IDEC Evaluation Report 2016–2017

Results Again Show Strong Impact on Student Learning

Richard G. Lomax, The Ohio State University, International Data Evaluation Center

This report shares the results of Reading Recovery® and Descubriendo la Lectura in the United States for the 2016–2017 school year. As described herein, the interventions have continued to maintain strong outcomes, both in terms of progress across the length of the intervention and as contrasted against comparison groups. These results are also comparable to those of the 2015–2016 school year (Lomax, 2017).

Summary of Reading Recovery Outcomes

Characteristics of participants

During the 2016–2017 school year, Reading Recovery was implemented by 17 university training centers responsible for overseeing the intervention in schools located in 42 states (Table 1). More than 38,000 children were selected to participate in the one-to-one intervention. There were 4,863 teachers trained in Reading Recovery who delivered the intervention, with support from 274 teacher leaders in 220 training sites serving over 1,000 school districts. There were 3,190 schools implementing Reading Recovery.

The Observation Survey was administered to Reading Recovery students, a random sample of comparison students, and tested-not-instructed

(TNI) students at fall, mid-year, and year-end. As shown in Table 1, a total of 2,885 random sample students and 6,153 TNI students were tested.

The following demographics describe the Reading Recovery participants from 2016–2017. Of those students, 54% were boys, 70% were eligible for free or reduced lunch, and there were considerable Title I children (77% schoolwide Title I, 23% individual Title I). Children represented different ethnic backgrounds, including 57% White, 19% Hispanic, 16% African American, 2% Asian American, 1% Native American, and 5%

representing either multiple races or other ethnic backgrounds. School locations were in 27% urban, 33% suburban, and 40% rural areas.

For the Reading Recovery students:

- 17% (n = 6,518) were still in lessons at year-end without enough time in the school year to complete the intervention.

- 4% (n = 1,400) moved during the school year while still enrolled in lessons.

Of the remaining students who completed the intervention (n = 29,491):

- 72% (n = 21,165) reached at least average levels of reading and writing as their intervention programs were successfully discontinued.

- 28% (n = 8,326) made progress, but not at a sufficient enough level to reach average levels of reading and writing.

The latter students were recommended for consideration of additional intensive intervention. Most notable were 5,665 who were recommended for small-group literacy instruction or intervention other than special education, and 1,835 who were recommended for literacy-related special education services.

Table 1.	Participation in Reading Recovery in the United States, 2016–2017
Entity	**n**
University Training Centers	17
Teacher Training Sites	220
States	42
School Systems	1,038
School Buildings	3,190
Teacher Leaders	274
Teachers	4,863
Reading Recovery Students	38,448
Random Sample for RR	2,885
Tested-Not-Instructed for RR	6,153

NOTE: Some students in the Control Group of the random assignment study did not receive Reading Recovery. Their data are excluded from results in other tables in this report but included here.

The professional experience of the trained teachers consisted of the following:

- Mean 20.6 years of teaching experience

- Mean 8.8 years of Reading Recovery and/or Descubriendo la Lectura teaching experience

- Taught from 1 to 9 Reading Recovery children on a daily basis (mode = 4), while teaching a mean of 7.6 Reading Recovery children across the school year, and a mean of 39.9 children in other teaching roles, for a mean total of 47.5 children

Results

Two comparison groups were utilized—the random sample and the TNI sample—which helped to address two critical questions on the effectiveness of Reading Recovery. The first is whether Reading Recovery students reach average levels of literacy achievement at the end of first grade as compared to all other first-grade children who do not receive the intervention. Here the average Observation Survey scores of Reading Recovery students were compared against all random sample students (the Observation Survey being one metric for literacy achievement level). A second question is whether Reading Recovery students performed better at the end of the intervention than they would have performed if they were not provided the intervention. Here the average Observation Survey scores of Reading Recovery students were compared against the TNI students' scores.

The Total Score scale of the Observation Survey 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 tasks from fall, mid-year, and year-end were used to create the total measure. The six tasks are Text Reading Level, Writing Vocabulary, Hearing and Recording Sounds in Words, Letter Identification, Ohio Word Test, and Concepts About Print. 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 year-end Observation Survey scores were chosen from each group, respectively, to represent an independent sample of students from the three time points during the school year. The six tasks were treated as partial credit 'items' in a Rasch-based item response theory (IRT) analysis to convert the total raw scores to log-odd values ranging from approximately -4 to 4. Those values were then converted through a linear transformation to create the final 0 to 800-point scale. As student scores were from various test points during the school year, the scale reflects yearlong growth. Thus, for example, a Total Score of 500 indicates the same literacy achievement level at any time point. Additional details on the Observation Survey (e.g., scale construction, reliability and validity evidence, normality, equal interval scales, unidimensionality) are described in D'Agostino (2012) and D'Agostino, Rodgers, and Mauck (2017).

Figure 1 illustrates the mean Total Scores for successfully discontinued Reading Recovery students served first (fall entry) during the school year, Reading Recovery students served second (spring entry), random sample students, and 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, 2016–2017**
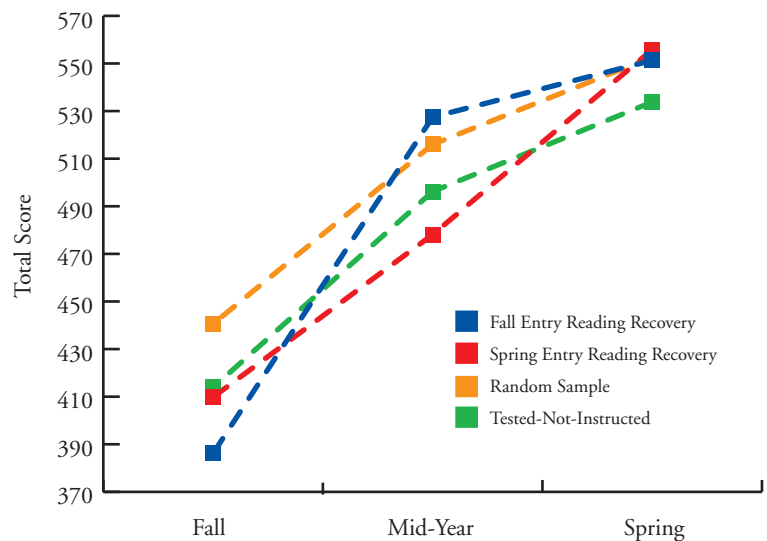
From mid-year to year-end, the average growth rate of the Reading Recovery fall entry students was slightly less than the average random sample growth rate over the same period, but the two groups finished the year at about the same achievement level and both groups were considerably higher than TNI students.

Only students with valid scores at all three tests points were included in the analysis. As in past years, the TNI group had a slightly higher fall mean score relative to fall and spring entry Reading Recovery students, but not as high as the random sample students.

Consider first the fall entry Reading Recovery students. By mid-year, these students had a greater mean gain than spring entry, TNI, and random sample students. Thus, the fall entry Reading Recovery students—whose mean Observation Survey score was the lowest of all groups—was the highest by mid-year. From mid-year to year-end, the average growth rate of the Reading Recovery fall entry students was slightly 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.

Consider next the spring entry Reading Recovery students. These students had a smaller fall-to-mid-year mean gain than TNI students. This was to be expected, as this group does not receive the intervention until the second half of the school year. Thus during the fall, the spring entry stu-

dents serve as an additional control. Once they begin their intervention in the second half of the year, spring entry students had the largest growth rate. In addition, the fall entry, spring entry, and random sample means were approximately the same at year-end testing, indicating that the Reading Recovery students had caught up to their random sample peers.

Figure 2 shows the results for the same four groups across the same three time points for Text Reading

Level. The general trend as shown in Figure 2 is quite similar to that for the Observation Survey Total Score. Note, however, that the Reading Recovery discontinued students (both fall and spring entry) at year-end testing had reached grade level and nearly achieved the text reading level of the random sample students.

The means and magnitude of mean differences (effect sizes) at fall and year-end testing between the Reading Recovery students and the random sample or TNI students were considered next. Tables 2 and 3 display the total and individual task scores of fall entry and spring entry Reading Recovery discontinued students pooled together as compared with the random sample and TNI students, respectively. For both tables, the far right-hand columns denote the effect sizes in terms of standardized mean differences (positive values indicate that the Reading Recovery mean was greater than the comparison mean

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, 2016–2017
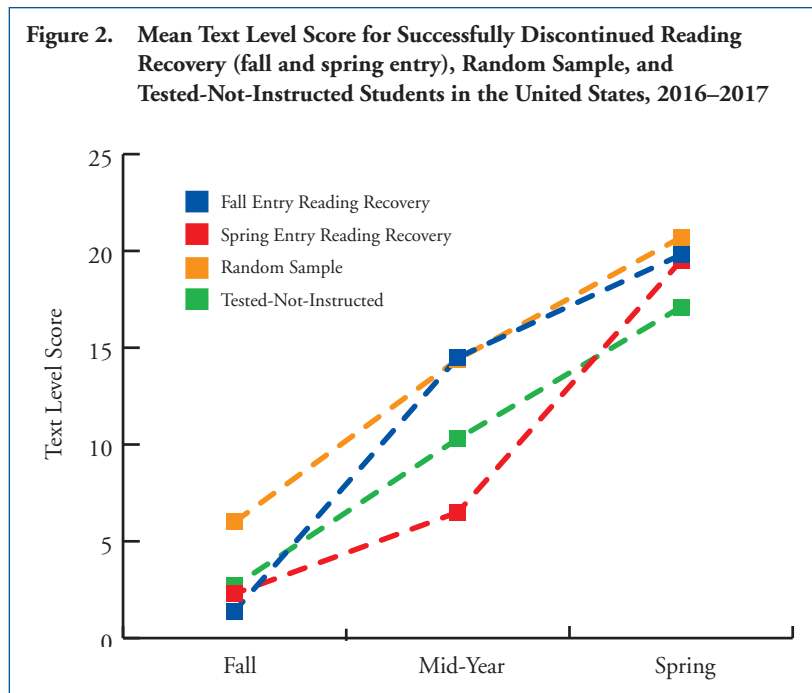
Table 2. Mean Fall and Year-End Total Scores with Effect Sizes for Successfully Discontinued Reading Recovery and Random Sample Students, 2016–2017

Observation Survey Task	Discontinued		Random Sample		Effect Size (*d*)	
	Fall	Year-End	Fall	Year-End	Fall	Year-End
Total Score	**395.5**	**552.9**	**441.8**	**552.8**	**-1.03**	**0.01**
Text Reading Level	1.7	19.6	6.0	20.7	-.91	-.19
Writing Vocabulary	12.8	56.5	21.1	56.2	-.81	0.02
Hearing and Recording Sounds in Words	23.7	36.0	29.0	35.7	-.65	0.13
Letter Identification	49.2	53.5	51.1	53.3	-.41	0.11
Ohio Word Test	4.6	19.2	10.1	18.9	-1.11	0.15
Concepts About Print	13.1	21.0	15.3	20.6	-.68	0.17

value). Note that the effect size measure utilized was Cohen's *d* (Cohen, 1988; Lomax & Hahs-Vaughn, 2012), which can be thought of in the metric of a standard deviation. Thus, a value of *d* = +1.00 would indicate that the Reading Recovery children had a mean score of one standard deviation above the comparison group. A common standard to judge *d* is that .2 is a small effect size, .5 a medium effect size, and .8 a large effect size.

During fall testing, mean Reading Recovery scores on all measures were substantially lower than the random sample, with medium to very large effect sizes (ranging from -.41 to -1.11). By year-end testing, there were relatively small effect sizes in favor of the Reading Recovery students (ranging from .01 to .17), except for Text Reading Level (-.19). Thus, the Reading Recovery sample began in the fall substantially below the random sample and by year-end had surpassed them for all but Text Reading Level. More specifically, the effect size changes from fall to year-end were as follows: Total Score (1.04), Text Reading Level (0.72), Writing Vocabulary (0.83), Hearing and Recording Sounds in Words (0.78), Letter Identification (0.52), Ohio Word Test (1.26), and Concepts About Print

(0.85). Thus, the Reading Recovery sample, as compared to the random sample, increased by approximately one standard deviation unit from fall to year-end across the measures (an average effect size change of 0.86). The fall and year-end test scores for Reading Recovery discontinued students (fall and spring entry combined) and TNI students are shown in Table 3. In fall testing, the Reading Recovery sample Total Score mean and individual task means were all lower than the comparison TNI group, with effect sizes ranging from -.19 (small) to -.52 (medium). By year-end testing, the Reading Recovery students had surpassed the TNI

Table 3. Mean Fall and Year-End Total Scores with Effect Sizes for Successfully Discontinued Reading Recovery and Tested-Not-Instructed Students, 2016–2017

Observation Survey Task	Discontinued		Tested-Not-Instructed		Effect Size (*d*)	
	Fall	Year-End	Fall	Year-End	Fall	Year-End
Total Score	**395.5**	**552.9**	**414.3**	**534.0**	**-.43**	**0.52**
Text Reading Level	1.7	19.6	2.7	17.1	-.50	0.45
Writing Vocabulary	12.8	56.5	16.1	49.8	-.40	0.42
Hearing and Recording Sounds in Words	23.7	36.0	26.5	34.9	-.35	0.40
Letter Identification	49.2	53.5	50.2	53.2	-.19	0.17
Ohio Word Test	4.6	19.2	6.7	18.2	-.52	0.43
Concepts About Print	13.1	21.0	13.9	19.5	-.25	0.60

students on all measures, with effect sizes ranging from .17 (small) to .60 (medium). Thus, the Reading Recovery sample began in the fall substantially below the TNI sample and by year-end had surpassed them for all measures. More specifically, the effect size changes from fall to year-end were as follows: Total Score (0.95), Text Reading Level (0.95), Writing Vocabulary (0.82), Hearing and Recording Sounds in Words (0.75), Letter Identification (0.36), Ohio Word Test (0.95), and Concepts About Print (0.85). Thus, the Reading Recovery sample, as compared to the TNI sample, increased by nearly one standard deviation unit from fall to year-end averaged across the measures (an average effect size change of 0.80).

A few other results should be noted. First, on the Observation Survey Total Score, the discontinued students moved from the 25th percentile in the fall to the 49th percentile at year-end. Second, when considering classroom reading group placement, the discontinued students increased from 16% in the average or higher reading group in the fall to 85% by year-end. Finally, in terms of special education services received, the pool of discontinued students had included 1,407 special education students when Reading Recovery began, and this number was reduced to 146 when Reading Recovery ended. These are additional indications of the efficacy of the Reading Recovery intervention, as discontinued students (a) move to the middle of the distribution on the Total Score, (b) move to the average, above average or well above average reading groups, and (c) are much less likely to need special education services.

> Strong effects such as these would not be possible without the strong commitment of Reading Recovery and Descubriendo la Lectura trainers, teacher leaders, and teachers, who consistently seek to improve their teaching craft. The efforts of these educators continue to result in outstanding literacy success for participating students.

Summary of Descubriendo la Lectura Outcomes

Characteristics of participants
The Descubriendo la Lectura intervention—the reconstruction of Reading Recovery in Spanish—is designed for first graders who receive their initial literacy instruction in Spanish. Table 4 provides details about participation in Descubriendo la Lectura in the United States. For the 2016–2017 school year,

Table 4. Participation in Descubriendo la Lectura in the United States, 2016–2017	
Entity	**n**
University Training Centers	4
Teacher Training Sites	23
States	8
School Systems	28
School Buildings	72
Teacher Leaders	27
Teachers	70
DLL Students	504
Random Sample for DLL	231
Tested-Not-Instructed for DLL	0

NOTE: Some students in the Control Group of the random assignment study did not receive Descubriendo la Lectura. Their data are excluded from results in other tables in this report but included here.

504 children were instructed by 70 teachers. These Descubriendo la Lectura students attended 72 schools in 28 school districts located in 8 states. These teachers were supported by 27 teacher leaders. Fifty-six percent of Descubriendo la Lectura students were boys, 97% were Hispanic, 94% qualified for free or reduced lunch, and there were considerable Title I children (86% schoolwide Title I, 14% individual Title I). Schools were located in 60% urban areas, 32% suburban areas, and 8% rural areas.

For students served in Descubriendo la Lectura, 44% reached the average reading levels of their peers and lessons were successfully discontinued. Another 24% were recommended for further evaluation, 2% moved, and 26% received incomplete interventions. Of the students who completed the intervention (both discontinued and referred students), 65% were discontinued. Of the referred students, of note were 84 recommended for small-group literacy instruction or intervention other than special education, and 18 for literacy-related special education services. Trained teachers had a mean of 19.2 years of teaching experience and 7.9 years of Reading Recovery and/or Descubriendo la Lectura teaching experi-

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, 2016–2017
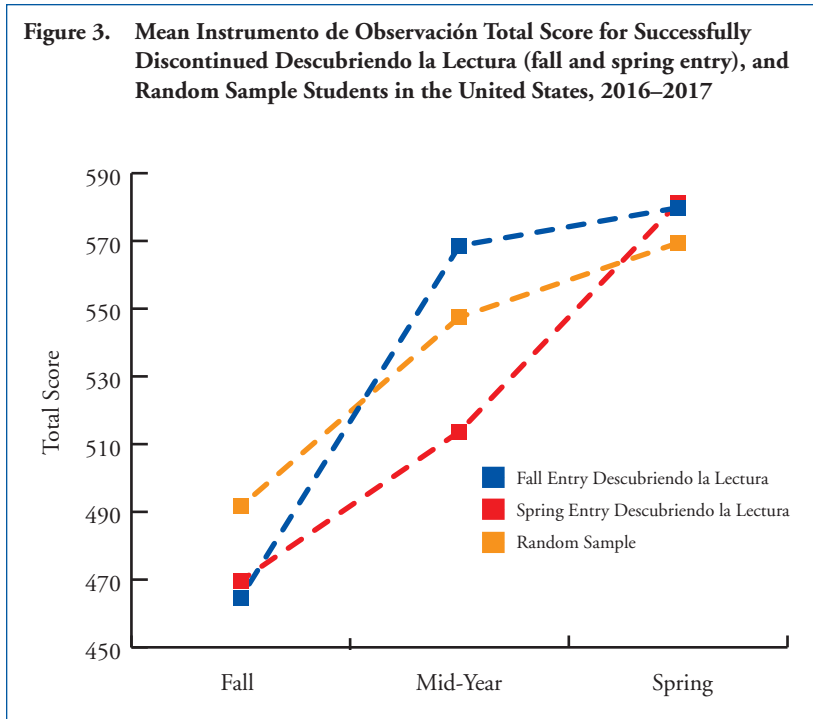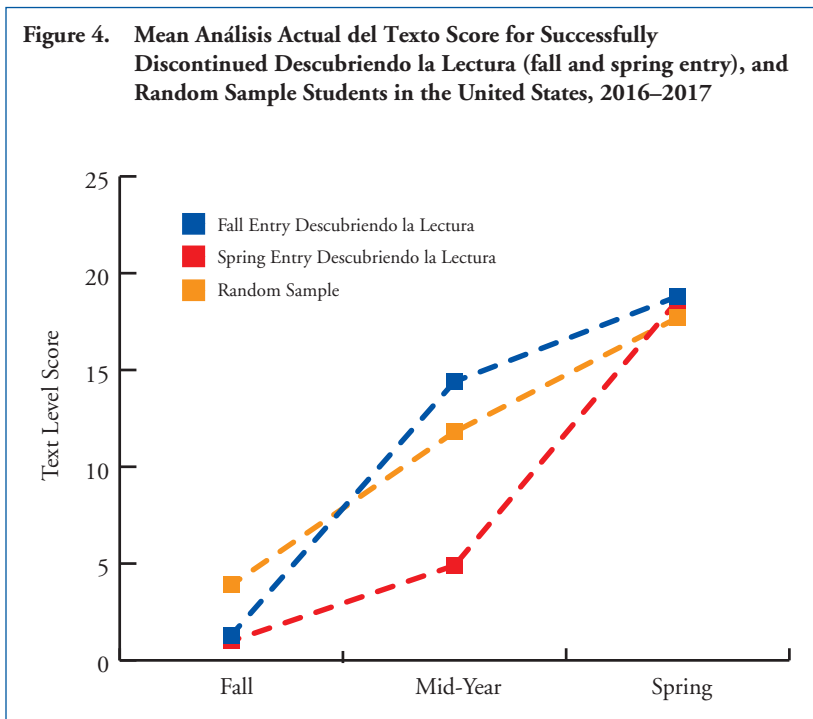

Figure 4. Mean Análisis Actual del Texto Score for Successfully Discontinued Descubriendo la Lectura (fall and spring entry), and Random Sample Students in the United States, 2016–2017

ence. These teachers taught from 2 to 8 children on a daily basis (mode = 4), while teaching a mean of 6.9 children across the school year, and a mean of 24.2 children in other teaching roles, for a mean total of 31.0 children.

Results

Two students per participating Descubriendo la Lectura school were administered the Instrumento de Observación in fall, mid-year, and at year-end in half of the schools randomly assigned. Those students combined represented the random sample. Descubriendo la Lectura schools had last collected TNI data in 2011–2012, but due to very small samples in subsequent years leading to unstable average scores, IDEC did not conduct Descubriendo la Lectura TNI testing. Descubriendo la Lectura random sample students' scores on the six tasks of the Instrumento de Observación were utilized, as was done for Reading Recovery students, to create a 0 to 800-point Total Score measure that reflected literacy development throughout the school year.

Among the fall entry, spring entry, and random sample groups, the largest growth from fall to mid-year on the Instrumento de Observación Total Score was for the fall entry Descubriendo la Lectura students (see Figure 3). From mid-year to year-end, the largest growth was for the spring entry students. Together these results indicate that the greatest gain was during the respective intervention periods. Spring entry Descubriendo la Lectura and random sample students showed approximately the same gain from fall to mid-year. However, from mid-year to year-end, the spring entry

Table 5. Mean Fall and Year-End Total Scores with Effect Sizes for Successfully Discontinued Descubriendo la Lectura and Random Sample Students, 2016–2017

Instrumento de Observacíon Task	Discontinued		Random Sample		Effect Size (*d*)	
	Fall	Year-End	Fall	Year-End	Fall	Year-End
Total Score	**466.7**	**580.2**	**491.2**	**569.4**	**-.68**	**0.40**
Análisis Actual del Texto	1.3	18.7	3.9	17.7	-.82	0.19
Escritura de Vocabulario	11.2	48.9	17.0	45.1	-.59	0.25
Oír y Anotar los Sonidos en las Palabras	24.4	38.3	29.9	37.3	-.54	0.31
Identificacíon de Letras	46.1	59.1	51.0	57.8	-.42	0.32
Prueba de Palabras	8.2	19.7	11.7	18.7	-.54	0.46
Conceptos del Texto Impreso	11.2	20.3	13.3	19.6	-.61	0.21

Descubriendo la Lectura students outgained the random sample. The trend for Text Reading Level (see Figure 4) was very similar to the Total Score trend. By year-end testing, both fall and spring entry Descubriendo la Lectura students had substantially surpassed the scores on both measures as compared to the random sample group. In other words, both Descubriendo la Lectura groups began the school year behind the random sample but caught up to and exceeded the random sample group by the end of the year.

In Table 5 are the mean scores and effect sizes (Cohen's *d*) for fall and spring entry Descubriendo la Lectura discontinued students combined, as well as the Descubriendo la Lectura random sample students at both fall and end of year testing. In fall testing, the Descubriendo la Lectura sample Instrumento de Observación Total Score mean and individual task means were all lower than the comparison random sample group, with effect sizes ranging from -.42 (medium) to -.82 (large). By year-end testing, the Descubriendo la Lectura

students had surpassed the random sample students on all measures, with effect sizes ranging from .19 (small) to .46 (medium). Thus, the Descubriendo la Lectura sample began the fall substantially below the random sample and by year-end had surpassed them on all measures.

More specifically, the effect size changes for the Descubriendo la Lectura students and random sample from fall to year-end were as follows: Instrumento de Observación Total Score (1.08), Analisis Actual del Texto (1.01), Escritura de Vocabulario (0.84), Oir y Anotar los Sonidos en las Palabras (0.85), Identificacion de Letras (0.74), Prueba de Palabras (1.00), and Conceptos del Texto Impreso (0.82). Overall, the Descubriendo la Lectura sample, as compared to the random sample, increased by nearly one standard deviation unit from fall to year-end averaged across the measures (an average effect size change of 0.91).

A few other results should be noted. First, on the Instrumento de Observación Total Score, the discontinued students moved from the 24th percentile in the fall to the 53rd per-

centile at year-end. Second, when considering classroom reading group placement, the discontinued students increased from 15% in the average or higher reading group in the fall to 97% by year-end. Finally, in terms of special education services received, the pool of discontinued students had 247 in special education and this was reduced to only one when Descubriendo la Lectura lessons ended. These are additional indications of the efficacy of the Descubriendo la Lectura intervention, as discontinued students (a) move to the middle of the distribution on the Total Score, (b) move to the average, above average or well above average reading groups, and (c) are much less likely to need special education services.

Conclusion

These results, as well as prior results (e.g., Lomax, 2017), indicate that Reading Recovery and Descubriendo la Lectura continue to be among a very small list of educational interventions with strong impacts on student learning in the United States.

In the 33rd year of implementation in 2016–2017, students receiving these interventions continue to generate strong outcomes. On the Observation Survey Total Score for both Reading Recovery and Descubriendo la Lectura, the average discontinued student caught up to and surpassed the average of the random sample.

Strong effects such as these would not be possible without the strong commitment of Reading Recovery and Descubriendo la Lectura trainers, teacher leaders, and teachers, who consistently seek to improve their teaching craft. The efforts of these educators continue to result in outstanding literacy success for participating students.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

D'Agostino, J. V. (2012). Technical review committee confirms highest NCRTI ratings for Observation Survey of Early Literacy Achievement. *The Journal of Reading Recovery, 11*(2), 53–56.

D'Agostino, J. V., Rodgers, E., & Mauck, S. (2017). Addressing inadequacies of the Observation Survey of Early Literacy Achievement. *Reading Research Quarterly*, doi: 10.1002/rrq.181

Lomax, R. G. (2017). Student outcomes among strongest in 32-year history. *The Journal of Reading Recovery, 16*(2), 47–53.

Lomax, R. G., & Hahs-Vaughn, D. L. (2012). *An introduction to statistical concepts* (3rd ed.). New York: Routledge.

About the Author

Dr. Richard G. Lomax is director of research for Reading Recovery and Descubriendo la Lectura and professor emeritus of educational studies at The Ohio State University, where he was previously associate dean for research and administration in the College of Education and Human Ecology. His research primarily focuses on multivariate analysis and models of literacy acquisition. He has published textbooks and in diverse journals including *Reading Research Quarterly, Parenting: Science and Practice, Understanding Statistics: Statistical Issues in Psychology, Education and the Social Sciences, Violence Against Women, Journal of Early Adolescence, The Journal of Negro Education, International Journal of Computer Science in Sport,* and *International Journal of Sports Medicine.* Named an AERA Fellow, he has served as a Fulbright Scholar on three different occasions; worked on numerous funded projects; and received several teaching, research, service, and book awards.