IDEC Evaluation Report 2014–2015

Annual Results Confirm Strength of Reading Recovery

Jerome V. D'Agostino and Katherine Brownfield, International Data Evaluation Center

The 2014–2015 school year represents the beginning of the fourth decade of data collection and evaluation of Reading Recovery® in the United States. For Descubriendo la Lectura, 2014–2015 represents the 22nd year that it has been evaluated by IDEC. The school year also was critical for both interventions in that it was the last cohort of i3-funded teachers and schools. Over the 5-year grant period, university training centers recruited more urban high-need schools, more rural schools, and schools with large proportions of English language learners. About half of the active teachers in Reading Recovery and Descubriendo la Lectura were recruited with the support of the i3 grant, which means that not only are there more Reading Recovery and Descubriendo la Lectura teachers and schools that serve at-risk students, but there are more teachers with less years of experience offering the interventions nationwide.

One may suspect that a greater proportion of new Reading Recovery and Descubriendo la Lectura teachers serving greater proportions of at-risk students may lower the outcomes for Reading Recovery and Descubriendo la Lectura compared to prior years. The 2014–2015 outcomes, however, do not support such hypotheses — the results were maintained even with

the demographic changes, revealing the strength of Reading Recovery and Descubriendo la Lectura in getting students back on track toward successful literacy learning.

Summary of Reading Recovery Outcomes

Characteristics of participants

Reading Recovery was implemented by 19 university training centers in schools located in 42 states nationwide (see Table 1). There were over 46,000 children who were selected and participated in the one-to-one intervention. The 5,875 teachers trained in Reading Recovery also on

Table 1. Participation in Reading Recovery in the United States 2014–2015	
Entity	**n**
University Training Centers	19
Teacher Training Sites	243
States and Federal Entities*	42
Districts	1,205
Schools	3,735
Teacher Leaders	298
Teachers	5,875
Reading Recovery Students	46,849
Random Sample for RR	3,118
Tested-Not-Instructed for RR	6,175

*including Bureau of Indian Affairs, Department of Defense Domestic, and Department of Defense Overseas

average worked with an additional 40 students during the school year. These teachers were supported by 298 teacher leaders from 243 training sites that served just over 1,200 school districts. Reading Recovery was implemented in 3,735 schools, for an average of 1.60 teachers per building.

The Observation Survey was administered to Reading Recovery, random sample, and tested-not-instructed (TNI) students in fall, mid-year, and spring. As can be seen from Table 1, 3,118 random sample and 6,175 TNI students were tested.

Among the Reading Recovery participants from 2014–2015, 56% were boys and 69% were eligible for free or reduced lunch. Children were from a diversity of ethnic backgrounds, including 56% White, 17% African American, 19% Hispanic, 2% Asian American, 1% Native American, and 4% that represented multiple races or other ethnic backgrounds.

Among the Reading Recovery students:

- 17% (*n* = 8,107) were still in lessons at year-end without enough time in the school year to complete the intervention.

- 4% (*n* = 1,876) moved during the school year while they were enrolled in lessons.

- 3% (*n* = 1,344) were removed from the intervention by someone other than the Reading Recovery teacher.

Of the remaining students who had a complete intervention (*n* = 35,488):

- 72% (*n* = 25,718) reached average levels of reading and writing and their programs were successfully discontinued.

- 28% (*n* = 9,770) made progress but not sufficient enough to reach average levels of reading and writing. They were recommended for consideration of a more-intensive intervention.

Observation Survey results

The comparison groups, random sample and TNI, served to address two fundamental questions regarding the effectiveness of Reading Recovery. One key question is whether Reading Recovery students reach average levels of literacy achievement at the end of first grade relative to all other first-grade children who do not receive the intervention. The Observation Survey scores of all random sample students, including those that received Reading Recovery, were used to compute average achievement levels. A second key question relates to whether Reading Recovery students performed better than how they would have performed if not provided the intervention. TNI students' scores were used to address that research question.

The total score scale was created based on 2009–2010 random sample student data (including the random sample students who received Reading Recovery). Students' Observation Survey scores on all six subtests

from fall, mid-year, and spring were used to create the measure. Instead of using the Observation Survey scores of each student from the three time points, the random sample was divided into three randomly assigned groups. The fall, mid-year, or spring Observation Survey scores were chosen from each group, respectively, to represent a sample of students from the three time points during the school year. The six Observation Survey subtasks were treated as partial credit "items" in a Rasch-based IRT analysis to convert the total raw scores to log odd values that ranged from about -4 to 4. Those values were converted using a linear transformation to create the final 0 to 800-point scale. Because student scores were from various test points during the school year, the scale reflects yearlong growth. Hence, a score such as 500 indicates the same literacy achievement level at any time point.

Figure 1 presents the mean total scores for successfully discontinued Reading Recovery students who were served first (fall entry) during the school year, Reading Recovery students served second (spring entry), random sample, and TNI students. Only students with valid scores at all three tests points were included in the analysis. As expected, the TNI group had a slightly larger fall mean score relative to fall and spring entry Reading Recovery students, but less than the random sample students. By mid-year, fall entry Reading Recovery students had a significantly greater mean gain than spring entry students, TNI, and random sample students. From mid-year to spring, the average growth rate of the Reading Recovery fall entry students was less than the average random sample growth rate over the same period, but the two groups finished the year at about the same achievement level and both groups were considerably higher than TNI students.

Figure 1. Mean Observation Survey Total Score for Successfully Discontinued Reading Recovery (fall and spring entry), Random Sample, and Tested-Not-Instructed Students in the United States, 2014–2015
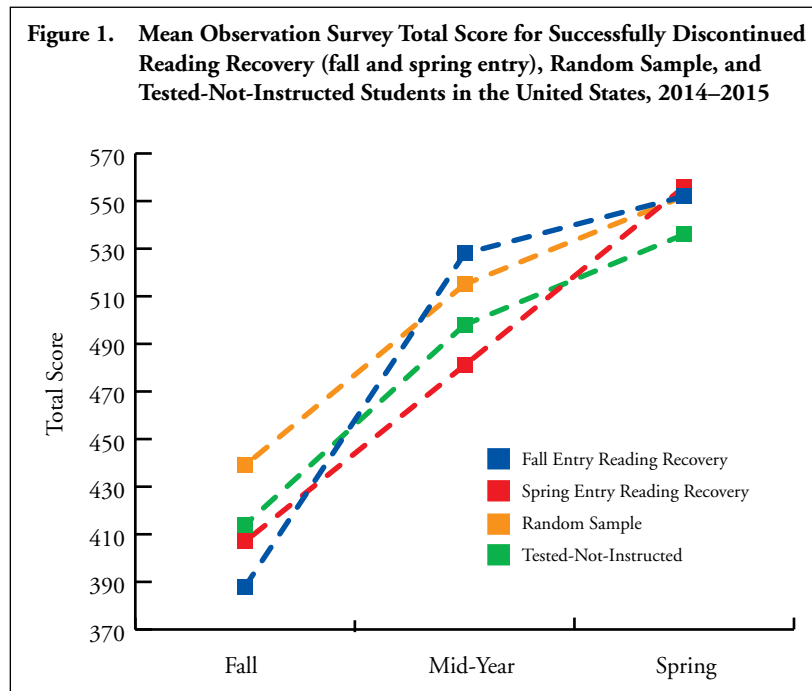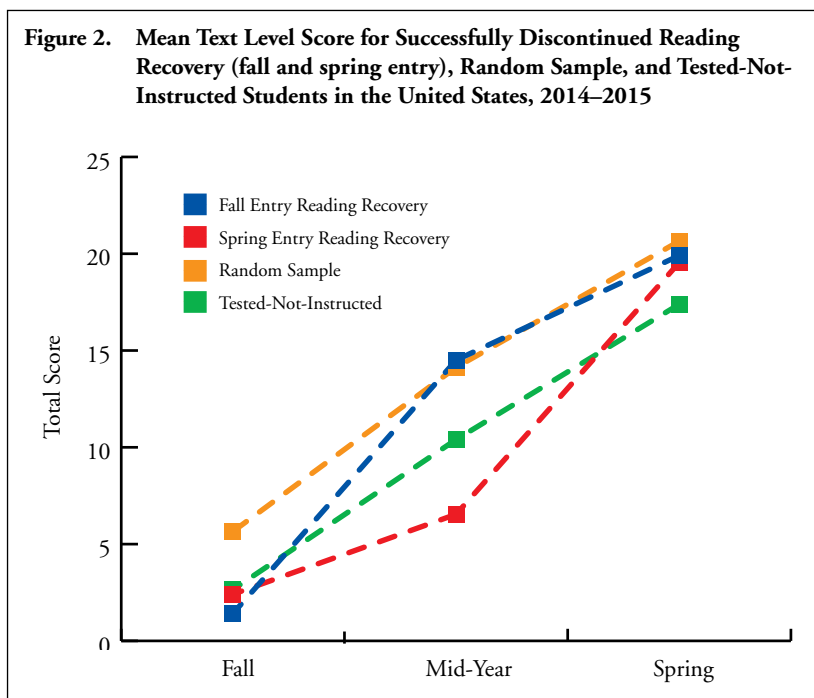
Note that spring entry students had a significantly smaller fall-to-mid-year mean gain than TNI students. This finding is critical to strengthen the inference that Reading Recovery is an effective intervention for three reasons. One, it may indicate that Reading Recovery teachers accurately identify and provide the treatment to the students most in need. On average, the students served in the second round are those who are falling behind the TNI group. Two, one possible explanation for the larger fall to mid-year gain for fall entry students is that their scores regressed more to the mean than TNI or random sample students. If that explanation were true, however, one would also expect the spring entry students to regress more toward the mean given their lower fall mean score. As can be seen from Figure 1, their growth rate in the first half of the year does not reflect greater regression. Three, spring entry students essentially serve as another (even more similar)

One key question is whether Reading Recovery students reach average levels of literacy achievement at the end of first grade relative to all other first-grade children who do not receive the intervention. The Observation Survey scores of all random sample students were used to compute average achievement levels.

comparison group for fall entry students at least in the first part of the year to address the question, "What would happen to the achievement levels of Reading Recovery students if they did not receive the treatment?" Clearly, the growth rate for fall entry students would be considerably lower without the treatment, as reflected in the spring entry student fall to mid-year growth. During the time of their intervention in the second half of the year, spring entry students had the largest growth rate.

Figure 2 presents the same group comparison method at three time points during the year (fall, mid-year, spring) on Text Reading Level. The general trends depicted in Figure 2 were similar to those for the total score, except for spring testing, where it is evident that Reading Recovery students whose lessons were discontinued did not entirely close the achievement gap between themselves and random sample students.

The magnitude of mean differences (effect sizes) in fall and spring between Reading Recovery and random sample or TNI students was examined. Tables 2 and 3 present the mean total and Observation Survey task scores of fall entry and spring entry Reading Recovery students whose lessons were discontinued pooled together, and random sample and TNI students, respectively. In both tables, the right columns provide the effect sizes in terms of standardized mean differences (positive values indicate that the Reading Recovery mean was greater than the comparison mean value) and the percentile standing of the average Reading Recovery child in the comparison-group distribution (in parentheses). As expected, the mean Reading Recovery scores in fall

Figure 2. **Mean Text Level Score for Successfully Discontinued Reading Recovery (fall and spring entry), Random Sample, and Tested-Not-Instructed Students in the United States, 2014–2015**

Table 2. Mean Fall and Spring Total Scores with Effect Sizes for Successfully Discontinued Reading Recovery and Random Sample Students 2014–2015

Observation Survey Task	Discontinued (n = 18,158)		Random Sample (n = 2,756)		Effect Size Difference	
	Fall	Spring	Fall	Spring	Fall	Spring
Total Score	395.76	553.79	440.44	552.62	-.30 (38)	+.03 (51)
Text Reading Level	1.70	19.76	5.79	20.73	-.65 (26)	-.14 (44)
Writing Vocabulary	13.00	56.63	21.20	56.01	-.69 (25)	+.03 (51)
Hearing and Recording Sounds in Words	23.63	35.97	29.13	35.65	-.69 (25)	+.11 (54)
Letter Identification	49.42	53.49	51.03	53.42	-.30 (38)	+.04 (52)
Ohio Word Test	4.63	19.18	9.88	18.87	-.86 (19)	+.12 (55)
Concepts About Print	13.13	21.98	15.31	20.68	-.62 (27)	+.12 (55)

ranged from the 19th to 38th percentile, with the latter value likely due to an apparent ceiling effect of Letter Identification in the random sample. By year-end, the effect size differences decreased significantly, indicating the closing of the achievement gap.

On the total score, the average Reading Recovery student performed at a level slightly above that of the average random sample student, indicating not only a complete closure of the achievement gap, but that the typical Reading Recovery student surpassed the average of the random sample group. In 2013–2014, the average

Reading Recovery student performed at the 50th percentile in the random sample distribution, and in 2012–2013, the average Reading Recovery student scored at the 47th percentile of the random sample on the total score. Thus, the spring outcome for Reading Recovery students is improving over time, which is remarkable given the demographic and teacher changes over the i3 grant period.

Also by year-end, on Concepts About Print, Hearing and Recording Sounds in Words, Letter Identification, the Ohio Word Test, and Writing Vocabulary, the mean Reading Recovery

score was slightly higher than the average random sample value. On Text Reading Level, the average Reading Recovery student was at the 44th percentile, and on Writing Vocabulary the average Reading Recovery student was at the 51st percentile, an increase of one percentile point over the 2013–2014 school year. Positive changes over the two years, particularly on Writing Vocabulary, on those two measures contributed greatly to the Reading Recovery group surpassing the random sample group on total score achievement in 2014–2015.

Table 3. Mean Fall and Spring Total Scores with Effect Sizes for Successfully Discontinued Reading Recovery and Tested-Not-Instructed Students 2014–2015

Observation Survey Task	Discontinued (n = 18,158)		Tested-Not-Instructed (n = 5,586)		Effect Size Difference	
	Fall	Spring	Fall	Spring	Fall	Spring
Total Score	395.76	553.79	414.23	536.37	-.45 (32)	**+.41 (66)**
Text Reading Level	1.70	19.76	2.65	17.38	-.42 (34)	+.35 (64)
Writing Vocabulary	13.00	56.63	16.26	50.73	-.35 (36)	+.34 (63)
Hearing and Recording Sounds in Words	23.63	35.97	26.38	35.04	-.34 (37)	+.26 (60)
Letter Identification	49.42	53.49	50.32	53.23	-.17 (43)	+.14 (56)
Ohio Word Test	4.63	19.18	6.58	18.29	-.44 (33)	+.30 (62)
Concepts About Print	13.13	21.98	14.06	19.72	-.30 (38)	+.46 (68)

The fall and spring test scores for Reading Recovery discontinued students (fall and spring entry combined) and TNI children are provided in Table 3. In fall, the Reading Recovery total score mean was at the 32nd percentile in the fall TNI distribution, indicating the greater initial proficiency of the TNI group. The Reading Recovery and TNI students, on average, were the most comparable on Letter Identification, as indicated by the smallest fall effect size difference among the measures. By spring, Reading Recovery students outperformed the TNI students on all six tasks and the total score; in other words, Reading Recovery students started the year below the TNI group and surpassed them by the end of the year. The average Reading Recovery student scored at the 66th percentile in the TNI group distribution on the total score, reflecting a sizable end-of-year achievement gap in favor of Reading Recovery.

Summary of Descubriendo la Lectura Outcomes

Descubriendo la Lectura, the reconstruction of Reading Recovery in Spanish, is for first graders who receive their initial literacy instruction in Spanish. Table 4 provides basic descriptive information about Descubriendo la Lectura implementation in the U.S. During the 2014–2015 school year, 569 Descubriendo la Lectura children were taught by 82 teachers. The students were from 81 schools in 26 school districts located in 8 states. The teachers received professional development support from 28 teacher leaders. Fifty-seven percent of Descubriendo la Lectura students were boys, 98% were Hispanic, and 99% qualified for free or reduced lunch costs.

> On the total score, the average Reading Recovery student performed at a level slightly above that of the average random sample student, indicating not only a complete closure of the achievement gap, but that the typical Reading Recovery student surpassed the average of the random sample group.

Among all children served in Descubriendo la Lectura, 47% reached the average reading levels of their peers and their lessons were discontinued successfully. Another 29% were recommended for further evaluation, 2% moved, and 20% received incomplete interventions. Among the students who completed the intervention (discontinued and referred students), 62% were discontinued.

Two students per participating Descubriendo la Lectura school were administered the Instrumento de Observación in fall, mid-year, and at the end of year in half of the schools

Table 4.	Participation in Descubriendo la Lectura in the United States 2014–2015	
Entity		**n**
University Training Centers		6
Teacher Training Sites		23
States		8
Districts		26
Schools		81
Teacher Leaders		28
Teachers		82
DLL Students		569
Random Sample for DLL		256

assigned at random. Those students combined represented the random sample. Descubriendo la Lectura schools had collected TNI data in 2011–2012, but due to very small samples and thus uninterpretable average scores, IDEC decided to forgo Descubriendo la Lectura TNI testing.

Descubriendo la Lectura random sample students' score on the six tasks of the Instrumento de Observación across multiple years were combined as was done for Reading Recovery to create a 0 to 800-point total score measure that reflected literacy development throughout the school year. Note that although this measure was developed using the same methods, a score of the same value on each measure should not be interpreted to indicate the same degree of literacy achievement (the tests contain different items and were scaled on different random samples).

Figure 3 presents the mean scores for both fall entry and spring entry successfully discontinued students and all Descubriendo la Lectura random sample participants on the total score at each time point, and Figure 4 provides the average scores for the same groups at the same time points on text reading level. The trends for Descubriendo la Lectura on the total

score were similar to the Reading Recovery results presented in Figure 1 with some differences. Descubriendo la Lectura students had considerably lower total scores than random sample students, on average, in fall, but by the end of year, the two Descubriendo la Lectura groups surpassed the random sample.

The greatest growth of any group was fall entry Descubriendo la Lectura students from fall to mid-year, followed by spring entry Descubriendo la Lectura students from mid-year to spring, indicating that gain was greatest during the intervention periods. Spring entry Descubriendo la Lectura and random sample students gained about the same amount from fall to mid-year, but from mid-year to spring, the spring entry Descubriendo la Lectura students outgained the random sample, indicative of a predictable growth pattern during the treatment period. The trend for text level (Figure 4) was similar to the total score trend (Figure 3) except for one difference; spring entry Descubriendo la Lectura students did not, on average, make comparable fall to mid-year gains relative to the random sample. Instead, the spring entry Descubriendo la Lectura students had considerably lower growth rates in the first part of the year without the intervention. During the second part of the year, they caught the random sample and the Descubriendo la Lectura discontinued students who received the intervention in the fall. Therefore, both Descubriendo la Lectura groups started the school year behind the random sample but caught the comparison group by the end of the year.

Figure 3. **Mean Instrumento de Observación Total Score for Successfully Discontinued Descubriendo la Lectura (fall and spring entry), and Random Sample Students in the United States, 2014–2015**
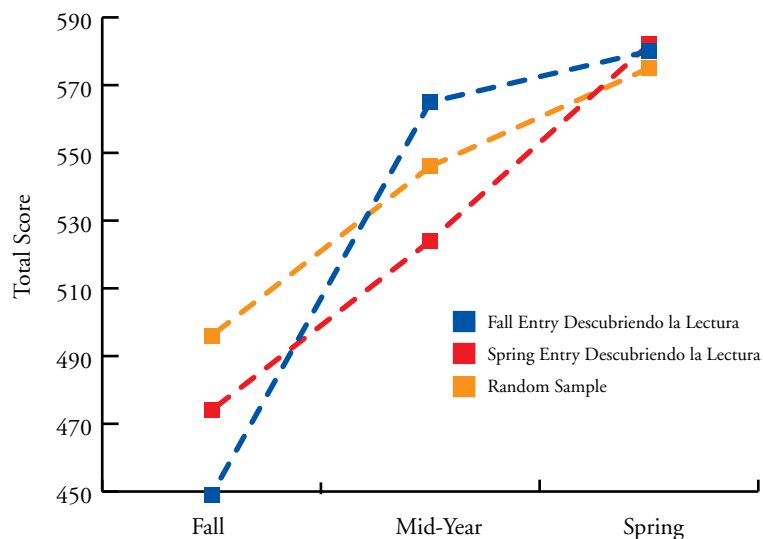
Figure 4. **Mean Análisis Actual del Texto Score for Successfully Discontinued Descubriendo la Lectura (fall and spring entry), and Random Sample Students in the United States, 2014–2015**
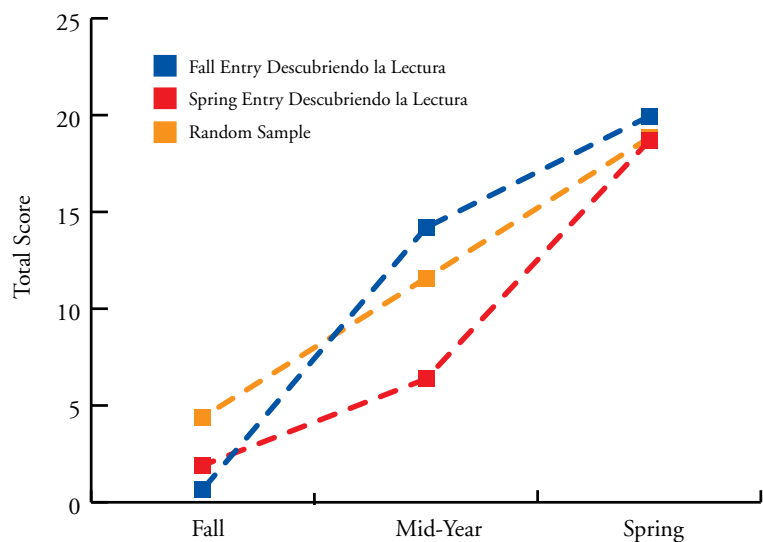
Table 5. Mean Fall and Spring Total Scores with Effect Sizes for Successfully Discontinued Descubriendo la Lectura (DLL) and DLL Random Sample Students 2014–2015

Instrumento de Observacíon Task	Discontinued (n = 221) Fall	Discontinued (n = 221) Spring	Random Sample (n = 265) Fall	Random Sample (n = 265) Spring	Effect Size Difference Fall	Effect Size Difference Spring
Total Score	458.44	581.10	495.48	574.88	-.95 (17)	+.21 (58)
Análisis Actual del Texto	1.14	19.59	4.49	18.78	-.70 (24)	+.12 (55)
Escritura de Vocabulario	10.47	49.60	18.13	47.94	-.72 (24)	+.10 (54)
Oír y Anotar los Sonidos en las Palabras	21.91	38.52	30.32	38.11	-.83 (20)	+.18 (57)
Identificacíon de Letras	43.72	58.63	52.01	58.49	-.93 (18)	+.04 (52)
Prueba de Palabras	6.90	19.62	12.60	19.08	-.84 (20)	+.22 (59)
Conceptos del Texto Impreso	9.69	19.95	12.48	19.03	-.73 (23)	+.30 (62)

Table 5 consists of the mean scores and effect sizes for fall and spring entry Descubriendo la Lectura discontinued students combined and random sample students in fall and at the end of year. It can be seen from the table that the average discontinued Descubriendo la Lectura student performed at the 58th percentile of random sample students on the total test in spring. Discontinued Descubriendo la Lectura students equaled or outperformed the random sample on all of the Instrumento de Observación tasks in spring. These average score differences reveal strong effects for Descubriendo la Lectura.

Conclusion

The list of educational interventions that have had the effect on student learning and program longevity in the United States compared to Reading Recovery and Descubriendo la Lectura is very small. In its 31st year of implementation during 2014–2015, students in the intervention posted perhaps the strongest outcomes experienced to date. On the total score for both Reading Recovery and Descubriendo la Lectura, the average discontinued student surpassed the average of the random sample.

These findings reflect the strong commitment of Reading Recovery and Descubriendo la Lectura trainers, teacher leaders, and teachers to persistently strive to improve their practices. Their hard work and engagement are paying off in terms of greater student literacy success.

About the Authors

Jerome D'Agostino is a professor in the Quantitative Research, Evaluation, and Measurement program at The Ohio State University and research director of the International Data Evaluation Center. He specializes in assessment, measurement, and intervention evaluation.
Dr. D'Agostino also served as director of the i3 project to scale up Reading Recovery.

Katherine Brownfield is a PhD candidate in the Reading and Literacy in Early and Middle Childhood program in The Ohio State University College of Education and Human Ecology. She currently serves as a graduate research associate for the International Data Evaluation Center. Her research interests center on teacher-student interactions that support literacy learning.