizes for TN/2 scores demonstrate that the performance of RR students lagged behind the performance of their random sample peers.

Similar to findings for comparison with random sample students, significant differences in the TN/2 scores between RR and similar students were found in favor of similar students in the three cohorts. The associated small-to-medium negative effect sizes for association between participation in RR and TN/2 scores verify that the performance of RR lagged behind the performance of similar students who did not participate in RR.

Indeed, the performance of RR students by spring of Grade 2 was consistent with the districtwide performance of the students in their schools and respective demographic profile (Chesney & Fink, 2007; Douglas, 2006; Zhao, 2006, 2007). These districtwide trends show that less than 50% of students receiving special services and students from the African American and Hispanic subgroups meet the Grade 2 end-of-year reading benchmark or score above the 50th NCE on the TN/2. In addition, the greatest variability among schools and within subgroups in the district occurs for the Grade 2 end-of-year reading benchmark and on TN//2.

The trends in the district suggest that the decline in performance rates of Grade 2 students from certain subgroups on the Grade 2 reading measures is related to the performance of RR students. As such, issues related to acquisition of literacy skills in Grade 2 should be more of a concern than the findings that the RR students are performing at the same level as similar students who did not participate in the intervention.

The findings for Grade 3 reading outcomes were mixed and varied by measure and cohort being examined. As expected, by Grade 3, the majority of the RR and comparison students performed at or above proficiency on the Grade 3 MSA. However, the RR students achieved proficiency at significantly lower rates than their random sample peers, whereas their performance levels were comparable with those of similar students who did not participate in RR. The small effect sizes indicated that the performance of RR students lagged behind the performance of the students in the random sample.

While results are slightly negative on the MSA reading scores for cohort 1 RR students, after controlling for initial differences, the results showed that the MSA and MAP-R scores for RR students were not significantly different from scores for similar students who did not receive RR. The negligible effect sizes also corroborate that the performance of RR students on the MSA reading scores was comparable with the performance of similar students who did not participate in RR. In addition, the small negative effect sizes signify that the performance of RR students lagged behind that of their random sample grade-level peers and similar students who did not receive RR on the MSA and MAP-R RIT scores.

Conversely, the performance level of cohort 2 RR students on the MSA and MAP-R was significantly lower than the performance levels of students in the random sample and similar students who did not receive RR. In addition, the small negative effect sizes signify that the performance of RR students lagged behind that of their random sample grade-level peers and similar students who did not receive RR on the MSA and MAP-R RIT scores.

Conclusion

The RR program has been in place in MCPS since the 1997–1998 school year. This study examined the performance students who participated in RR from the 2003–2004 through 2005–2006 school years.

The MCPS RR students accelerated their reading to and beyond the grade-level benchmark by the end of Grade 1.

The finding pertaining to short-term performance of RR students in this study is consistent with the RR program theory and literature (Briggs & Young, 2003; Forbes & Szymczuk, 2006). In general, some research findings indicate sustained effects of RR while others do not. The findings relating to the performance of RR students on Grades 2 and 3 reading measures counter those reported by others (Askew, 2003; Schmitt et al., 2005, USDE, 2007). In these studies, higher proportions of RR students were performing at higher rates than grade-level peers in reading as they progressed to Grades 3, 4, and 5. The students in this study did not show this pattern. Because this was a quasi-experimental study and nonequivalent groups were used for comparison, many intervening variables across time, such as the quality of subsequent teaching, service receipt variables, and the context within which RR is situated within a school may deem it impracticable to attribute students' achievements in literacy beyond the intervention year to their participation in RR in Grade 1 (Askew et al., 1998).

The performance of Grade 2 RR students was either on par with their peers, as would be expected, or lagged behind. The trends observed in this study are similar to findings by other researchers (Wasik & Slavin, 1993; Glynn, Crooks, Bethune, Ballard & Smith 1989; Shanahan & Barr, 1995). These studies found that the learning rate of RR students slowed when the students returned to the classroom and was much lower than for other low-achieving students. By Grade 3, in one cohort, the proficiency rates for RR students were comparable with stratified samples of random and similar peers, indicating that the students were performing at comparable levels with their peers. Notably, the RR program developers make no promises that successfully discontinued RR students will remain within the average classroom range without further intense intervention (D'Agostino & Murphy, 2004).

Indeed, researchers also have questioned whether it is reasonable to hold RR accountable for the long-term progress of RR participants (D'Agostino & Murphy, 2004). D'Agostino and Murphy argued that, while RR seemed to have a lasting effect on the TRC level at least until the end of the Grade 2, expecting all RR students to maintain average class achievement levels without the availability of additional support might be unrealistic. Besides, RR, an intervention that is designed for Grade 1 students, should not "shoulder" the burden for continued interventions over the years. Therefore, other reading interventions may be better designed to meet the needs of older students in the years following RR. An additional reason for questioning the expectations that RR should be held as a sufficient intervention for ongoing progress of RR students relates to factors outside RR's spheres of influence.

Therefore, placing RR within a coordinated school literacy approach that emphasizes high-quality literacy instruction prior to RR, during RR, and follow up after RR, is critical. RR should not be viewed in isolation from other schoolwide literacy approaches. Further, the findings from this study suggest ongoing monitoring of the progress of RR students is crucial. Ongoing monitoring would create schoolwide ownership and valuing of RR. Literature suggested that the schools that had great success with RR had systems in place for supporting students who were not maintaining their gains from RR (Mcdowall, Boyd, Hodgen & Vliet, 2005).

Attrition of the students from the RR sample and the completeness of local and district assessment data for students in Grades 1 and 2, and particularly students in the RR sample, were limitations in this study. There are several explanations for why students do not have complete data on each assessment used in this evaluation, including the following: a) students were not

assessed for a variety of reasons (they were ESOL instructional level I students), b) students were in schools implementing the Reading First program, and/or c) data entry was incomplete.

Finally, this study did not attempt to determine which aspects of the RR intervention were effective. Future evaluations should explore the unique components of the program (e.g., one-on-one intervention, daily lessons, and high levels of professional development) to determine what factor or group of factors is most effective for RR students and for student subgroups. Even without this information, the results speak to the effectiveness of RR in raising spring Grade 1 TRC levels.

Recommendations

- **Investigate factors that contribute to low reading proficiency levels of RR students and students receiving special services on the Grade 2 end-of-year benchmarks and TN/2 assessments.** Reasons for RR students falling behind in Grade 2 are unclear, particularly since the findings for Grade 2 counter findings reported by others (United States Department of Education [USDE, 2007]). In general, districtwide trends show that less than 50% of students receiving special services, and students from the African American and Hispanic subgroups, meet the Grade 2 end-of-year reading benchmark or score above the 50th NCE on the TN/2. Further, the greatest variability among schools and within subgroups in the district occurs for Grade 2 end-of-year reading benchmark and on TN//2 (Chesney & Fink, 2007; Douglas, 2006; Zhao & Von Secker, 2007). Therefore, understanding and addressing issues related to reading performance in Grade 2 is critical for closing the achievement gap among student subgroups and schools.
- **Establish targeted training of Grade 2 teachers and reading specialists to support former RR students and students receiving special services.**
- **Consider the development of an appropriate Grade 2 reading intervention program for former RR students and other low-performing students.** In particular, appropriate interventions matched with specific needs of students for schools with less than 75% of students achieving the spring Grade 1 and 2 end-of-year reading benchmark are needed.
- **Add a marker to OASIS to identify students who received RR each year.** This will enable schools and researchers to track and monitor performance of RR students over time.
- **Monitor progress of RR students in Grade 2 and provide additional support as needed.** Eliciting classroom teachers' observations of RR students before RR, changes in students after RR, and areas of further needs for RR students after the intervention will help teachers target their support to the specific needs of the students at each stage.

References

- Adkins, D. (2007). Northwest Evaluation Association: *A study of the alignment of NWEA RIT scale score with Maryland Assessment System*. Lake Oswego, OR: North West Evaluation Association.
- Ashdown, J., & Simic, O. (2000). Is early literacy intervention effective for English language learners? Evidence from Reading Recovery. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, 5(1), 27–42.
- Askew, B. J., Kaye, E., Frasier, D. E., Mobasher, M., Anderson, N., & Rodriguez, Y. G. (2002). Making a case for prevention in education. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, 6(2), 43–73.
- Askew, B. J., Fountas, I.C., Pinnell, G.S., & Schmitt, M.C. (1998). *Reading Recovery review: Understandings, outcomes, and implications*. Columbus, OH: Reading Recovery Council of North America.
- Becker, L. (2000). *Effect size measures for two independent groups*. Retrieved on 15 July, 2007, from <http://web.uccs.edu/lbecker/Psy590/es.htm#II.%20independent>.
- Briggs, C., & Young, B. (2003). Does Reading Recovery work in Kansas? A retrospective longitudinal study of sustained effects. *The Journal of Reading Recovery*, 3(1), 59–64.
- Bordens, K & Abbott, B. (1999). *Research design and methods: A process approach* (4th Ed.). Mountain View:CA: Mayfield Publishing Company.
- Chesney, G., & Fink, S. (2007). *Results for the 2007 Spring Reading Benchmark Assessments*. Retrieved January 25, 2007, from www.montgomeryschoolsmd.org/departments/sharedaccountability/.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Lawrence Hillsdale, NJ: Erlbaum and Associates.
- D'Agostino, J., & Murphy, J. (2004). A meta-analysis of Reading Recovery in United States schools. *Educational Evaluation and Policy Analysis*, 26 (1), 23–28.
- D'Agostino, R. B. (1998). Propensity score method for biased reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17 (19), 2265–2281.
- Douglas, A. (2006). *Trend results for spring reading benchmark*. Retrieved November 10, 2007, from ps.k12.md.us/departments/sharedaccountability/reports/2007/.
- Forbes, S., & Szymczuk, M. (2004). *Putting Reading First with Reading Recovery: Sustained effects of Reading Recovery in Iowa*. Retrieved January 15, 2008, from www.rrci.us/Results/SAI%20Report%207%20-%2011%20-%2004.pdf.
- Glynn, T., Crooks, T., Bethune, N., Ballard, K., & Smith, J. (1989). Reading Recovery in context: Implementation and outcomes. *Educational Psychology*, 12(3&4), 249–261.

- Kline, R. B. (2004). *Beyond significance testing*. Washington, D.C.: American Psychological Association.
- Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher*, 31(1), 3–12.
- Luellen, J. K., Shadish, W.R., & Clark, M.H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review*, 29, 530–558.
- Maryland State Department of Education. (2005). *Division of Accountability and Assessment: HSA, MSA, and Alt MSA HOSS, LOSS, Cut Scores and Proficiency Level Information*. Retrieved February 10, 2008, from www.montgomeryschoolsmd.org/departments/sharedaccountability/testing/PDFs/Cut%20Scores_102605.pdf.
- McDowall, S., Boyd, S., Hodgen, E., & van Vliet, T. (2005). *Reading Recovery in New Zealand: Uptake, implementation, and outcomes, especially in relation to Māori and Pasifika students*. Wellington: Ministry of Education. Retrieved December 20, 2007, from www.nzcer.org.nz/default.php?cPath=130_131&products_id=1604.
- Montgomery County Public Schools. (2007). *Description of Reading Recovery teacher position*. Retrieved May 2, 2007, from www.montgomeryschoolsmd.org/departments/titleone/documents/innovative_position/reading_recovery.pdf.
- 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.
- Pinnell, G.S. (1989). Reading Recovery: Helping at-risk children learn to read. *The Elementary School Journal*, 90 (2), 159–181.
- Randall, L. (2004). *Reading Recovery site report: Montgomery County 2003–2004*, Rockville, MD: Montgomery County Public Schools.
- Randall, L. (2005). *Reading Recovery site report: Montgomery County 2004–2005*, Rockville, MD: Montgomery County Public Schools.
- Randall, L. (2006). *Reading Recovery site report: Montgomery County 2005–2006*, Rockville, MD: Montgomery County Public Schools.
- Randall, L. (2007). Personal communication.
- Rodgers, E., Wang, C., & Gómez-Bellengé, F. (2004). *Closing the literacy achievement gap with early intervention*. Retrieved February 19, 2005, from paper%20Closing%20the%20Achievement%20Gap.pdf.
- Rubin, D. B. (1997). *Estimation from nonrandomized treatment comparisons using subclassification on propensity scores*. *Nonrandomized Comparative Clinical Studies*. Retrieved April 16, 2007, from www.symposium.com/nrccs/rubin.htm.

- Schmitt, M. C., Gregory, A. E., & Younts, T. (2001, April 28–May 3). *The impact of early literacy interventions: Where are the children now?* Paper presented at the National Reading Conference, San Antonio, TX.
- Schmitt, M. C., Gregory, A. E., & Younts, T. (2005). The impact of early literacy interventions: Where are the children now? *Literacy Teaching and Learning*, 10 (1), 1–20.
- Schwartz, R. (2005). Literacy learning of at-risk first grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, (97), 257–267.
- 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(4), 958–996.
- Springer, R. (2006). Using effect sizes in NSSF survey reporting. *Research and Practice in Assessment*, 1 (1), 1–19.
- Talwar, J., & Hill, S. (1993). *Interim report. Reading Recovery: Halifax. 1989–1992 (Tech. Rep.)*. Halifax, Nova Scotia: Halifax District School Board.
- Timperley, H., Fung, I., Wilson, A., & Barrar, H. (2006, April 7–16). *Professional learning and development: A best evidence synthesis of impact on student outcomes*. Paper presented to the annual meeting of the American Educational Research Association, San Francisco, California.
- U.S. Department of Education. (2001). *The Nation's report card: Fourth-grade reading 2000*. Retrieved January 25, 2005, from <http://nces.ed.gov/nationsreportcard/data/ppt/1>.
- U.S. Department of Education. (2007). What Works Intervention Report: Reading Recovery. Institute of Education Services: *The nation's report card: Fourth-grade reading 2000*. Retrieved August 2007, from http://www.readingrecovery.org/pdf/rrcna/news/WWC_Media_Release-3-19-07.pdf.
- Wasik, B., & Slavin, R. (1993). Preventing early reading failure with one-on-one tutoring: A review of five programs. *Reading Research Quarterly*, 28, 179–200.
- West, J., Denton, K., & Reaney, L. M. (2000). The kindergarten year: Findings from the early childhood longitudinal study, kindergarten class of 1998–99. *Education Statistics Quarterly*, 2(4), 25–30.
- Zhao, H., & Von Secker, C. (2007). *Performance of the Montgomery County Public Schools Grade 2 Students on the 2007 TerraNova Second Edition*. Retrieved December 20, 2007, from www.montgomeryschoolsmd.org/departments/sharedaccountability/reports/2007/TN2%20BriefFinal%202007101507.pdf.

ⁱ. The author would like to thank Ms. Natalie Wolanin for compiling the assessment data analyzed in this report, Ms. Maria Jose Allendes for technical assistance, and Mrs. Ada Miranda-Morales for providing extensive editorial and technical assistance on the formatting of this report.

Appendix A: Elementary Schools Implementing Reading Recovery

Table A1
Schools Implementing Reading Recovery

School	Title I School	Implementation Status	Number of Students		
			2003–2004 <i>N</i>	2004–2005 <i>N</i>	2005–2006 <i>N</i>
Beall	No	Partial	8	8	9
Broad Acres	Yes	Full ^c	21	20	20
Brooke Grove	No	Partial	8	8	8
Brookhaven	No	Partial	8	8	7
Brown Station	No	Partial	8	9	9
Burnt Mills	Yes	Partial	9	9	8
Cannon Road	No	Partial	8	8	8
Clopper Mill	No	Partial	16	8	
Dr. Charles R. Drew	No	Partial	9	8	7
East Silver Spring	Yes	Partial	13	--	7
Fairland	No	Partial	6	7	7
Gaithersburg	Yes	Partial	12	14	11
Galway	No	Partial	6	6	6
Glenallan	No	Partial	8	9	8
Greencastle	No	Partial	8	--	8
Harmony Hills	Yes	Partial	8	^a	^a
Highland	Yes	Partial	8	^a	^a
Highland View	No	Partial	8	8	12
Jackson Road	Yes	Partial	12	12	12
Kemp Mill	Yes	Partial	17	17	17
Montgomery Knolls	Yes	Partial	8	8	25
New Hampshire Estates	Yes	Full ^c	35	31	34
Rock Creek Valley	No	Partial		8	9
Rolling Terrace	Yes	Full ^c	20	28	29
Ronald McNair	No	Partial	9	8	-2
Rosemont	Yes	Partial	9	^a	^a
Sargeant Shriver	Yes	Partial		^b	^b
South Lake	Yes	Partial			8
Strawberry Knoll	No	Partial	8	9	8
Summit Hall	Yes	Partial	8	^a	^a
Takoma Park	No	Partial	17	14	10
Twinbrook	Yes	Partial	8	8	7
Viers Mill	Yes	Partial	8	8	9
Watkins Mill	No	Partial	18	17	8
Weller Road	Yes	Partial	16	16	17
Wheaton Woods	Yes	Partial	8	^a	^a
William Tyler Page	No	Partial	8	--	--
Woodlin	No	Partial	9	9	--
Total		Partial	390	323	338

^a Reading First schools. ^b New school; no Reading Recovery students during these years.

^b School opened in 2006–2007 school year.

^c 2003–2004: Broad Acres and New Hampshire Estates Elementary Schools; 2004–2005: Broad Acres, New Hampshire Estates, and Rolling Terrace Elementary Schools; 2005–2006: New Hampshire Estates and Rolling Terrace Elementary Schools.

Appendix B: Demographic Characteristics of Students by Cohort

Table B1

Demographic Characteristics of Students by Sample and Cohort Year

Cohort	Characteristic		Sample					
			Random		Reading Recovery		Non-RR	
			<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
2003–2004 (1)	Race	African American	128	31.5	60	28.4	136	33.1
		American Indian	2	n/a	0	n/a	2	n/a
		Asian American	32	7.9	18	8.5	27	6.6
		Hispanic	195	48.0	104	49.3	198	48.2
		White	49	12.1	28	13.3	49	11.9
	Gender	Female	175	43.1	91	42.5	178	43.3
		Male	231	56.9	122	57.0	233	56.7
	ESOL	No	214	52.7	101	47.2	214	52.1
		Yes	192	47.3	113	52.8	197	47.9
	FARMS	No	119	31.6	67	32.1	128	31.1
		Yes	257	68.4	142	67.9	283	68.9
	Special Education	No	343	91.2	194	92.8	341	83.0
		Yes	33	8.8	15	7.2	70	17.0
	2004–2005 (2)	Race	African American	91	28.8	43	25.1	103
American Indian								
Asian American			19	6.0	13	7.6	25	6.6
Hispanic			152	48.1	82	48.0	196	51.7
White			54	17.1	33	19.3	55	14.5
Gender		Female	153	48.4	67	39.0	175	46.2
		Male	163	51.6	102	59.3	204	53.8
ESOL		No	192	60.8	95	55.2	228	60.2
		Yes	124	39.2	77	44.8	151	39.8
FARMS		No	128	40.5	69	40.1	155	40.9
		Yes	188	59.5	103	59.9	224	59.1
Special Education		No	279	88.3	137	95.1	324	85.5
		Yes						
2005–2006 (3)		Race	African American	124	33.7	53	27.6	130
	American Indian		3	n/a	3	n/a	3	n/a
	Asian American		28	7.6	12	6.3	30	7.7
	Hispanic		186	50.5	117	60.9	202	51.5
	White		27	7.3	10	5.2	27	6.9
	Gender	Female	154	41.8	85	44.0	161	41.1
		Male	214	58.2	107	55.4	231	58.9
	ESOL	No	169	45.9	71	36.8	169	43.1
		Yes	199	54.1	122	63.2	223	56.9
	FARMS	No	120	32.6	52	26.9	122	31.1
		Yes	248	67.4	141	73.1	270	68.9
	Special Education	No	247	86.1	152	93.8	235	74.8
		Yes	40	13.9	10	6.2	79	25.2

Note. Percentage for subgroups with fewer than five students is not reported. RR= Reading recovery

Appendix C: Variables in the Analysis of Covariance Model

Table C1

Variables used in the Analysis of Covariance Models by Grade Level

Grade level	Measure of initial reading ability	Dependent	Control variables
Grade 1 Cohort 1 Cohort 2 Cohort 3	Fall Grade 1 TRC level Fall Grade 1 TRC level Fall maximum word recognition score	Spring Grade 1 TRC level	Race/ethnicity; FARMS, special education, ESOL, propensity score.
Grade 2 Grade 3	Grade 1 spring TRC level	Grade 2 spring TRC level; TN/2 scale score	Race/ethnicity; FARMS, special education, ESOL, propensity score.
	TN/2 scale score	MSA scale score, MAP-R RIT score	Race/ethnicity; FARMS, special education, ESOL, propensity score.

Note. A separate propensity score was computed for each cohort. This reduction of covariates from many characteristics to one composite characteristic helped simplify the pre-existing characteristics from many variables to one and helped balance the groups on the list of covariates (Rubin, 1997). Stratification of the propensity score is as efficient as matching for balancing non-equivalent groups and eliminates approximately 90% of bias in each of the covariates that could be removed by matching (Lueleen, Shadish, & Clark, 2006; D'Agostino, 1998).

Table C2

Variables in the Logistic Regression Model by Cohort by Grade Level

Grade Level	Measure of initial ability	Dependent Variables	Covariates
Grade 1 Cohort 1 Cohort 2 Cohort 3	End of K TRC level Fall Grade 1 TRC level Fall maximum word recognition score	Grade 1 benchmark status	Race/ethnicity; FARMS, special education, ESOL; propensity score.
Grade 2	Fall Grade 2 TRC	Grade 2 end-of-year benchmark status, at or above 50 th NCE	Race/ethnicity; FARMS, special education, ESOL, propensity score.
Grade 3	TN/2 scale score	At or above proficient on Grade 3 reading MSA	Race/ethnicity; FARMS, special education, ESOL, propensity score.

Appendix D: Calculation of Effect Sizes

Effect sizes (ES) for comparing groups on continuous outcome. In this study, mean differences on reading measures (Treatment – comparison groups’ scores) were converted to a measure of effect size called Cohen’s d (Cohen, 1988). The formula used to calculate magnitude of d in this study, is $d = \sqrt{F(n_1+n_2)/n_1n_2}$. F is $n_1=n$ for RR sample, n_2 = number of students in the comparison sample.

Cohen’s d is a descriptive measure, which indicates the relative importance of the difference in the measures between two groups. By convention, the subtraction, $M_1 - M_2$, is done so that the difference is positive if it is in the direction of *improvement* or in the predicted direction, and negative if in the direction of *deterioration*. In this study, negative effect sizes indicate that the reading achievement for RR students lags behind that of the comparison group.

Conventional thresholds for the size of Cohen’s d are 0.20 for a small effect, 0.50 for a moderate effect, and 0.80 for a large effect (Cohen, 1988).

ES can also be interpreted in terms of the percent of non-overlap of the treated group’s scores with those of the untreated group (Cohen, 1988). An ES of 0.0 indicates that the distribution of scores for the treated group overlaps completely with the distribution of scores for the untreated group; there is 0% of nonoverlap. An ES of 0.8 indicates a nonoverlap of 47.4% in the two distributions. An ES of 1.7 indicates a nonoverlap of 75.4% in the two distributions (Becker, 2000).

An effect size also is interpreted as exactly equivalent to a “Z-score” of a standard Normal distribution. For example, an effect size of 0.8 means that the score of the average person in the RR (experimental) group is 0.8 standard deviations above the average person in the comparison group (random or non-RR), and hence exceeds 79% of the students in the control group (Springer, 2006)..

Effect sizes can also be thought of as the average percentile standing of the average treated (or experimental) participant relative to the average untreated (or control) participant. An ES of 0.0 indicates that the mean of the treated group is at the 50th percentile of the untreated group. An ES of 0.8 indicates that the mean of the treated group is at the 79th percentile of the untreated group. An effect size of 1.7 indicates that the mean of the treated group is at the 95.5 percentile of the untreated group (Springer, 2006; Becker 2002).

Effect sizes for comparing groups across categorical measures. For the categorical outcomes (achieving MCPSAP-PR Grade 1 and 2 benchmark, being at the 50th percentile or above on TN/2 assessment, and achieving proficient and above on MSA), a logistic regression model was used to compute odds ratio for meeting these benchmarks. Then, the odds ratio was used to compute a logit d . Logit d is comparable to Cohens d (Kline, 2004).

Kline (2004) provides a formula for converting an odds ratio to a logit d , which is then comparable to standardized mean difference effect sizes expressed as d . This puts effect sizes, Cohen’s d and the odds ratio, on a common metric. That formula is:

$$\text{logit } d = \frac{\ln(OR)}{\pi / \sqrt{3}}$$

For example, the odds of Cohort 1 RR students meeting the Grade 1 benchmark were 1.069 of random sample of grade level peers students who were eligible for RR but did not attend. Values greater than 1 indicate greater odds of RR students achieving the benchmark whereas values less than 1 indicate greater odds of non-RR achieving the benchmark. The further the odds ratio is from 1, the larger the effect. Converting this odds ratio to a logit d results in:

$$\text{logit } d = \frac{\ln(1.069)}{\pi / \sqrt{3}} = \frac{0.03}{1.81} = 0.02$$

This value is negligible and indicates that 50% of the comparison group would be below the average student in the RR group—essentially, the performance of the groups is comparable.

Appendix E: Reading Outcomes for Random and Reading Recovery Students

Table E1a

Measures of Effect Sizes for Association Between Grade 1 End-of-Year Benchmark Status and Participation in Reading Recovery by Cohort

Cohort	Odds ratio	Logit <i>d</i> (Effect Sizes)
(1) 2003–2004	1.64	0.27
(2) 2004–2005	4.804	0.87
(3) 2005–2006	2.521	0.51

Table E1b

Measures of Effect Sizes for Association Between Grade 2 End-of-Year Benchmark Status and Participation in Reading Recovery by Cohort

Cohort	Odds ratio	Logit <i>d</i> (Effect Sizes)
(1) 2003–2004	0.576	-0.30
(2) 2004–2005	0.509	-0.37
(3) 2005–2006	0.427	-0.47

Table E1c

Measures of Effect Sizes for Association Between Being at or Above the 50th NCE and Participation in Reading Recovery by Cohort

Cohort	odds ratio	Logit <i>d</i> (Effect Sizes)
(1) 2003–2004	0.485	-0.40
(2) 2004–2005	0.143	-1.07
(3) 2005–2006	0.343	-0.59

Table E1d

Measures of Effect Sizes for Association Between MSA Proficiency Level and Participation in Reading Recovery by Cohort

Cohort	odds ratio	Logit <i>d</i> (Effect Sizes)
(1) 2003–2004	0.601	-0.28
(2) 2004–2005	0.632	-0.29

Table E2
Means and Standard Deviations for Grade 1 Reading Measures by Cohort Year

Measure	Statistic	Cohort 1 (2003–2004)		Cohort 2 (2004–2005)		Cohort 3 (2005–2006)	
		Random	RR	Random	RR	Random	RR
Fall Word Recognition	<i>N</i>	--	--	--	--	263	184
	SD	--	--	--	--	16.5	20.0
	Mean	--	--	--	--	55.0	47.4
Fall Text Level	Mean	6.1	2.6	4.1	1.7	--	--
	<i>N</i>	281	186	316	172	--	--
	SD	4.6	1.0	6.0	3.2	--	--
Spring Grade 1 TRC ^a	Mean	14.3	14.3	4.4	1.9	5.9	2.2
	<i>N</i>	360	209	316	172	317	189
	SD	6.9	3.4	14.3	15.6	14.2	15.1
TRC change in Grade 1	Mean	9.7	12.0	3.7	2.1	--	--
	<i>N</i>	276	186	316	172	--	--
	SD	5.0	3.2	8.5	12.3	--	--

^a Grade 1 benchmark TRC level is 16. Fall TRC level should be levels 5–7. Change in TRC scores is computed only for students with both fall and spring TRC scores.

Figures E1

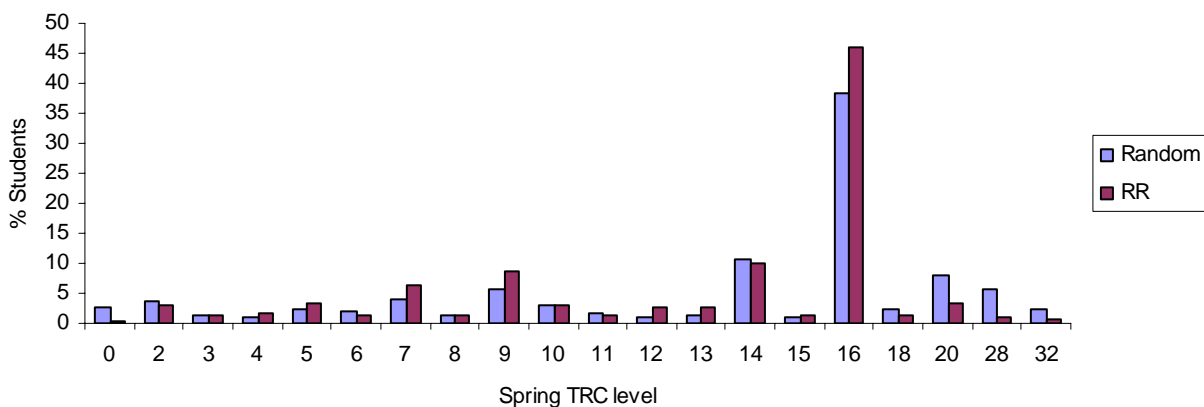


Figure E1a. Frequency distribution for Grade 1 spring TRC levels for Grade 1 cohort 1 students.

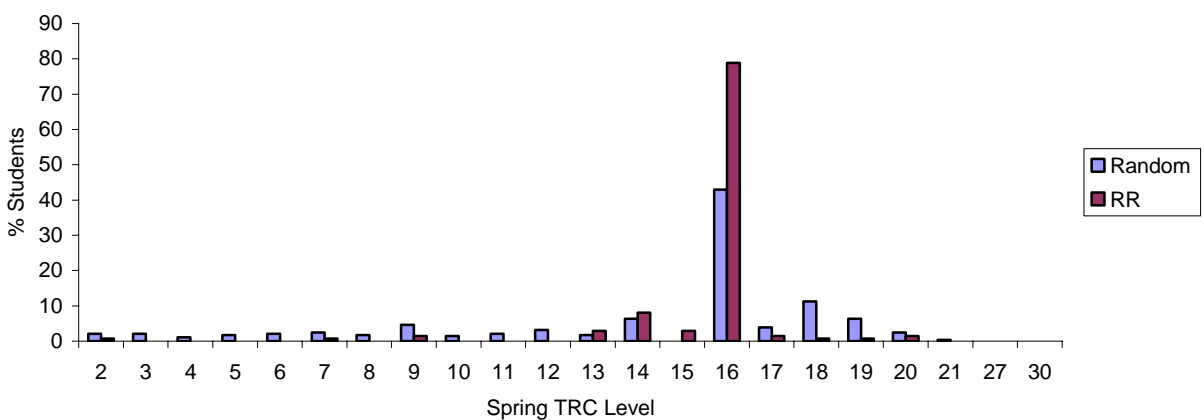


Figure E1b. Frequency distribution for Grade 1 spring TRC levels for Grade 1 cohort 2 students.

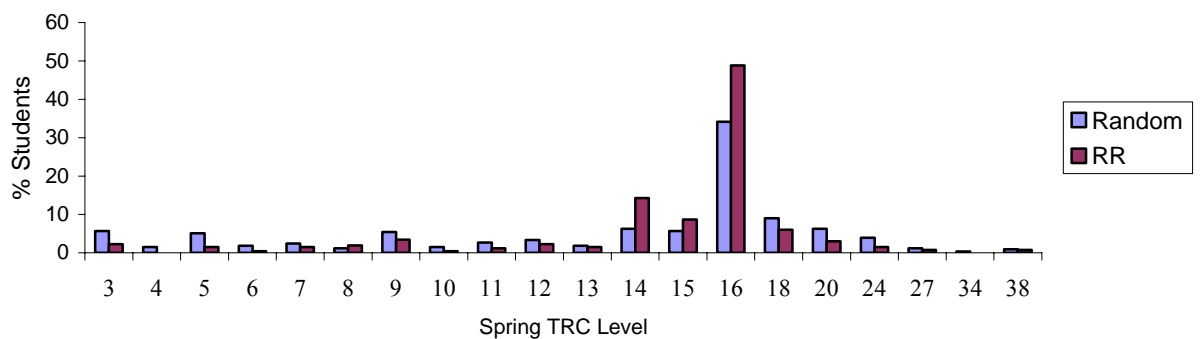


Figure E1c. Frequency distribution for Grade 1 spring TRC levels for Grade 1 cohort 3 students.

Table E3
Means and Standard Deviations for Fall and Spring Grade 1 TRC Levels for Cohort 1 (2003–2004) by Subgroup

Subgroup		Random						RR					
		Fall TRC			Spring TRC			Fall TRC			Spring TRC		
		Mean	N	SD	Mean	N	SD	Mean	N	SD	Mean	N	SD
Race	African American	7.2	128	4.9	15.4	128	7.0	2.6	60	0.9	14.5	60	3.4
	American Indian	--	--	--	--	--	--	--	--	--	--	--	--
	Asian American	6.1	32	3.9	16.8	32	6.2	2.8	18	1.6	14.3	18	4.7
	Hispanic	4.7	195	3.5	12.5	195	6.0	2.5	104	1.0	14.2	104	3.4
	White	7.8	49	5.9	17.4	49	8.0	2.4	28	0.6	14.6	28	2.7
Gender	Female	6.3	175	5.0	14.3	175	7.3	2.5	91	0.7	14.5	91	3.2
	Male	5.9	231	4.2	14.3	231	6.5	2.6	122	1.1	14.2	122	3.6
FARMS	Non-FARMS	7.5	119	5.6	16.6	119	7.5	2.5	67	1.1	14.7	67	3.3
	FARMS	5.3	257	3.8	13.2	257	6.3	2.6	142	0.9	14.2	142	3.5
ESOL	No	7.6	214	5.1	16.8	214	6.4	2.7	101	1.1	14.6	101	2.9
	Yes	4.0	192	2.6	11.6	192	6.3	2.4	113	0.9	14.1	113	3.8
Special Ed.	No	6.2	343	4.7	14.8	343	6.7	2.5	194	0.9	14.3	194	3.4
	Yes	4.3	33	2.4	8.7	33	6.6	3.3	15	2.0	14.3	15	3.4

Note. Results for subgroups with fewer than five students are not reported.

Table E4
Means and Standard Deviations for Fall and Spring Grade 1 TRC Levels for Cohort 2 (2004–2005) by Subgroup

Subgroup		Random						RR					
		Fall TRC			Spring TRC			Fall TRC			Spring TRC		
		Mean	N	SD	Mean	N	SD	Mean	N	SD	Mean	N	SD
Race	African American	6.7	91	4.3	14.8	91	4.4	2.8	43	0.8	15.7	43	1.2
	American Indian	--	--	--	--	--	--	--	--	--	--	--	--
	Asian American	7.3	19	4.4	15.2	19	4.1	3.1	13	1.0	15.5	13	1.0
	Hispanic	4.7	152	2.8	13.3	152	4.3	3.3	82	1.5	15.7	82	1.4
	White	8.3	54	5.3	15.9	54	4.3	3.8	33	3.0	15.2	33	3.4
Gender	Female	6.1	153	4.2	14.4	153	4.4	3.1	67	1.3	15.8	67	1.2
	Male	5.9	163	4.0	14.2	163	4.4	3.3	102	2.0	15.5	102	2.0
FARMS	No FARMS	7.4	128	4.7	15.7	128	3.8	3.2	69	2.2	15.5	69	2.3
	FARMS	5.1	188	3.4	13.4	188	4.6	3.3	103	1.3	15.6	103	1.4
ESOL	No	7.1	192	4.5	15.3	192	4.0	3.3	95	2.0	15.7	95	2.0
	Yes	4.5	124	2.8	13.0	124	4.6	3.1	77	1.3	15.3	77	1.7
Special Ed.	No	6.1	279	4.1	14.7	279	4.0	3.2	137	1.7	15.7	137	1.3
	Yes	5.2	37	4.3	11.2	37	6.0	4.6	7	2.4	15.3	7	1.3

Note. SD= standard deviation. TRC=text reading comprehension. Results for subgroups with fewer than five students are not reported.

Table E5
Means and Standard Deviation for Fall, Spring, and Change in Grade 1 TRC Levels for Cohort 3 (2005–2006) by Subgroup

Subgroups	Random						RR						
	Fall TRC			Spring TRC			Fall TRC			Spring TRC			
	Mean	N	SD	Mean	N	SD	Mean	N	SD	Mean	N	SD	
Race	African American	2.3	124	0.9	14.9	124	6.7	2	53	0.0	15.0	53	2.5
	American Indian	--	--	--	--	--	--	--	--	--	--	--	--
	Asian American	4.0	28	4.5	18.1	28	7.8	--	12	0.0	15.4	12	1.3
	Hispanic	2.6	186	1.7	12.8	186	4.8	2	117	0.0	15.3	117	2.2
	White	2.4	27	0.5	15.1	27	5.9	--	10	--	15.0	10	2.6
	Hispanic	2.6	186	1.7	12.8	186	4.8	2	117	0.0	15.3	117	2.2
Gender	Female	2.6	154	1.6	15.2	154	5.6	2	85	0.0	15.3	85	2.0
	Male	2.5	214	1.9	13.4	214	6.1	2	107	0.0	15.1	107	2.4
FARMS	No	3.0	120	2.5	15.3	120	6.7	--	52	--	15.3	52	2.1
	Yes	2.2	248	0.8	13.6	248	5.5	2	141	0.0	15.1	141	2.3
ESOL	No	2.8	169	2.0	15.9	169	6.2	--	71	--	15.1	71	2.4
	Yes	2.0	199	0.0	12.6	199	5.3	2	122	0.0	15.2	122	2.2
Special ed.	No	2.7	247	2.0	15.4	247	5.6	2	152	0.0	15.1	152	2.3
	Yes	2.0	40	0.0	9.0	40	4.6	--	10	--	15.7	10	0.7

Note. TRC=text reading comprehension. Results for subgroups with less than five students are not reported. --=data not available for majority of students.

Table E6
Adjusted Mean Differences for Grade 1 TRC Levels and Measures of Effect Sizes for Association Between Grade 1 Spring TRC Level and Participation in Reading Recovery

Cohort	Mean Difference	p-value for mean difference (2-tailed t-test)	Standard Error of estimate	Effect Size
Cohort 1 (2003–2004)	1.42	0.00	0.43	0.32
Cohort 2 (2004–2005)	2.067	0.00	0.43	0.76
Cohort 3 (2005–2006)	2.441	0.00	0.75	0.68

Note: ES=Effect Size

Appendix F: Grade 2 Outcomes for Random and Reading Recovery Students

Table F1

Means and Standard Deviation for Grade 2 Reading Measures by Cohort Year

Measure	Statistic	Cohort					
		Cohort 1 2003–2004		Cohort 2 2004–2005		Cohort 3 2005–2006	
		Random	RR	Random	RR	Random	RR
Grade 2 fall TRC level	Mean	14.2	14.4	14.3	15.7	14.91	14.40
	<i>N</i>	308	176	316	172	274	153
	SD	4.6	2.5	5.1	1.6	6.42	3.17
Grade 2 spring TRC level	Mean	26.2	24.7	25.3	23.4	23.76	22.36
	<i>N</i>	89	99	316	172	277	161
	SD	7.8	5.4	7.8	5.4	7.03	4.81
TN/2 scale score	Mean	609.3	589.4	614.4	588.5	607.17	590.36
	<i>N</i>	308	179	316	172	308	162
	SD	38.6	27.9	55.8	57.7	32.22	24.20
TN/2 NCE	Mean	52.6	42.1	52.6	39.6	48.08	38.41
	<i>N</i>	308	179	316	172	308	162
	SD	19.6	14.2	20.8	16.0	16.49	13.20

Note. Grade 2 benchmark TRC level is 27 (M). Fall TRC level should be levels 18–20 (J–K).

Table F2

Adjusted Mean Differences for Grade 2 Reading Measures and Measures of Effect sizes for Association Between Grade 2 Reading Measures and Participation in Reading Recovery by Cohort

Outcome measure	Mean Difference	<i>p</i> -value for Mean Difference (2-tailed <i>t</i> -test)	Standard Error of Estimate	Effect Size
Cohort 1(2003–2004)				
Spring Grade 2 TRC level	0.57	0.28	1.28	0.09
TN/2 scale score	-14.18	0.01	3.13	-0.46
Cohort 2(2004–2005)				
Spring Grade 2 TRC level	-2.40	0.00	0.10	-0.42
TN/2 scale score	-29.07	0.00	8.58	-0.44
Cohort 3(2005–2006)				
Spring Grade 2 TRC level	-1.13	0.00	0.56	-0.26
TN/2 scale score	-10.49	0.00	2.84	-0.37

Figures F

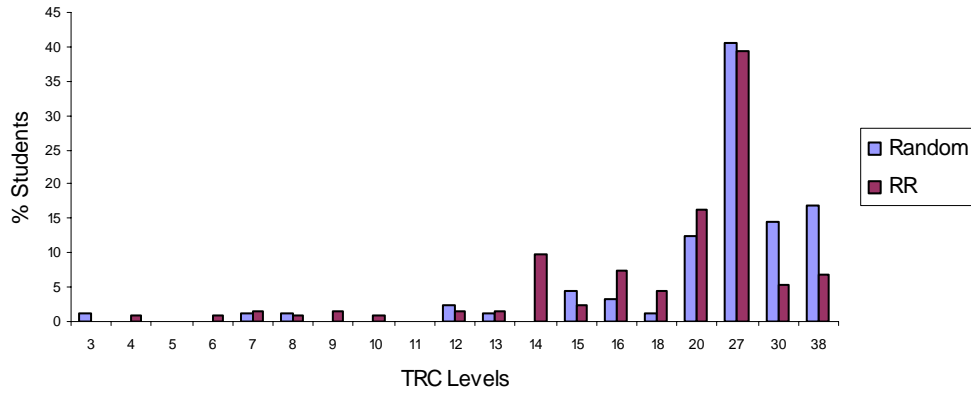


Figure F1. Frequency distribution of students across spring TRC levels for Grade 2 cohort 1 students.

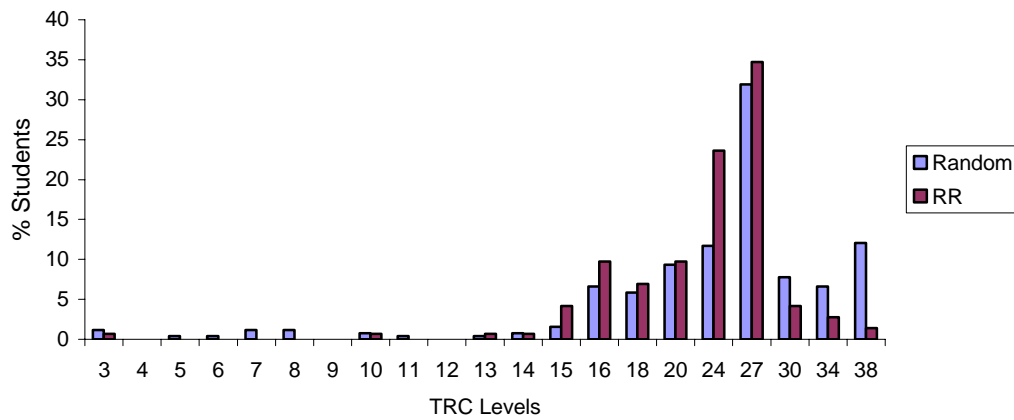


Figure F2. Frequency distribution of students across spring TRC levels for Grade 2 cohort 2 students.

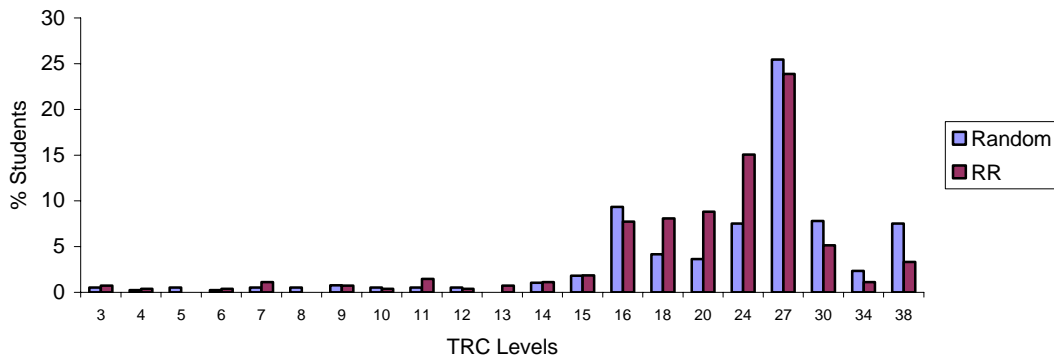


Figure F3. Frequency distribution of students across spring TRC levels for Grade 2 cohort 3 students.

Table F3

Summary Statistics for Grade 3 Reading Measures for Random and Reading Recovery Students

Measure	Cut-off Score for MSA Proficiency	Statistic	Cohort			
			Cohort 1 (2003–2004)		Cohort 2 (2004–2005)	
			Random	RR	Random	RR
Grade 3 MSA Scale Score	388	Mean	407.4	389.6	416.3	395.1
		<i>N</i>	313	168	316	172
		SD	33.5	25.9	39.3	31.2
Grade 3 MAP-R RIT_SCORE	190	Mean	196.7	187.6	198.4	188.4
		<i>N</i>	293	160	316	172
		SD	15.5	11.6	15.5	14.5

Table F4

Adjusted Mean Differences Between Random and Reading Recovery Students for Grade 3 Reading Measures and Measures of Effect Sizes for Association Between Reading Achievement and Participation in Reading Recovery by Cohort

Outcome measure	Mean Difference	<i>p</i> -value (2-tailed <i>t</i> -test)	Standard Error of estimate	Effect Size
Cohort 1(2003–2004)				
MSA Scale score	-6.60	0.02	2.85	-0.25
MAP-RIT score	-3.74	0.00	1.19	-0.34
Cohort 2 (2004–2005)				
MSA Scale score	-10.177	0.01	5.37	-0.29
MAP-R RIT score	-6.794	0.00	2.27	-0.50

Appendix G: Outcomes for Reading Recovery and non-Reading Recovery^a Students

Table G1

Summary Statistics for Grade 1 and 2 TRC Level, TN/2, MSA, and MAP-R RIT Scale Scores for Reading Recovery and non-Reading Recovery Students

Measure	Reading Recovery			Non-RR		
	Unadjusted Mean	N	SD	Unadjusted Mean	N	SD
Cohort 1 (2003–2004)						
Grade 1 Fall TRC	2.55	186	0.99	2.00	411	0.00
Grade 1 Spring TRC - Change in Grade 1 TRC	14.34	209	3.40	10.72	408	4.55
Fall Grade 2 TRC	11.98	186	3.21	8.72	408	4.55
Spring Grade 2 TRC	14.39	176	2.50	10.83	349	4.63
Change Grade 2 TRC	24.74	99	5.45	19.36	87	7.77
CTBS scale score	7.51	169	5.39	6.54	339	4.28
CTBS NCE	589.44	179	27.90	592.76	344	33.96
MSA Scale score	42.07	179	14.19	43.78	344	17.65
RIT score	389.65	168	25.94	389.76	311	28.10
	187.63	160	11.59	188.64	289	15.31
Cohort 2 (2004–2005)						
Grade 1 Fall Text Level	3.22	158	1.74	4.11	301	3.98
Grade 1 Spring TRC - Change in Grade 1 TRC	15.56	137	1.88	11.60	298	5.70
Fall Grade 2 TRC	12.35	130	2.06	8.06	269	4.63
Spring Grade 2 TRC						
Change in Grade 2 TRC	23.42	144	5.44	22.02	253	8.12
TN/2 scale score	--	--	--	--	--	--
TN/2 NCE	588.53	141	57.74	602.41	265	36.02
MSA Scale score						
RIT score	395.10	124	31.22	407.00	247	35.88
	188.45	112	14.54	192.70	233	16.26
Cohort 3 (2005–2006)						
Grade 1 Fall Text Level	--	--	--	--	--	--
Grade 1 Spring TRC - Change in Grade 1 TRC	15.16	189	2.24	10.88	294	4.92
Fall Grade 2 TRC	--	--	--	--	--	--
Spring Grade 2 TRC	14.40	153	3.17	10.17	280	5.39
Change in Grade 2 TRC	22.36	161	4.81	18.70	293	7.42
TN/2 scale score	--	--	--	--	--	--
TN/2 NCE	590.36	162	24.20	594.53	261	35.84
	38.41	162	13.20	41.74	261	15.92

^a Students reading grade-level as indicated by spring K TRC benchmark or fall Grade 1 mClass 3D TRC level=similar to Reading Recovery students in initial reading ability.

Figures G

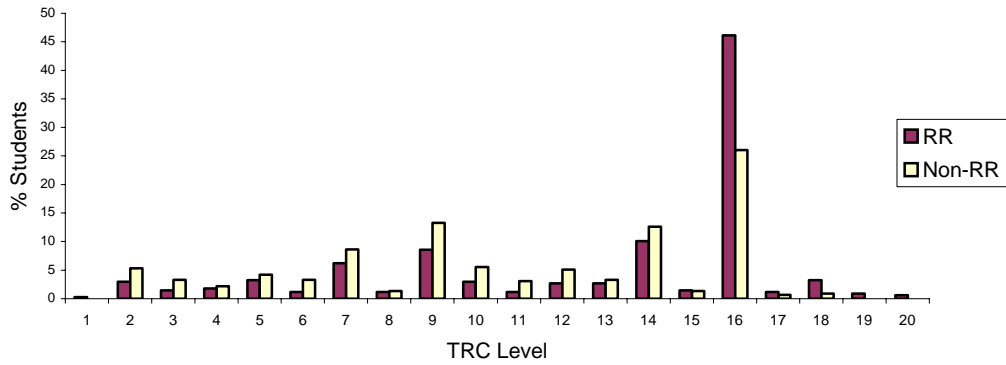


Figure G1a. Frequency distribution of students across TRC levels for Grade 1 cohort 1 students.

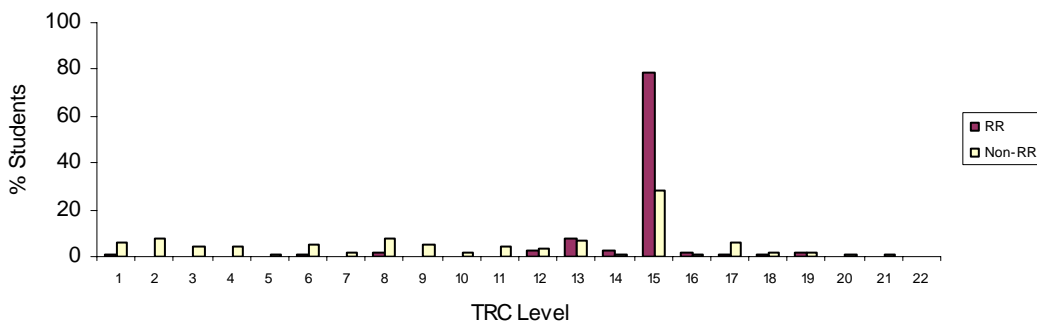


Figure G1b. Frequency distribution of students across TRC levels for Grade 1 cohort 2 students.

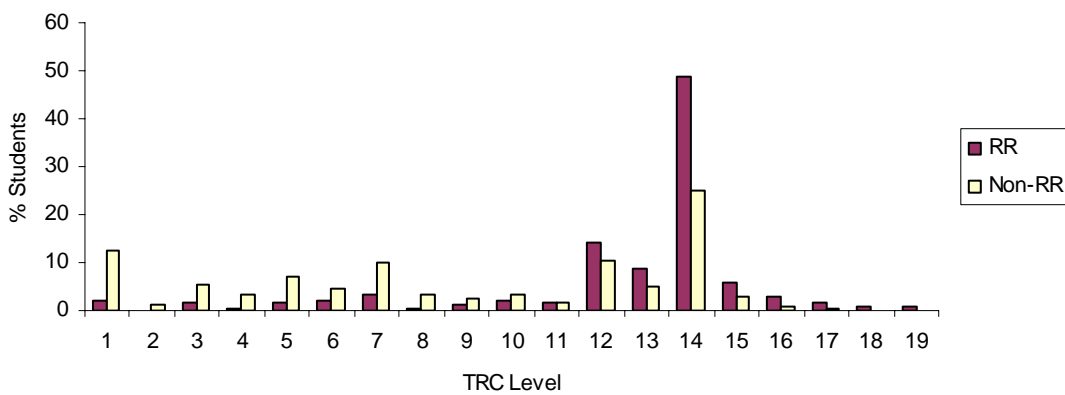


Figure G1c. Frequency distribution of students across TRC levels for Grade 1 cohort 3 students.

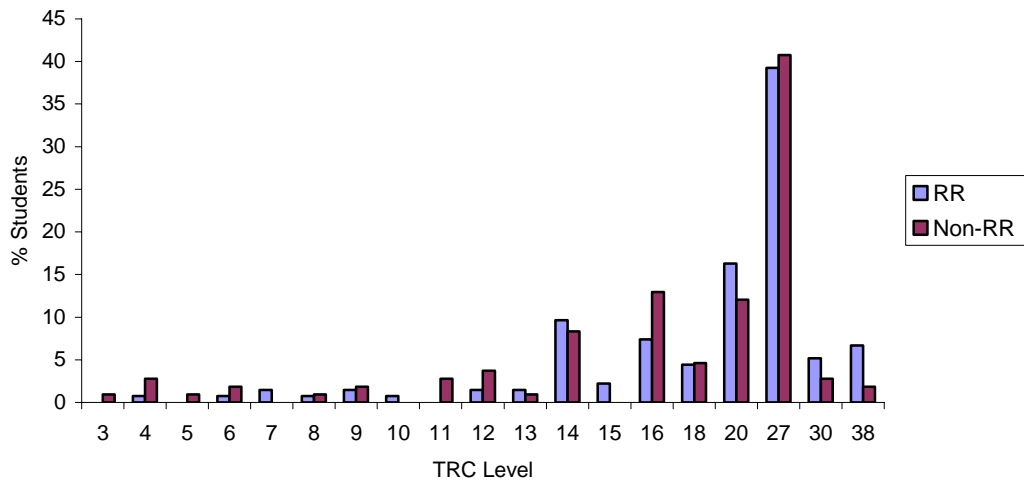


Figure G2a. Frequency distribution of students across TRC levels Grade 2 cohort 1 students.

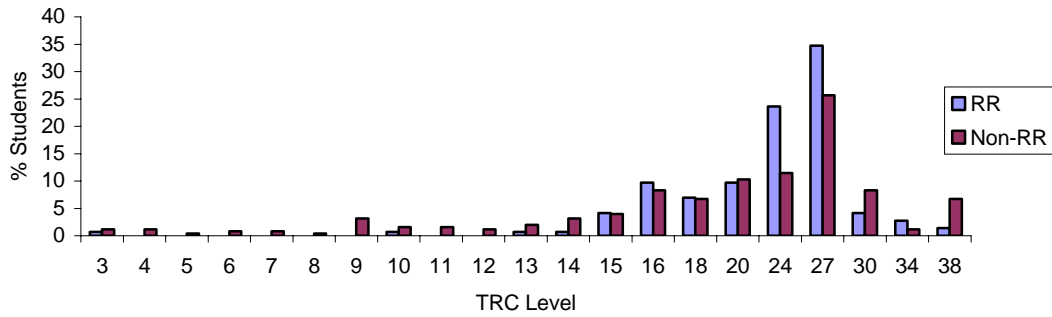


Figure G2b. Frequency distribution of students across TRC levels Grade 2 cohort 2 students.

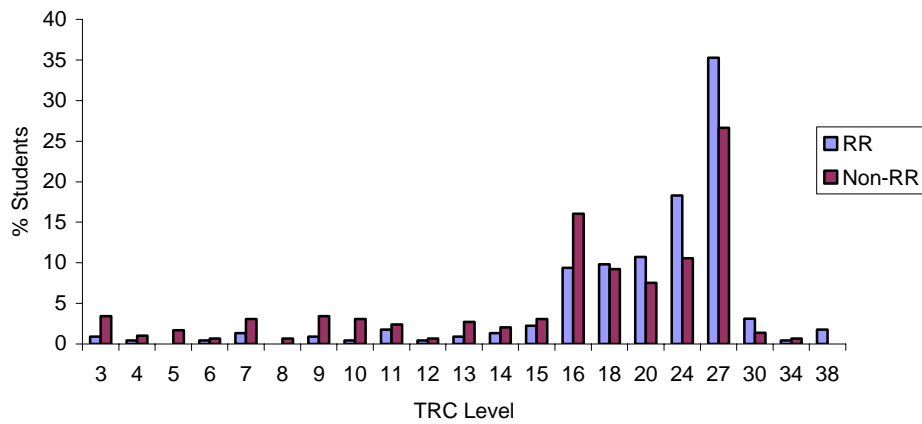


Figure G2c. Frequency distribution of students across TRC levels Grade 2 cohort 3 students.

Appendix H: Assessment Data

Table H1

Number and Percentage of Students Meeting Grade 1 End-of-Year Benchmark for 2007 by Elementary School and Implementation Status

School	Male		Female		African American		Asian American		Hispanic		White		FARMS		SP ED		ELL		Total	
	N	Met %	N	Met %	N	Met %	N	Met %	N	Met %	N	Met %	N	Met %	N	Met %	N	Met %	N	Met %
Beall	57	91.2	44	93.2	17	94.1	31	93.5	9	66.7	43	95.3	24	83.3	8	75.0	13	69.2	101	92.1
Broad Acres	32	75.0	26	73.1	9	88.9	11	90.9	38	65.8	0	N/A	50	70.0	4	N/A	37	62.2	58	74.1
Brooke Grove	37	75.7	30	83.3	20	60.0	7	100.0	5	100.0	35	82.9	14	57.1	14	7.1	11	81.8	67	79.1
Brookhaven	26	84.6	28	78.6	14	92.9	5	80.0	29	72.4	6	100.0	31	74.2	7	28.6	29	69.0	54	81.5
Brown Station	31	58.1	28	78.6	30	63.3	7	71.4	12	75.0	10	70.0	33	63.6	3	N/A	13	46.2	59	67.8
Burnt Mills	26	88.5	21	95.2	34	91.2	1	N/A	8	100.0	4	N/A	23	91.3	4	N/A	14	92.9	47	91.5
Cannon Road	30	83.3	31	93.5	24	83.3	16	100.0	18	83.3	3	N/A	38	84.2	11	81.8	19	94.7	61	88.5
Dr. Charles R. Drew	18	83.3	21	90.5	23	78.3	7	100.0	4	N/A	5	100.0	17	88.2	6	50.0	9	88.9	39	87.2
East Silver Spring	36	75.0	31	74.2	34	67.6	7	100.0	16	68.8	10	90.0	32	75.0	11	36.4	17	58.8	67	74.6
Fairland	52	76.9	41	95.1	54	81.5	13	100.0	11	81.8	15	86.7	37	75.7	3	N/A	26	80.8	93	84.9
Gaithersburg	41	53.7	27	70.4	23	60.9	2	N/A	34	50.0	9	88.9	54	53.7	5	20.0	24	50.0	68	60.3
Galway	57	73.7	55	89.1	69	81.2	19	84.2	14	71.4	10	90.0	46	80.4	14	14.3	30	73.3	112	81.3
Glenallan	37	86.5	27	92.6	26	96.2	8	100.0	22	77.3	8	87.5	31	83.9	7	14.3	29	82.8	64	89.1
Greencastle	30	90.0	54	92.6	60	93.3	10	100.0	9	77.8	4	N/A	42	92.9	2	N/A	10	90.0	84	91.7
Harmony Hills	34	64.7	35	77.1	24	79.2	3	N/A	37	64.9	5	80.0	57	64.9	5	40.0	38	60.5	69	71.0
Highland View	31	87.1	20	85.0	9	88.9	4	N/A	13	84.6	25	84.0	23	82.6	7	57.1	16	81.3	51	86.3
Jackson Road	36	72.2	37	78.4	35	71.4	10	90.0	17	64.7	11	90.9	35	65.7	3	N/A	22	77.3	73	75.3
Kemp Mill	44	61.4	43	69.8	32	65.6	6	66.7	36	55.6	13	92.3	46	54.3	12	41.7	37	56.8	87	65.5
Montgomery Knolls	48	81.3	44	93.2	30	83.3	22	95.5	26	76.9	14	100.0	58	84.5	5	60.0	34	79.4	92	87.0
New Hampshire Estates	40	80.0	45	86.7	17	82.4	8	87.5	51	82.4	9	88.9	59	78.0	6	33.3	45	80.0	85	83.5
Rolling Terrace	54	88.9	58	89.7	27	92.6	6	66.7	59	86.4	19	100.0	67	85.1	2	N/A	47	83.0	112	89.3
South Lake	41	70.7	38	60.5	25	72.0	15	93.3	30	46.7	9	66.7	53	58.5	6	83.3	26	42.3	79	65.8
Strawberry Knoll	35	82.9	31	83.9	21	76.2	8	100.0	19	73.7	18	94.4	23	78.3	6	66.7	18	88.9	66	83.3
Takoma Park	71	83.1	77	89.6	52	80.8	12	91.7	15	66.7	69	94.2	37	62.2	4	N/A	20	50.0	148	86.5
Twinbrook	36	86.1	41	90.2	10	90.0	11	100.0	33	78.8	21	95.2	41	85.4	12	75.0	31	74.2	77	88.3
Viers Mill	34	70.6	40	62.5	10	60.0	12	83.3	36	55.6	15	80.0	47	55.3	6	33.3	26	42.3	74	66.2
Washington Grove	19	68.4	32	78.1	8	62.5	6	100.0	23	56.5	13	100.0	24	54.2	2	N/A	21	52.4	51	74.5
Watkins Mill	55	85.5	41	92.7	33	81.8	18	94.4	33	87.9	12	100.0	51	80.4	6	50.0	45	88.9	96	88.5
Weller Road	39	64.1	24	87.5	6	100.0	10	70.0	38	63.2	9	100.0	37	64.9	2	N/A	43	65.1	63	73.0
Woodlin	36	97.2	39	100.0	24	95.8	11	100.0	7	100.0	33	100.0	12	100.0	5	100.0	9	100.0	75	98.7

Table H2

Number and Percentage Meeting Grade 2 End-of-Year Benchmark for 2007 by Elementary School and Implementation Status

School	Male		Female		African American		Asian American		Hispanic		White		FARMS		SP ED		ELL		Total		
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	
	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	
Beall	41	43.9	49	55.1	18	27.8	27	59.3	14	28.6	31	64.5	28	28.6	10	0.0	11	27.3	90	50.0	
Broad Acres	32	68.8	31	54.8	16	62.5	5	60.0	42	61.9	0	N/A	56	58.9	1	N/A	28	46.4	63	61.9	
Brooke Grove	34	64.7	31	48.4	15	46.7	10	60.0	9	55.6	31	61.3	12	16.7	11	27.3	8	50.0	65	56.9	
Brookhaven	26	42.3	30	83.3	29	79.3	3	N/A	22	45.5	2	N/A	37	62.2	10	10.0	26	46.2	56	64.3	
Brown Station	23	65.2	29	69.0	23	69.6	10	80.0	14	50.0	5	80.0	27	63.0	3	N/A	10	30.0	52	67.3	
Burnt Mills	29	55.2	24	58.3	38	55.3	2	N/A	9	44.4	4	N/A	30	50.0	5	0.0	11	9.1	53	56.6	
Cannon Road	34	64.7	29	69.0	24	62.5	8	100.0	16	37.5	15	86.7	28	46.4	10	40.0	15	20.0	63	66.7	
Dr. Charles R. Drew	31	32.3	31	48.4	29	31.0	9	44.4	9	33.3	15	60.0	21	28.6	11	18.2	4	N/A	62	40.3	
East Silver Spring	32	78.1	28	82.1	37	83.8	9	77.8	6	33.3	8	100.0	32	68.8	5	20.0	9	33.3	60	80.0	
Fairland	37	18.9	41	26.8	41	19.5	10	40.0	13	15.4	13	23.1	34	23.5	7	14.3	18	5.6	78	23.1	
Gaithersburg	41	24.4	31	45.2	25	28.0	4	N/A	35	34.3	8	50.0	49	30.6	8	12.5	26	11.5	72	33.3	
Galway	61	65.6	51	76.5	67	76.1	15	60.0	17	58.8	12	66.7	43	62.8	14	21.4	24	41.7	112	70.5	
Glenallan	32	43.8	23	60.9	17	58.8	7	42.9	20	40.0	9	55.6	27	37.0	7	14.3	14	35.7	55	50.9	
Greencastle	40	37.5	52	53.8	65	46.2	9	55.6	12	58.3	5	20.0	37	32.4	12	8.3	5	20.0	92	46.7	
Harmony Hills	45	68.9	36	61.1	30	70.0	6	100.0	37	59.5	8	50.0	62	58.1	10	50.0	25	36.0	81	65.4	
Highland View	28	53.6	22	86.4	12	50.0	3	N/A	17	64.7	18	77.8	19	57.9	1	N/A	10	30.0	50	68.0	
Jackson Road	38	47.4	37	62.2	34	55.9	6	66.7	25	48.0	10	60.0	45	48.9	3	N/A	18	27.8	75	54.7	
Kemp Mill	46	26.1	39	38.5	33	24.2	8	37.5	35	22.9	9	88.9	53	22.6	6	16.7	31	9.7	85	31.8	
Montgomery Knolls	44	65.9	39	76.9	39	66.7	8	100.0	26	61.5	10	90.0	45	55.6	4	N/A	20	40.0	83	71.1	
New Hampshire																					
Estates	44	70.5	47	74.5	20	85.0	12	75.0	55	67.3	4	N/A	66	68.2	6	33.3	38	55.3	91	72.5	
Rolling Terrace	36	72.2	52	80.8	21	76.2	3	N/A	40	67.5	24	91.7	41	58.5	4	N/A	17	29.4	88	77.3	
South Lake	42	47.6	37	75.7	33	54.5	8	75.0	30	53.3	7	100.0	54	57.4	8	12.5	24	37.5	79	60.8	
Strawberry Knoll	28	64.3	46	71.7	28	71.4	11	81.8	19	57.9	16	68.8	32	62.5	3	N/A	14	35.7	74	68.9	
Takoma Park	78	82.1	60	76.7	47	66.0	4	N/A	26	61.5	61	98.4	40	57.5	10	40.0	26	38.5	138	79.7	
Twinbrook	32	37.5	35	62.9	10	60.0	11	45.5	29	41.4	15	66.7	35	45.7	8	12.5	24	29.2	67	50.7	
Viers Mill	28	39.3	25	80.0	7	57.1	7	85.7	29	48.3	8	62.5	30	46.7	10	20.0	16	37.5	53	58.5	
Washington Grove	19	68.4	26	50.0	8	75.0	7	71.4	19	42.1	11	63.6	24	45.8	5	20.0	23	39.1	45	57.8	
Watkins Mill	33	48.5	40	55.0	36	58.3	8	75.0	19	36.8	10	40.0	38	52.6	2	N/A	14	21.4	73	52.1	
Weller Road	34	73.5	29	75.9	10	50.0	7	85.7	39	74.4	7	100.0	44	68.2	4	N/A	31	64.5	63	74.6	
Woodlin	31	80.6	35	80.0	23	65.2	3	N/A	8	62.5	32	93.8	15	66.7	10	40.0	8	62.5	66	80.3	

Table H3
Number and Percent Meeting Grade 1 End-of-Year Benchmark for 2006 by Elementary School and Implementation Status

School	Male		Female		African American		Asian American		Hispanic		White		FARMS		SP ED		ELL	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Beall	46	73.9	45	93.3	21	76.2	23	100.0	13	53.8	34	88.2	28	75.0	10	40.0	17	70.6
Broad Acres	41	53.7	41	48.8	12	58.3	8	62.5	62	48.4	n/r	n/r	76	53.9	5	20.0	43	37.2
Brooke Grove	33	72.7	31	90.3	13	84.6	10	80.0	9	77.8	32	81.3	9	44.4	12	41.7	10	70.0
Brookhaven	20	65.0	26	80.8	24	83.3	n/r	n/r	17	64.7	n/r	n/r	32	65.6	8	37.5	28	60.7
Brown Station	25	52.0	37	64.9	29	58.6	9	77.8	17	41.2	7	85.7	26	46.2	5	0.0	17	35.3
Burnt Mills	37	62.2	37	81.1	51	72.5	n/r	n/r	16	62.5	n/r	n/r	39	66.7	5	80.0	16	56.3
Cannon Road	29	79.3	28	60.7	24	58.3	8	87.5	12	66.7	13	84.6	19	57.9	12	41.7	21	47.6
Dr. Charles R. Drew	27	63.0	30	70.0	25	56.0	8	87.5	10	70.0	13	76.9	18	50.0	10	30.0	7	57.1
East Silver Spring	43	67.4	28	78.6	41	70.7	7	85.7	9	55.6	12	83.3	38	60.5	7	14.3	12	25.0
Fairland	36	91.7	44	93.2	44	90.9	13	92.3	7	85.7	15	100.0	33	90.9	5	100.0	18	83.3
Gaithersburg	30	60.0	36	77.8	19	89.5	n/r	n/r	42	61.9	n/r	n/r	51	62.7	5	40.0	25	40.0
Galway	66	62.1	56	80.4	72	73.6	22	54.5	18	66.7	10	90.0	40	55.0	12	50.0	40	50.0
Glenallan	31	80.6	23	87.0	18	77.8	n/r	n/r	20	80.0	10	90.0	30	80.0	n/r	n/r	22	77.3
Greencastle	42	81.0	64	81.3	77	79.2	12	83.3	10	80.0	6	100.0	47	68.1	11	54.5	16	81.3
Harmony Hills	50	54.0	44	54.5	30	56.7	9	88.9	45	44.4	10	60.0	62	40.3	14	42.9	34	41.2
Highland View	32	65.6	24	75.0	17	64.7	n/r	n/r	20	60.0	17	88.2	22	54.5	n/r	n/r	13	30.8
Jackson Road	39	59.0	37	70.3	33	63.6	9	55.6	21	57.1	13	84.6	47	53.2	n/r	n/r	18	38.9
Kemp Mill	42	57.1	48	68.8	32	75.0	9	88.9	38	42.1	11	81.8	53	52.8	5	20.0	46	47.8
New Hampshire Estates	51	64.7	55	83.6	21	81.0	11	81.8	69	71.0	5	80.0	77	71.4	11	36.4	61	65.6
Rolling Terrace	40	87.5	59	93.2	22	86.4	n/r	n/r	49	89.8	24	100.0	51	84.3	6	100.0	32	87.5
South Lake	50	34.0	40	40.0	35	34.3	10	70.0	37	29.7	7	28.6	51	27.5	8	25.0	35	28.6
Strawberry Knoll	29	69.0	42	83.3	27	74.1	13	69.2	13	76.9	18	88.9	24	62.5	7	57.1	22	68.2
Takoma Park	82	87.8	63	82.5	47	78.7	7	85.7	31	74.2	60	96.7	40	70.0	12	66.7	28	67.9
Twinbrook	37	89.2	36	83.3	13	84.6	14	92.9	29	79.3	14	100.0	42	83.3	6	16.7	31	80.6
Viers Mill	52	40.4	42	61.9	23	87.0	9	55.6	49	34.7	12	33.3	59	44.1	8	12.5	45	28.9
Washington Grove	22	72.7	28	71.4	10	90.0	7	57.1	22	63.6	11	81.8	25	60.0	6	16.7	27	59.3
Watkins Mill	42	78.6	42	85.7	36	83.3	9	88.9	25	80.0	14	78.6	37	83.8	6	33.3	20	75.0
Weller Road	45	51.1	36	72.2	15	46.7	6	66.7	53	60.4	7	85.7	54	53.7	5	40.0	46	54.3
Woodlin	33	72.7	40	82.5	24	58.3	5	80.0	9	88.9	33	90.9	21	66.7	6	50.0	13	76.9

Table H4

Number and Percentage Meeting Grade 2 End-of-Year Benchmark for 2006 by Elementary School and Implementation Status

School	Male		Female		African American		Asian American		Hispanic		White		FARMS		SPED		ELL	
	%		%		%		%		%		%		%		%		%	
	N	Met	N	Met	N	Met	N	Met	N	Met	N	Met	N	Met	N	Met	N	Met
Beall	48	35.4	58	43.1	17	29.4	19	47.4	13	30.8	57	42.1	27	22.2	10	0.0	12	16.7
Broad Acres	30	13.3	39	12.8	21	14.3	8	50.0	40	5.0	n/r	n/r	62	12.9	n/r	n/r	23	0.0
Brooke Grove	26	26.9	29	31.0	12	16.7	10	50.0	7	14.3	26	30.8	6	33.3	11	0.0	6	0.0
Brookhaven	29	37.9	26	65.4	23	52.2	7	57.1	24	45.8	n/r	n/r	30	50.0	9	11.1	18	38.9
Brown Station	25	36.0	29	44.8	20	35.0	7	71.4	18	27.8	9	55.6	25	36.0	n/r	n/r	10	20.0
Burnt Mills	42	50.0	38	36.8	55	49.1	n/r	n/r	19	21.1	n/r	n/r	38	36.8	6	16.7	11	18.2
Cannon Road	33	24.2	30	50.0	30	36.7	6	33.3	19	26.3	8	62.5	24	29.2	8	0.0	10	20.0
Dr. Charles R. Drew	24	41.7	37	32.4	29	48.3	15	40.0	6	0.0	11	18.2	19	21.1	11	18.2	7	0.0
East Silver Spring	32	50.0	26	53.8	34	64.7	n/r	n/r	14	21.4	7	57.1	26	34.6	5	0.0	7	0.0
Fairland	53	18.9	43	27.9	56	17.9	16	37.5	13	7.7	11	45.5	33	9.1	11	9.1	9	0.0
Gaithersburg	32	34.4	32	21.9	23	30.4	n/r	n/r	31	19.4	6	50.0	36	16.7	9	11.1	17	0.0
Galway	59	45.8	47	44.7	63	42.9	18	61.1	14	28.6	11	54.5	46	28.3	6	0.0	18	22.2
Glenallan	38	42.1	29	37.9	24	41.7	9	11.1	21	42.9	13	53.8	31	25.8	13	30.8	11	9.1
Greencastle	51	58.8	38	60.5	69	65.2	6	33.3	11	36.4	n/r	n/r	38	57.9	8	37.5	7	42.9
Harmony Hills	45	66.7	39	56.4	24	58.3	6	83.3	46	60.9	8	62.5	58	60.3	6	16.7	23	47.8
Highland View	28	25.0	21	38.1	13	23.1	n/r	n/r	18	11.1	15	66.7	22	13.6	6	16.7	12	8.3
Jackson Road	46	41.3	45	62.2	46	47.8	13	46.2	19	47.4	13	76.9	41	41.5	n/r	n/r	10	10.0
Kemp Mill	45	22.2	42	26.2	24	20.8	13	30.8	37	10.8	13	61.5	58	19.0	10	0.0	34	17.6
Montgomery Knolls	45	80.0	35	68.6	28	71.4	10	80.0	29	65.5	13	100.0	43	72.1	5	40.0	19	52.6
New Hampshire Estates	32	75.0	44	68.2	17	82.4	8	62.5	44	63.6	6	100.0	59	64.4	9	33.3	24	58.3
Rolling Terrace	44	47.7	53	60.4	15	33.3	8	87.5	47	40.4	27	81.5	49	24.5	15	33.3	19	15.8
South Lake	43	55.8	42	61.9	29	55.2	12	91.7	35	54.3	9	44.4	46	56.5	11	27.3	18	38.9
Strawberry Knoll	40	37.5	31	51.6	22	40.9	10	50.0	21	38.1	18	50.0	26	42.3	6	0.0	13	7.7
Takoma Park	76	82.9	57	80.7	48	79.2	10	80.0	24	70.8	51	90.2	41	70.7	13	69.2	12	33.3
Twinbrook	51	17.6	35	40.0	20	20.0	6	16.7	37	24.3	21	38.1	44	18.2	12	0.0	22	13.6
Viers Mill	43	60.5	52	51.9	26	42.3	13	61.5	50	58.0	6	83.3	63	47.6	9	22.2	33	42.4
Washington Grove	36	33.3	32	37.5	11	36.4	11	36.4	32	21.9	14	64.3	33	30.3	n/r	n/r	29	13.8
Watkins Mill	55	25.5	53	24.5	39	23.1	n/r	n/r	37	16.2	26	38.5	51	15.7	9	0.0	23	26.1
Weller Road	50	54.0	38	68.4	14	78.6	10	100.0	54	46.3	10	70.0	55	49.1	7	57.1	29	24.1
Woodlin	51	66.7	40	62.5	29	62.1	7	42.9	8	75.0	46	67.4	22	50.0	13	15.4	5	60.0

Table H5

Number and Percentage at or Above the 50th National Curve Equivalent on TN/2 Assessments by Elementary School and Year

School	2004 ^a		2005 ^a		2006		2007	
	<i>N</i>	% Met	<i>N</i>	% Met	<i>N</i>	% Met	<i>N</i>	% Met
Beall	72	69.4	94	68.1	101	64.4	89	71.9
Broad Acres	70	57.1	75	49.3	56	30.4	59	45.8
Brooke Grove	87	71.3	74	68.9	54	63.0	66	75.8
Brookhaven	59	54.2	64	42.2	49	59.2	55	34.5
Brown Station	49	65.3	54	61.1	48	50.0	52	53.8
Burnt Mills	66	45.5	87	47.1	72	45.8	50	48.0
Cannon Road	53	67.9	57	70.2	59	37.3	61	39.3
Dr. Charles R. Drew	73	57.5	53	69.8	59	54.2	62	51.6
East Silver Spring	67	44.8	72	41.7	53	60.4	61	49.2
Fairland	79	55.7	80	52.5	89	51.7	77	51.9
Gaithersburg	95	55.8	85	58.8	56	53.6	67	40.3
Galway	100	63.0	106	64.2	100	49.0	110	44.5
Glenallan	64	59.4	71	62.0	61	42.6	51	66.7
Greencastle	131	44.3	96	53.1	84	54.8	92	37.0
Harmony Hills	69	49.3	72	43.1	76	53.9	79	41.8
Highland View	50	64.0	44	68.2	47	51.1	47	57.4
Jackson Road	68	47.1	64	60.9	91	47.3	73	46.6
Kemp Mill	100	59.0	95	57.9	80	46.3	82	45.1
Montgomery Knolls	93	41.9	103	43.7	76	42.1	83	50.6
New Hampshire Estates	87	40.2	81	48.1	74	50.0	88	40.9
Rolling Terrace	104	53.8	122	63.1	91	58.2	84	64.3
South Lake	91	56.0	78	57.7	75	42.7	77	35.1
Strawberry Knoll	96	52.1	70	55.7	66	50.0	74	50.0
Takoma Park	118	76.3	119	72.3	127	70.1	136	66.9
Twinbrook	87	55.2	76	52.6	82	40.2	64	42.2
Viers Mill	103	62.1	111	55.9	93	54.8	52	50.0
Washington Grove	63	69.8	57	73.7	63	60.3	43	58.1
Watkins Mill	97	58.8	104	73.1	105	50.5	75	60.0
Weller Road	95	44.2	88	61.4	82	48.8	62	56.5
Woodlin	68	80.9	84	71.4	89	66.3	66	68.2

^aTerraNova Comprehensive Test of Basic Skills (CTBS) was administered.

Table H6
Number and Percentage of Student Proficient and Above on Reading MSA by Elementary School and Test Year

School	2004		2005		2006	
	N	% Met	N	% Met	N	% Met
Beall	90	84.4	70	91.4	60	88.1
Broad Acres	80	75.0	70	64.3	46	82.0
Brooke Grove	103	76.7	87	79.3	50	55.3
Brookhaven	38	68.4	55	70.9	31	65.0
Brown Station	65	63.1	46	82.6	34	97.6
Burnt Mills	101	66.3	53	60.4	54	75.7
Cannon Road	70	71.4	55	78.2	34	73.1
Dr. Charles R. Drew	57	80.7	76	72.4	38	62.4
East Silver Spring	79	63.3	a	a	a	a
Fairland	81	67.9	89	66.3	48	85.2
Gaithersburg	66	66.7	82	74.4	49	63.2
Galway	131	65.6	100	78.0	63	94.8
Glenallan	64	73.4	67	77.6	37	84.3
Greencastle	85	60.0	119	61.3	59	80.9
Harmony Hills	61	57.4	67	67.2	42	77.6
Highland View	50	86.0	55	61.8	29	85.5
Jackson Road	76	71.1	83	54.2	47	87.5
Kemp Mill	99	58.6	92	70.7	64	94.9
Montgomery Knolls	a	a	a	a	a	a*
New Hampshire Estates	a	a	a	a	a	a
Rolling Terrace	112	69.6	106	67.9	65	88.4
South Lake	84	69.0	83	60.2	49	86.1
Strawberry Knoll	93	73.1	97	71.1	53	70.0
Takoma Park	a	a	a	a	a	a
Twinbrook	90	73.3	74	77.0	49	88.8
Viers Mill	101	72.3	91	90.1	88	91.2
Washington Grove	57	64.9	63	82.5	39	80.0
Watkins Mill	89	71.9	86	79.1	57	95.9
Weller Road	85	77.6	96	70.8	52	96.0
Woodlin	73	78.1	60	85.0	43	100.0

^a Primary schools (K-2)