Effective Early Intervention: Lessons from the i3 Evaluation of Reading Recovery

Robert M. Schwartz, Oakland University

Parents, teachers, educational and political leaders have long recognized the need and continually strived to ensure that all of our nation’s children enjoy the benefits of early literacy learning (Chall, 1996; Flesch, 1955; Snow, Burns, & Griffin, 1998). Despite widespread concern, the educational community has failed to sufficiently support 2 out of 3 children by fourth grade to reach literacy levels we define as proficient, with our highest failure rates among children from poor Black, Latino, and American Indian families (Annie E. Casey Foundation, 2016).

While every commercial reading program promises to bring the latest scientific research to bear on this problem, even well-funded scale-up attempts have had little to no effect. One exception to this pattern is the independent evaluation of the Reading Recovery® scale-up grant. This research demonstrates that it is possible to widely implement an early intervention approach, based in professional development, intended to prevent reading difficulties for 70–80% of the lowest-performing first-grade students with 12–20 weeks of individual, daily 30-minute lessons (Clay, 2005a, 2005b). Reading Recovery also aims to identify the strengths and needs of children requiring additional long-term literacy support (Clay, 1987; Jones, Johnson, Schwartz, & Zalud, 2010; Vellutino, 2010). A more-detailed description of Reading Recovery’s theoretical base, professional development, assessment, progress monitoring, and instructional practices can be found in Changing Futures: The Influence of Reading Recovery in the United States (Schmitt, Askew, Fountas, Lyons, & Pinnell, 2005).

In 2010, The Ohio State University, in partnership with 18 Reading Recovery university training centers, received an Investing in Innovation (i3) scale-up grant. The grant totaled $45 million, with an additional $10.1 million raised from private sources, to cover the expansion of Reading Recovery around the U.S. (See Rodgers & D’Agostino, 2012 for a description of the project.) Of this, $4 million was earmarked for the completion of the independent evaluation of the scale-up effort between 2011 and 2015 (D’Agostino, personal communication, May 9, 2016).

Scale-up grants were awarded to innovations with the strongest evidence of effectiveness to determine whether similar results could be demonstrated in a variety of schools, demographic groups, and implementation conditions. The scale-up goal for the Reading Recovery grant was “training 3,675 new Reading Recovery teachers; providing one-to-one Reading Recovery lessons to an additional 67,264 students; and, delivering other instruction—generally classroom or small-group instruction—to 302,688 more students via teachers trained in Reading Recovery during the scale-up” (May, Sirinides, Gray, & Goldsworthy, 2016, p. 2).

Scale-up Outcomes
The Reading Recovery network was very successful in meeting these targets. The evaluators (May et al., 2016) report that the scale-up surpassed its goals for the number of Reading Recovery teachers trained (3,747 or 102% of the scale-up goal),
and the number of non-Reading Recovery students served by those teachers (325,458 or 108% of the scale-up goal). In the third area, students served with one-to-one Reading Recovery lessons provided by teachers trained with i3 funds, the project achieved 92% of its goal, with 61,992 students (p. 2). The i3 grant initiated or expanded Reading Recovery in over 1,300 schools. These schools were classified as suburban (29.8%), urban (27.9%), rural (27.2%), and towns (13.8%). Also in 20% of the i3 schools, ELL students comprised more than 20% of the student population (May et al.).

Impact Outcomes
The major goal of the independent evaluation was to provide experimental evidence that the intervention continued to demonstrate a strong impact on the literacy learning of the lowest-performing first-grade students as the Reading Recovery community expanded. The final evaluation report combines data across the 4-year scale-up implementation to conduct both a randomized controlled trial (RCT) and a regression discontinuity (RD) analysis. Both the RCT and RD studies are designed to reveal the causal effect of the Reading Recovery intervention on students’ literacy learning.

The RCT study was conducted across the first half of the school year so control group students could receive Reading Recovery service, if needed, during the second half of the year. The RCT included 3,444 matched pairs of low-performing first-grade students from 1,122 schools. The RD design also provides an estimate of Reading Recovery’s impact, but in this design the comparison is between outcomes for children initially below or above a school’s cut score on the screening measures used to determine who receives intervention services. Statistical procedures were used to determine if the group receiving intervention services performed higher than expected on the outcome measures relative to the comparison group that did not receive the intervention.

Based on both types of designs, May et al. (2016) report learning impacts that were positive and impressive due to the Reading Recovery intervention. The main outcome measures were the Iowa Test of Basic Skills (ITBS) Total Reading score, which includes subscales for Comprehension and Word Reading, and an Observation Survey of Early Literacy Achievement (OS) Total Score (Clay, 2005c; see D’Agostino, 2012 for a description of the Total Score). For each set of scores, the means are one-third to one-half of a standard deviation larger for the treatment group. With the RCT, these are the differences in mean percentile ranks between the Reading Recovery and control group:

- +18 for ITBS Total Reading score
- +16 for ITBS Reading Words
- +16 for ITBS Reading Comprehension
- +26 for OS Total Score (p. 39)

These results apply to the overall sample and are replicated with different schools in each of four school years (2011–2012 to 2014–2015; see May et al., 2016, p. 153). In addition, similarly large effects are shown in subgroup analyses for English language learners (ELL) and rural schools (pp. 39–41). This replication of strong effects across schools, years, and subgroups is a remarkable accomplishment seldom achieved in even tightly controlled experiments (Open Science Collaboration, 2015), let alone large complex scale-ups (Bryk, Gomez, Grunow, & LeMahieu, 2015).

The evaluators provided some benchmarks to help educators judge the impact of Reading Recovery:

- Typically, research on interventions at the elementary
level using “broad scope” standardized tests as outcomes found the average effect size was 0.08 standard deviations. So, at 0.37 standard deviations the Reading Recovery effect on the ITBS Total Reading score “was 4.6 times greater than average for studies that use comparable outcome measures” (p. 42).

- A meta-analysis of Title I programs (Borman & D’Agostino, 1996) posits the Reading Recovery impact was 3.5 times the average effect of Title I interventions.

- Finally, the norms for the ITBS Total Reading score indicate an additional 1.6 months of learning for the Reading Recovery group, a growth rate that is 31% greater than the national average for first grade (p. 43).

This impact is sufficiently large to close the achievement gap between many of our most-at-risk students and their average peers by the end of first grade.

Previous Attempts to Address Low Literacy Achievement

The Reading Recovery i3 grant is certainly not the first attempt to address the individual and societal costs of low literacy achievement. These problems persist despite extraordinary investments of state and federal resources.

Reading First

The Reading First component of the No Child Left Behind Act of 2001 (PL. 107-110, Title I, Part B, Subpart 1) mandated $1 billion per year to “promote instructional practices validated by scientific research” (Gamse, Jacob, Horst, Boulay, & Unlu, 2008, p. xvi). Over a 5-year period, subgrants were provided to 5,880 schools in 1,809 school districts. The funds were designated for curriculum materials, professional development, and diagnosis and prevention of reading difficulties.

What was the return on investment from this massive federal program? The Reading First Impact study (Gamse et al., 2008) reports substantial increases in the amount of time devoted to the scientifically identified components of instruction, support by full-time literacy coaches for classrooms teachers, and additional services for struggling readers. Although instructional practices changed, the benefits for children were less impressive. Game et al. found a small, but statistically significant increase on decoding scores for first-grade students in one school year. Unfortunately, for the 3 years of the study, they report no increase in student’s reading comprehension scores in Grades 1, 2, or 3.

Response to Intervention (RTI)

A recent evaluation of response to intervention (RTI) continues to demonstrate the difficulty of implementing large scale, research-based literacy initiatives (Balu et al., 2015). The “2004 reauthorization of the Individuals with Disabilities in Education Act (IDEA) allows states and school districts to use a portion of federal special education funds to provide coordinated early intervening services to students at risk of reading failure or other academic or behavioral problems” (p. ES-1). Since RTI services are provided prior to special education classification, schools also invest general education and Title I funds in support of these early intervention services. The 2011–2012 evaluation studied “a reference sample of schools representative of elementary schools in the 13 states included in the evaluation and an impact sample of 146 elementary schools with 3 or more years of implementing RTI approaches in reading” (p. ES-1).

The impact sample showed a high rate of implementation of the assessment, staffing, and data-based decision processes related to RTI. The effectiveness of these RTI practices were evaluated using an RD design that “compared the difference in reading outcomes between students whose fall screening test scores were just above the cut point for Tier 2 intervention set by the schools and those whose scores were just below” (p. ES-6). No statistical differences were found for children who received RTI services versus children who did not on measures of word recognition in Grades 1 or 2, and no differences on state reading tests at third grade after multiple years of RTI services. Surprisingly, the evaluation indicated a significant negative impact on reading comprehension for children receiving RTI services in first grade!

Can Early Literacy Gains Be Sustained? Exploration and Expectations

Reading Recovery is an RTI approach that was not evaluated in the Balu et al. (2015) study. Given the large, positive impacts on literacy learning in first grade demonstrated by Reading Recovery, the i3 evaluation attempted to go beyond assessing the immediate impact of the intervention by using an RD analysis to estimate the sustained effects, 2 years after the intervention in third grade. The
design and analysis used to examine sustained effects was considerably weaker than that used to examine the immediate impact of Reading Recovery (May et al., 2016).

This design required comparing outcomes for students who received the Reading Recovery intervention relative to students who scored slightly higher on the first-grade screening tasks and therefore did not receive the intervention. Only students from schools assigned to the RD study in the first 2 years of implementation had reached third grade by the final year of the grant and therefore had outcome scores on a state assessment test. Attrition, due to various causes, further reduced the number of schools available for this analysis from 325 to 77.

An additional complication in this exploration of sustained effects is that low-performing students in these schools could continue to receive small-group instruction from Reading Recovery trained teachers for 2 years after the intervention period ended. These continued RTI services could reduce any differences between the treatment and comparison group over time. Thus, it was not surprising that the overall RD analysis showed no effect on third-grade state tests.

Even with these design problems, a subgroup analysis of students judged to have made accelerated progress during the intervention period (those students whose programs were considered successfully discontinued) showed a marginally significant (p < .10, but > .05), positive effect of .19 standard deviations on the third-grade state tests.

The question for schools and school districts is not whether these children will sustain their gains relative to a control group receiving literacy support over the next few years, but rather under what conditions these children will continue to benefit from classroom instruction so we are successful in helping them become proficient readers.

Annual Reading Recovery national evaluations (see D’Agostino & Brownfield, 2016 or Gómez-Bellengé & Thompson, 2000 for examples) continually demonstrate that children who make accelerated progress during Reading Recovery interventions in the fall continued to make gains across the 3–5 months following the intervention. The question for schools and school districts is not whether these children will sustain their gains relative to a control group receiving literacy support over the next few years, but rather under what conditions these children will continue to benefit from classroom instruction so we are successful in helping them become proficient readers (Jesson & Limbrick, 2014; Schwartz, 2015).

Marie Clay’s expectation was that Children who successfully complete early literacy interventions like Reading Recovery should operate in reading and writing in ways that put them on track for being silent readers with self-extending systems during the next two years at school. With good classroom instruction and moderate personal motivation that should be achievable. (2005a, p. 52)

Effectiveness and Cost
It is worth noting that neither the Reading First study nor the RTI evaluation attempted to examine sustained effects since they were not able to demonstrate a positive immediate impact on a measure of reading comprehension and very minimal effects on word level reading. Because the i3 program differs considerably from the conditions involved in these legislative initiatives, it is more appropriate to compare the Reading Recovery i3 evaluation to the independent evaluation of Success for All’s (SFA) i3 scale-up grant (Quint, Zhu, Balu, Rappaport, & DeLaurentis, 2015).

Success for All
In the overview, the evaluators described SFA as “one of the best-known school reform initiatives. Combining a challenging reading program, whole-school reform elements, and an emphasis on continuous improvement, it seeks to ensure that every child learns to read well in the elementary grades” (p. iii). SFA used the grant to expand to 447 new schools and reached 276,000 students. The evaluation used an RCT design to compared student performance in 19 schools that
implemented SFA to students in 18 similar schools that used alternative reading programs (but not Reading Recovery).

The primary analysis examined effects for second-grade students who had participated in the SFA schools since kindergarten and compared them to students in the control group schools. This 3-year treatment resulted in no significant differences from the control group on

- a comprehension measure (Woodcock-Johnson Passage Comprehension),
- a measure of word reading efficiency (TOWRE2), or
- measures of letter and word identification (Woodcock-Johnson Letter-Word Identification).

The only significant difference was on a measure of nonsense-word reading (Woodcock-Johnson Word Attack), with a $t$ standard deviation effect (Quint et al., 2015).

The evaluators further break down this analysis to look at second-grade students in the SFA or control group who were identified as high or low performers based on their kindergarten reading level. For the initially low-performing students, this analysis resulted in a slightly larger effect on the nonsense-word reading task (.23 standard deviations), with marginally significant gains ($p < .10$, but $>.05$) on the Woodcock-Johnson Letter-Word Identification (.17 standard deviations) and the measure of word reading efficiency (.19 standard deviations).

There were no significant effects on any of these measures for the high-performing students in this analysis and neither high- nor low-performing SFA students showed significant gains on the comprehension measure relative to students in the control group schools (Quint et al., 2015).

**Comparative Costs**

Cost and cost-effectiveness estimates are always complex. In their executive summary, the evaluators claim “scale-up findings show that, for a modest investment, SFA reliably improves the decoding skills of students in kindergarten through second grade, and that it is especially beneficial for students who begin in the lower half in these skills” (Quint et al., 2015, pp. ES5–6). The SFA scale-up provides considerable information on the cost of implementing SFA relative to the costs of the literacy program in control schools within one district. Again in the executive summary, they report “program group schools spent about $227 worth of resources per student per year more than control group schools to implement their respective reading programs” (p. ES5). Of course, this is $227 per student per year for every student in Grades K–5, with the only effect shown for second-grade students reading nonsense words. For students in third, fourth, and fifth grades, after 3 years of SFA there were no effects favoring the SFA group, and one marginally significant difference ($-.13$ standard deviations) on the fourth-grade state test that favored the control group (see Table 5.6, p. 86).

The information in the SFA evaluation (Quint et al., 2015) provides a way to compare the relative cost of scaling up Reading Recovery and SFA. The funding for the SFA i3 grant was approximately equivalent to the Reading Recovery grant, about $50 million dollars over 5 years. With these funds, SFA served some 276,000 students enrolled in 447 schools over the first 4 years of the grant, and estimate that they will serve an additional 34,000 in the final grant year. Dividing the grant funds by the total students served yields a cost of $161 per student served in SFA schools.

The Reading Recovery final evaluation (May et al., 2016) indicates that teachers who received Reading Recovery professional development served 61,992 first-grade children in one-to-one Reading Recovery interventions and an additional 325,458 students in other contexts. Dividing the grant funds by these totals yields $807 per child served in the Reading Recovery intervention, and $129 per student served, in any context, by Reading Recovery i3 teachers. The higher per-student estimates of Reading Recovery intervention costs in previous studies (Shanahan & Barr, 1995; Hollands et al., 2013) included the portion of the intervention teacher’s salary devoted to one-to-one instruction. As in the SFA implementation, most schools did not hire new teachers to implement Reading Recovery. Rather, principals reassigned teachers to participate in the professional development and provide intervention services.

While salary is a very real cost for districts, it is not an additional cost for schools that are currently using literacy coaches and intervention teachers to provide the far less-effective services described in the SFA,
As others have repeatedly warned, there is no quick fix for the literacy challenges of our nation. The good news from the Reading Recovery i3 evaluation is that if we are willing to do the work, we can make a meaningful impact on the literacy learning of many of our most at-risk students.
This orientation "directs attention toward a more realistic appraisal of 'What works, for whom, and under what set of conditions?'" (p. 13–14).

RRCNA will continue to learn with teachers and administrators across the educational community as we build on the i3 scale-up research to further optimize early intervention services for young children and construct systems to support their ongoing literacy success (Askew, Pinnell, & Scharer, 2014; Arteberry & Bryk, 2011; Dorn & Schubert, 2010). We encourage new districts and professionals to participate in the scale-up of Reading Recovery, contribute to and learn from our Reading Recovery network, and join us in ensuring that all students gain the literacy skills needed for the 21st Century. Additional information about Reading Recovery may be found on the RRCNA website at http://readingrecovery.org.

References


---

**About the Author**

Bob Schwartz is an emeritus professor in the Department of Reading and Language Arts at Oakland University in Rochester, MI. He is a past president of and current research consultant for the Reading Recovery Council of North America. His research interests include self-monitoring in beginning reading, early literacy intervention, research design, and professional development for literacy teachers. In the What Works Clearinghouse 2007 review of 887 studies from 153 beginning reading programs, Dr. Schwartz’s Reading Recovery research was one of only 27 studies that met WWC’s standards without reservations.

Contact: rschwartz@oakland.edu