IDEC Evaluation Report 2013–2014

Annual Results Show Strongest Outcomes Experienced to Date

Jerome V. D'Agostino and Sinéad Harmey, International Data Evaluation Center

The 2013–2014 school year marked a special milestone for data collection and evaluation of Reading Recovery® in the United States. Beginning with the 1984–1985 pilot year in Columbus City Schools, the intervention reached its 30th year of implementation and evaluation in 2013–2014. With funding from the National Diffusion Network, IDEC (then NDEC) began operations with the evaluation of Reading Recovery data from the 1987–1988 school year, making 2013–2014 the 27th year of IDEC data analysis. Descubriendo la Lectura (DLL) has been evaluated for 21 years given that IDEC's first evaluation of the intervention was in 1993–1994. Over the years, many things have changed, including computer data entry, evaluation designs, and analytic methods, but one fact has remained constant; both interventions have consistently produced strong evidence of effectiveness. Students in Reading Recovery and DLL consistently made greater yearly gains relative to their respective random samples, and Reading Recovery students made gains relative to their respective similar comparison group, reflecting a substantial closing of the literacy gaps at year-end testing.

From 2005–2011, IDEC used a low-random sample, students who scored in the bottom 20% of the random sample distribution, to produce a "similar comparison group" in order to evaluate the progress of Reading Recovery and DLL students. That method, however, had certain methodological flaws. Because only two students per school were randomly chosen for the random sample—including students who received Reading Recovery—many schools did not have a student in the bottom 20%, and the schools that did were not representative of all Reading Recovery schools. Also, the distribution of outcome scores for the low-random sample was not similar to the Reading Recovery student distribution because it artificially had a ceiling at the 20th percentile.

To address these issues, IDEC randomly assigned half of the schools to employ the tested-not-instructed (TNI) approach beginning in the 2011–2012 school year. The TNI students represent a subgroup of children who are either eligible and would have been selected if teaching slots were available, or slightly more proficient than eligible Reading Recovery students. Thus, they are the closest students (achievement wise) at each school to the treatment children, and as such, they form the best possible comparison group at each school sampled to collect TNI data. The other half of the Reading Recovery and all DLL schools are asked to implement the extant method of testing two randomly chosen students (random sample children) in order to generate estimates of average student performance. The schools alternate between the TNI and random sample method each year. We report the general participation rates for Reading Recovery and DLL in 2013–2014, along with the basic outcomes such as students' end-of-treatment statuses and Observation Survey outcomes compared to TNI and random samples students.

> Perhaps the most-important finding from the 2013–2014 evaluation was that Reading Recovery students closed the achievement gap on two measures that have been historically difficult to change: Writing Vocabulary and Text Reading Level.

Summary of Reading Recovery Outcomes

Characteristics of participants

Reading Recovery was implemented by 19 university training centers in schools located in 41 states nationwide (see Table 1). There were over 47,000 children who were selected and participated in the one-to-one intervention. On average, the 5,982 teachers trained in Reading Recovery also worked with an additional 40 students during the school year. These teachers were supported by 305 teacher leaders from 252 training sites that served nearly 1,250 school districts. Reading Recovery was implemented in 3,736 schools, for an average of 1.60 teachers per building.

The Observation Survey was administered to Reading Recovery, random sample, and TNI students in fall, mid-year, and spring. As can be seen in Table 1, 3,213 random sample and 7,105 TNI students were tested.

Among the Reading Recovery participants from 2013–2014, 57% were boys, and 67% were eligible for free or reduced-price lunch. Children were from a diversity of ethnic backgrounds, including 57% White, 17% African American, 18% Hispanic, 2% Asian American, 1% Native American; and 4% that represented multiple races or other ethnic backgrounds.

Among the Reading Recovery students,

- 17% (n = 8,010) were still in lessons at year-end without enough time in the school year to complete the intervention.

- 4% (n = 1,817) moved during the school year while they were enrolled in lessons.

- 2% (n = 1,100) were removed from the intervention by someone other than the Reading Recovery teacher.

Of the remaining students who had a complete intervention (N = 36,338),

- 72% (n = 26,108) reached average levels of reading and writing and their programs were successfully discontinued.

- 28% (n = 10,229) 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, and 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.

Table 1.	Participation in Reading Recovery in the United States 2013–2014
Entity	**n**
University Training Centers	19
Teacher Training Sites	252
States and Federal Entities*	41
Districts	1,250
Schools	3,736
Teacher Leaders	305
Teachers	5,982
Reading Recovery Students	47,263
Random Sample for RR	3,213
Tested-Not-Instructed for RR	7,105

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

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.

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) 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

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, 2013–2014

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, 2013–2014
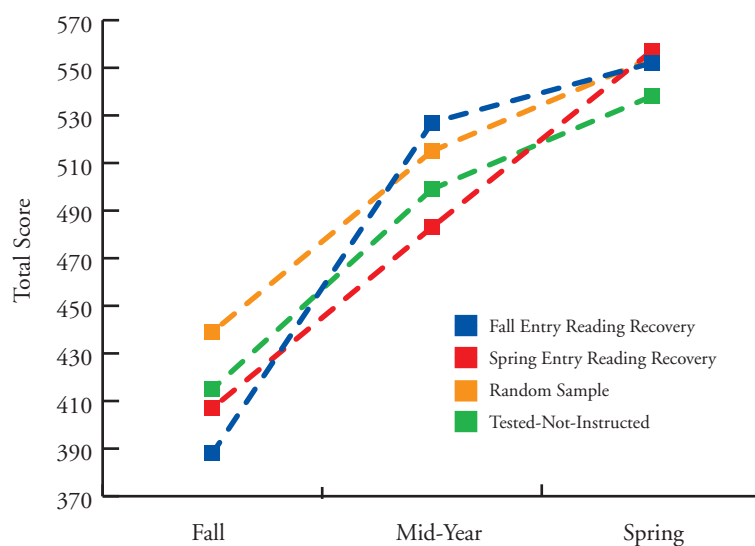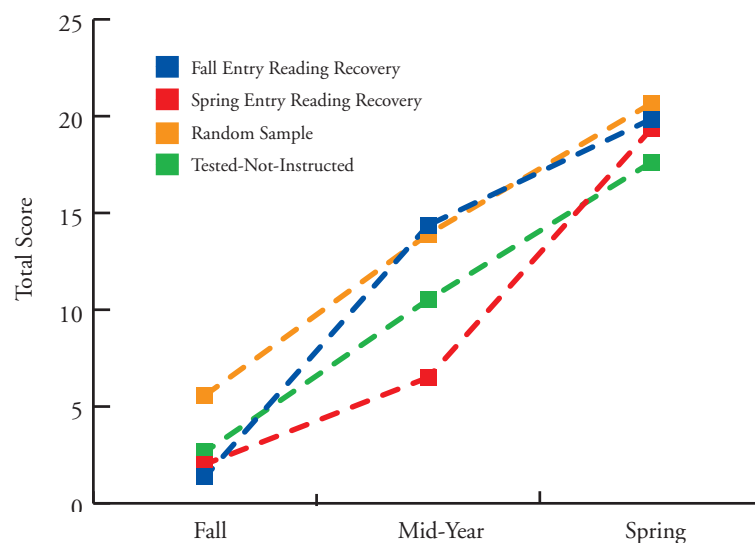
Table 2. Mean Fall and Spring Total Scores with Effect Sizes for Successfully Discontinued Reading Recovery and Random Sample Students 2013–2014

Observation Survey Task	Discontinued (n = 18,305)		Random Sample (n = 2,874)		Effect Size Difference	
	Fall	Spring	Fall	Spring	Fall	Spring
Total Score	**395.41**	**553.80**	**438.96**	**553.81**	**-.83 (21)**	**0.00 (50)**
Text Reading Level	1.68	19.71	5.58	20.66	-.78 (19)	-.14 (44)
Writing Vocabulary	12.84	56.65	21.09	56.65	-.67 (25)	0.00 (50)
Hearing and Recording Sounds in Words	23.42	35.98	28.78	35.72	-.68 (25)	+.09 (54)
Letter Identification	49.58	53.49	51.17	53.39	-.33 (37)	+.05 (52)
Ohio Word Test	4.56	19.17	9.60	18.92	-.82 (21)	+.11 (54)
Concepts About Print	13.52	21.03	15.36	20.78	-.61 (27)	+.09 (54)

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 (on the previous page) 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 discontinued students did not entirely close the achievement gap between themselves and random sample students (although the gap reduction improved slightly compared to 2012–2013, which is discussed further below).

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 discontinued

students pooled together and random sample and TNI students, respectively. In both tables, the right-hand 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 ranged from the 19th to 37th percentile, with the latter value likely due to an apparent ceiling effect of Letter Identification in the random sample.

Table 3. Mean Fall and Spring Total Scores with Effect Sizes for Successfully Discontinued Reading Recovery and Tested-Not-Instructed Students 2013–2014

Observation Survey Task	Discontinued (n = 18,305)		Tested-Not-Instructed (n = 6,322)		Effect Size Difference	
	Fall	Spring	Fall	Spring	Fall	Spring
Total Score	**395.41**	**553.80**	**415.28**	**538.43**	**-.52 (30)**	**+.38 (65)**
Text Reading Level	1.68	19.71	2.66	17.62	-.13 (45)	+.31 (62)
Writing Vocabulary	12.84	56.65	16.32	51.48	-.40 (34)	+.25 (60)
Hearing and Recording Sounds in Words	23.42	35.98	26.61	35.18	-.41 (34)	+.30 (62)
Letter Identification	49.58	53.49	50.60	53.27	-.23 (41)	+.12 (55)
Ohio Word Test	4.56	19.17	6.62	18.39	-.47 (32)	+.27 (61)
Concepts About Print	13.25	21.03	14.11	19.81	-.28 (39)	+.46 (68)

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 comparable to the 50th percentile random sample student, indicating a complete closure of the achievement gap. In 2012–2013, the average Reading Recovery student scored at the 47th percentile of the random sample on the total score, so the gap closed to a greater extent in 2013–2014. On Concepts About Print, Hearing and Recording Sounds in Words, Letter Identification and the Ohio Word Test, the mean Reading Recovery score was slightly larger than the average random sample value. On Writing Vocabulary and Text Reading Level, the average Reading Recovery students were at the 50th and 44th percentiles, respectively. The year prior (2012–2013), the average Reading Recovery student was at the 48th percentile for Writing Vocabulary, and the 43rd for Text Reading Level. Positive changes over the 2 years on those measures contributed greatly to the complete closure of the total score achievement gap in 2013–2014.

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 30th 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 Text Reading Level, 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. 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 65th 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 (DLL), 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 DLL implementation in the country. During the 2013–2014 school year, 632 DLL children were taught by 90 teachers. The students were from 85 schools in 25 school districts located in nine states. The teachers received professional development support from 29 teacher leaders. Of the total DLL students, 58% were boys, 99% were Hispanic, and 98% qualified for free or reduced lunch costs.

Table 4.	Participation in Descubriendo la Lectura in the United States 2013–2014	
Entity		**n**
University Training Centers		7
Teacher Training Sites		25
States		9
Districts		26
Schools		85
Teacher Leaders		29
Teachers		90
DLL Students		632
Random Sample for DLL		292

Among all children served in DLL, 49% reached the average reading levels of their peers and their lessons were discontinued successfully. Another 27% were recommended for further evaluation, 3% moved, and 19% received incomplete interventions. Among the students who completed the intervention (discontinued and referred), 65% were discontinued.

Two students per participating DLL school were administered the Instrumento de Observación in fall, mid-year, and at the end of year in half of the schools assigned at random. Those students combined represented the random sample. DLL schools had collected TNI data in 2011–2012, but due to very small samples and thus uninterpretable average scores, IDEC decided to forgo DLL TNI testing.

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

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, 2013–2014

The greatest growth of any group was fall entry DLL students from fall to mid-year, followed by spring entry DLL students from mid-year to spring, indicating that gain was greatest during the intervention periods. Spring entry DLL and random sample students gained about the same amount from fall to mid-year, but from mid-year to spring, the spring entry DLL 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 DLL students did not, on average, make comparable fall-to-mid-year gains relative to the random sample. Instead, the spring entry DLL 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 DLL discontinued students who received the intervention in the fall. Both DLL groups started the school year behind the random sample but caught up to them by the end of the year.

Table 5 consists of the mean scores and effect sizes for fall and spring entry DLL 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 DLL student performed at the 58th percentile of random sample students on the total test in spring. Discontinued DLL 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 DLL.

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, 2013–2014
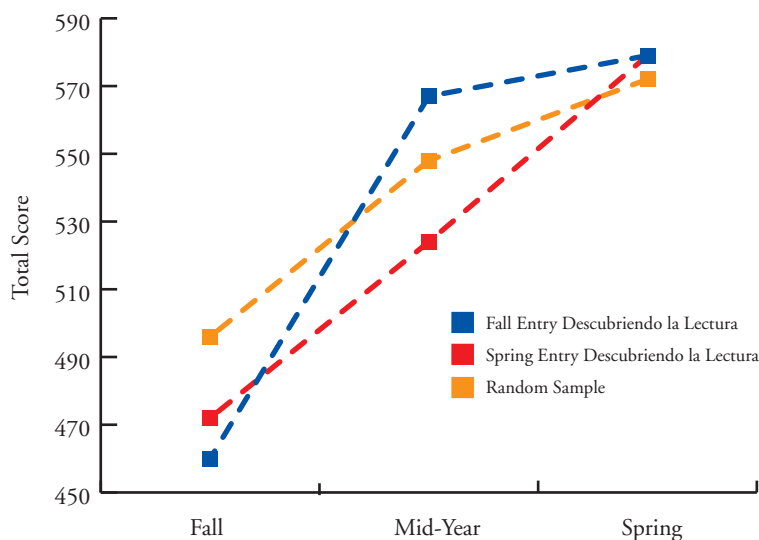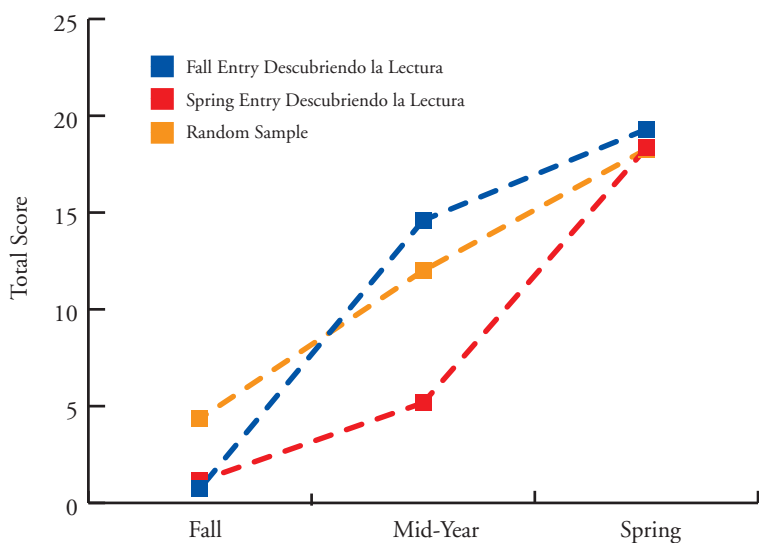
Table 5. **Mean Fall and Spring Total Scores with Effect Sizes for Successfully Discontinued Descubriendo la Lectura (DLL) and DLL Random Sample Students 2013–2014**

Instrumento de Observacíon Task	Discontinued (n = 221)		Random Sample (n = 265)		Effect Size Difference	
	Fall	Spring	Fall	Spring	Fall	Spring
Total Score	**465.10**	**579.65**	**496.45**	**572.88**	**-.82 (21)**	**+.21 (58)**
Análisis Actual del Texto	.98	19.04	4.36	18.27	-.71 (24)	+.11 (54)
Escritura de Vocabulario	11.01	49.74	18.65	46.62	-.69 (25)	+.18 (57)
Oír y Anotar los Sonidos en las Palabras	23.78	38.32	30.80	37.85	-.71 (24)	+.92 (82)
Identificacíon de Letras	46.44	58.93	51.85	58.50	-.61 (27)	+.14 (56)
Prueba de Palabras	7.37	19.70	12.51	19.09	-.75 (23)	+.25 (60)
Conceptos del Texto Impreso	10.61	19.36	13.12	19.30	-.66 (25)	+.02 (51)

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 30th year of implementation during 2013–2014, students in the intervention posted perhaps the strongest outcomes experienced to date. On the total score, the achievement gap was completely closed as indicated by the average discontinued Reading Recovery student being at the 50th percentile of the random sample in spring. In DLL, the average discontinued student surpassed the average of the random sample.

Perhaps the most-important finding from the 2013–2014 evaluation was that Reading Recovery students closed the achievement gap on two measures that have been historically difficult to change: Writing Vocabulary and Text Reading Level. For both measures, the average Reading Recovery student inched closer to catching the average random sample levels than had been seen in prior years. 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 is director of the i3 project to scale up Reading Recovery.

Sinéad Harmey is a doctoral student in reading and literacy in early and middle childhood in the School of Teaching and Learning at The Ohio State University. She is a graduate research associate for the International Data Evaluation Center and previously worked as a Reading Recovery teacher leader in Dublin, Ireland.