How Much Evidence Is Enough Evidence?

Richard L. Allington, University of Tennessee

With the current focus on evidence-based educational practice, how can one explain the lack of support for Reading Recovery? As reported in *Education Week* (Hoff, 2002), "A group of reading researchers has launched a campaign against Reading Recovery, contending that the popular one-on-one tutoring program fails to deliver the student achievement gains it promises" (p. 1). Further, it seems many states are not encouraging the use of Reading First funds to support Reading Recovery programs. Some seem openly antagonistic towards Reading Recovery.

In this short article I will summarize the research on Reading Recovery, focusing primarily on reports of independent researchers, and then also note the scientific evidence supporting tutoring programs generally as the most effective scheme for accelerating reading development. I close by questioning why federal policy has not been targeted to assure that children who would benefit from Reading Recovery and other tutoring programs have access to such intervention designs.

The Evidence for the Efficacy of Reading Recovery

To my knowledge only a single meta-analysis of the impacts of Reading Recovery is available, but that study demonstrated the positive impacts across a large number of studies of Reading Recovery effectiveness. D'Agostino and Murphy (2004) report on a meta-analysis of 36 studies of Reading Recovery from a sample of 109 studies located (they included only those studies that provided sufficient data for meta-analyses and effect size calculations). These studies provided 1,379 different effect sizes. Additionally, 11 of the most rigorous studies that provided pre- and posttest data on treatment and control subjects were analyzed separately

They assessed Reading Recovery effects on discontinued and not-discontinued students and control groups. When looking at all Reading Recovery students, gains from pre- to posttest were significantly larger than control groups on all measures except standardized tests. When examining only discontinued student gains, the Reading Recovery students improved significantly on all measures including standardized tests. The authors note that regression effects cannot account for observed gains.

One earlier evaluation did include estimated effects sizes across the domain of Reading Recovery studies then available and found Reading Recovery effective as an intervention for struggling readers. Shanahan and Barr (1995) included an analysis of all published evaluations of Reading Recovery available and unpublished evaluations that included sufficient data. Whenever possible, effect sizes were computed. Their analysis revealed that Reading Recovery pupils made greater-than-expected gains in reading, comparable to those of the most effective interventions. But these authors argued that Reading Recovery seemed less effective than promoters claimed and more costly, too. These researchers concluded that Reading Recovery merits continued support and offer several recommendations for improvement.

In an analysis of British intervention studies, including Reading Recovery, the research also demonstrated the effectiveness of Reading Recovery. The Brooks, Flanagan, Henkhuzens, and Hutchinson (1998) report analyzed 20 British studies providing evidence on the effectiveness of about 30 approaches to early reading interventions. Of particular interest is the evidence that Reading Recovery and Paired Reading produced sustained gains (evidence strongest for children receiving free lunch). They also noted that phonological-emphasis approaches were "substantially less effective than the main experimental approach; and the main approach was broader and incorporated work on phonological skills.…For the greatest impact with struggling readers, therefore, work on phonological skills should be embedded within a broad approach" (p. 9).

Similarly, Reading Recovery was found effective when compared to other tutoring interventions. Wasik and Slavin (1993) reported Reading Recovery as one of several expert tutoring approaches that produced reliable, positive effects on struggling readers' reading development. Likewise, Hiebert (1994) found Reading Recovery an effective intervention but provided a methodological critique of the studies evaluating the impact of Reading Recovery on a cohort of students within a school or district. I found her conclusions of Reading Recovery a bit severe, if only because no other intervention had been evaluated using the conservative criteria she proposed and then applied to Reading Recovery. Like Shanahan and Barr (1995), Hiebert seemed more concerned with detailing a rigorous evaluation model than disputing

the effect of Reading Recovery on struggling readers. Personally, I think Hiebert's model evaluation would be a useful standard for estimating the longer-term impact of any intervention. We simply know very little about how early intervention efforts impact the reading achievement of a cohort of children at the end of the elementary school years (Grade 5 or 6), and we have no information on effects beyond that point. Does early intervention reliably reduce the number of poor readers in a cohort when they enter high school? We can hope so, but there is little evidence even to debate.

Other studies reported on comparisons of the effects of Reading Recovery compared to control group children. In some of these studies the control groups participated in alternate intervention (which in some cases was a modified Reading Recovery lesson design). Center, Wheldall, Freeman, Outhred, and McNaught (1995) provided a conservative analysis of Reading Recovery effects. They found that Reading Recovery was able to accelerate the reading growth of 35% of children who would not reach the level of successful peers under other conditions. Reading Recovery participants performed significantly better than controls on word reading and word attack but not on a measure of phoneme awareness.

Iverson and Tunmer (1993) compared a modified Reading Recovery lesson framework (with added daily focus on phonological awareness) with standard Reading Recovery lessons. They found that children in the adapted Reading Recovery were discontinued with fewer lessons but report no other differences between groups on reading measures. Thus, the Reading Recovery intervention was found effective and with minor modifications (many incorporated later in a revised Reading Recovery lesson plan [Clay, 1993]) accelerated reading development a bit more quickly.

Finally, Stahl, Stahl, and McKenna (1999) found that the use of Elkonin boxes and Making and Breaking word analysis activities with magnetic letters were important components of Reading Recovery in development of phonological awareness. They examined outcomes for 12 of the lowest-achieving (on the Observation Survey) of 31 students ranked in the bottom 25% of their five first-grade classrooms. Reading Recovery students received traditional Reading Recovery with modified phonological awareness procedures as included by Clay in her 1993 text, *Reading Recovery: A Guidebook for Teachers in Training*. Two-thirds of the students were minority, mostly African-American. Reading Recovery subjects got daily 30-minute lessons from one of three Reading Recovery teachers scheduled during the school day. Control group students received classroom reading instruction in a 2-hour block from their classroom teacher.

The authors also developed a pseudoword reading measure and used the Yopp-Singer test as additional tests beyond the Observation Survey because some researchers had criticized the Observation Survey as too closely mirroring Reading Recovery lessons. All subjects were pretested on the Observation Survey plus the Yopp-Singer test and posttested on the Observation Survey, Yopp-Singer, and the pseudoword test. They reported no significant differences on pretest data between Reading Recovery and control students, even though the control group scored slightly higher on all measures. At posttest an ANCOVA was used to test for effects. Significant differences favoring the Reading Recovery students were found on Yopp-Singer (Yopp, 1995) and the Hearing and Recording Sounds in Words subtest of the Observation Survey (Clay, 2002). On the pseudoword reading task, the discontinued Reading Recovery students performed almost as well as the average first graders the experimenter-developed test was piloted on.

> If evidence—scientific research evidence—was the true standard for decisions, then Reading Recovery and other tutoring interventions would be available for every child who could benefit from them.

Tutoring Effects Generally

Contradicting the assertion of some researchers that there is no evidence supporting tutoring designs (e.g., see J. Fletcher's statements in the Hoff article in *Education Week* referenced above), the U.S. Department of Education has recently concluded that the evidence available on expert tutoring meets the rigorous gold standard required for federally funded interventions.

The U. S. Department of Education's report (2003), *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide* noted

> As illustrative examples of the potential impact of evidence-based interventions on educational outcomes, the following have been found to be effective in randomized controlled trials—research's 'gold standard' for establishing what works:
>
> • One-on-one tutoring by qualified tutors for at-risk readers in grades 1-3 (the average tutored student reads more proficiently than approximately 75% of the untutored students in the control group). (p. iii)

The evidence on expert tutoring is clear: tutoring is more effective than any other intervention design. Perhaps some researchers have been misled, or confused, by the National Reading Panel's report (NICHD, 2000) that suggested that tutoring in the meta-analysis they conducted was found no more effective than other intervention designs. But, as Camilli, Vargas, and Yurecko (2003) noted, that meta-analysis was procedurally flawed and when done correctly the impact of tutoring was substantially larger than the impact of systematic phonics instruction.

Critics of Reading Recovery have argued that it accelerates the development of only some of the children who participate. But that is true of all interventions for struggling readers. For instance, Joe Torgeson (2000), an NICHD-funded researcher, has pointed to the problem of *treatment resisters* in NICHD studies. He discusses five reading intervention studies focused on very low-achieving (lowest 15th–20th percentile in three of the studies) emergent readers. In three studies children with low IQs were excluded (IQ cutoffs ranged from 75 to 90). His analysis indicated that anywhere from 20–45% of the children in these interventions, many that included a tutoring component, had reading achievement below the 30 percentile at the conclusion of the intervention.

More recently, Torgeson (2002) has noted that the "ultimate goal" of reading instruction is for kids to be able to "comprehend printed material at a level commensurate with their general verbal ability or language comprehension skills" (p. 10). But he argues that "if we were to adopt a strict grade-level reading comprehension criteria *[sic]*,…this would imply an expectation for all children to have at least average verbal ability. Decades of cognitive intervention research suggests that it is unrealistic to expect all children to attain verbal ability estimates within the average range as a result of special instruction" (p. 10). Thus, Torgeson seems to be suggesting that a standard which requires an intervention to produce grade-level reading achievement is fundamentally flawed. I worry about this statement while also being aware of the difficulty researchers have had in designing any intervention that achieved the more rigorous standard of grade-level achievement for all participants. I worry because such a view may undermine attempts to provide sufficient support to develop the full capacity of all children.

Thus, the evidence from a variety of intervention studies show tutoring is effective, but even tutoring interventions have not produced a distribution of reading achievement where all students achieve grade-level proficiency (Allington, 2004).

Conclusions

There is a powerful research base supporting the efficacy of Reading Recovery specifically, and for expert, intensive tutoring interventions in general. It is wholly unclear to me, as a Reading Recovery outsider, how so many current state Reading First designs support the use of completely unproven interventions—Voyageur or Waterford Early Reading, for instance—while failing to encourage the use of federal funds to support Reading Recovery. I sense a triumph of ideology over evidence (Allington, 2002; Allington & Nowak, 2004) once again.

Let me close by noting that while I believe that Reading Recovery has more research evidence supporting its efficacy than any other intervention in the marketplace, I do think we can improve on Reading Recovery. I am concerned that the lack of fiscal support for Reading Recovery will undermine efforts to continue to fine-tune its design and to improve its implementation as part of a needed whole-day approach to effective reading instruction.

There are no easy answers in determining the future of Reading Recovery or in designing schooling so that literacy for all is the common standard. Unfortunately, research evidence doesn't seem to count for much in educational planning despite all the rhetoric about evidence-based educational design. If evidence—scientific research evidence—was the true standard for decisions, then Reading Recovery and other tutoring interventions would be available for every child who could benefit from them.

References

Allington, R. L. (2002). *Big brother and the national reading curriculum: How ideology trumped evidence.* Portsmouth, NH: Heinemann.

Allington, R. L. (2004). Setting the record straight. *Educational Leadership, 61*(6), 22–25.

Allington, R. L., & Nowak, R. (2004). 'Proven programs' and other unscientific ideas. In C. C. Block, D. Lapp, E. J. Cooper, J. Flood, N. Roser, & J. V. Tinajero (Eds.), *Teaching all the children: Strategies for developing literacy in an urban setting* (pp. 93–102). New York: Guilford.

Brooks, G., Flanagan, N., Henkhuzens, Z., & Hutchinson, D. (1998). *What works for slow readers? The effectiveness of early intervention schemes.* Slough, Berkshire UK: National Foundation for Educational Research.

Camilli, G., Vargas, S., & Yurecko, M. (2003). Teaching children to read: The fragile link between science and federal education policy. *Education Policy Analysis Archives, 11*(15). Retrieved May 20, 2003, from http://epaa.asu.edu/epaa/v11n15/.

Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly, 30*(2), 240–263.

Clay, M. M. (1993). *Reading Recovery: A guidebook for teachers in training.* Portsmouth, NH: Heinemann.

Clay, M. M. (2002). *An observation survey of early literacy achievement* (2nd ed.). Portsmouth, NH: Heinemann.

D'Agostino, J. V., & Murphy, J. A. (2004). A meta-analysis of Reading Recovery in United States schools. *Educational Evaluation and Policy Analysis, 26*(1), 23–38.

Hiebert, E. H. (1994). Reading Recovery in the United States: What difference does it make to an age cohort? *Educational Researcher, 23*(9), 15–25.

Hoff, D. J. (2002). Researchers urge officials to reject Reading Recovery. *Education Week, 21*(39) 1, 20.

Iverson, S., & Tunmer, W. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology, 85*(1), 112–126.

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Report of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing.

Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly, 30,* 958–996.

Stahl, K. A. D., Stahl, S. A., & McKenna, M. C. (1999). The development of phonological awareness and orthographic processing in Reading Recovery. *Literacy Teaching and Learning, 4*(1), 27–41.

Torgeson, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15*(1), 55–64.

Torgeson, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*(1), 7–26.

U. S. Department of Education (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide.* Washington, DC: U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Wasik, B. A., & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28*(2), 178–200.

Yopp, H. K. (1995). A test for assessing phonemic awareness in young children. *The Reading Teacher, 49*(1), 20–29.

About the Author

Dick Allington is a professor of education at the University of Tennessee. He is the president-elect of the International Reading Association and a former president of the National Reading Conference. His extensive publications include *Classrooms That Work: They Can All Read and Write*, and *Schools That Work: Where All Children Read and Write*, both co-authored with Pat Cunningham; *No Quick Fix: Rethinking Reading Programs in American Elementary Schools* with Sean Walmsley; *What Really Matters for Struggling Readers; Reading to Learn: Lessons From Exemplary 4th Grade Classrooms;* and *Big Brother and the National Reading Curriculum: How Ideology Trumped Evidence.*