



Is Reading Recovery an Effective Intervention for Students with Reading Difficulties? A Critique of the i3 Scale-Up Study

James W. Chapman & William E. Tunmer

To cite this article: James W. Chapman & William E. Tunmer (2016) Is Reading Recovery an Effective Intervention for Students with Reading Difficulties? A Critique of the i3 Scale-Up Study, *Reading Psychology*, 37:7, 1025-1042, DOI: [10.1080/02702711.2016.1157538](https://doi.org/10.1080/02702711.2016.1157538)

To link to this article: <https://doi.org/10.1080/02702711.2016.1157538>



Published online: 24 Mar 2016.



Submit your article to this journal [↗](#)



Article views: 569



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

IS READING RECOVERY AN EFFECTIVE INTERVENTION FOR STUDENTS WITH READING DIFFICULTIES? A CRITIQUE OF THE I3 SCALE-UP STUDY

JAMES W. CHAPMAN and WILLIAM E. TUNMER

Institute of Education, Massey University, Palmerston North, New Zealand

The recently reported i3 Scale-Up of Reading Recovery (May et al., 2015) found an effect size of + 0.69 in favor of Reading Recovery compared to the control group. We discuss four issues: (a) many of the lowest achieving students were excluded from participation in Reading Recovery; (b) the control group received a range of different experiences; (c) the successful completion rate of students in the program was modest; and (d) no data supported the claim that Reading Recovery leads to sustained literacy learning gains. We question the value of this study as the basis for widespread implementation of Reading Recovery.

In an earlier article in *Reading Psychology*, we discussed issues related to the theoretical underpinnings of the Reading Recovery (RR) program, along with concerns about the instructional strategies, assessment battery, and the nature of program delivery (Tunmer & Chapman, 2003). Very little has changed since those concerns were raised. Reading Recovery has become the most widely used early intervention program in the world for young children at risk for developing reading difficulties (Slavin, Lake, Davis, & Madden, 2011). Developed in New Zealand by Marie Clay (1985) in the late 1970s, the program was implemented in schools throughout the country during the 1980s. Since then, RR has been adopted for use in various parts of Australia, Canada, Europe, the United Kingdom, the United States, and South America.

Given the extent to which the program has been adopted, and considering the numerous criticisms of the program over the years (e.g., Center, Wheldall, Freeman, Outhred, & McNaught, 1995; Chapman, Greaney & Tunmer, 2007; Elbaum, Vaughn, Hughes, & Moody, 2000; Reynolds, Wheldall & Madelaine, 2009),

Address correspondence to James W. Chapman, Institute of Education, Massey University, Private Box 11 222, Palmerston North 4440, New Zealand. E-mail: j.chapman@massey.ac.nz

it is not surprising that there is considerable interest in research regarding its efficacy. Numerous studies have been reported on various aspects of the program since the 1980s, although no robust, well-designed study has been reported in support of the program's effectiveness in New Zealand (Chapman, Greaney & Tunmer, 2015). Indeed, relatively few well-controlled studies of the effectiveness of RR have been published in peer-reviewed journals. A recent What Works Clearinghouse (WWC) report on the RR program identified 202 studies that examined RR in terms of its effectiveness for at-risk beginning readers (U.S. Department of Education, 2013). According to WWC, only three met the evidence standards involving randomized controlled trials. This number is surprisingly low. However, WWC has been criticized for adopting a flawed approach in evaluating literacy interventions, resulting in a very restrictive set of conclusions about the efficacy of RR (e.g., Reynolds, Wheldall, & Madelaine, 2009; Stockard, 2010; Wood, 2014). Even if more studies were included in the WWC conclusions regarding the efficacy of the RR program, as suggested by Reynolds et al. (2009), the number would still be relatively few, considering the 30-year history of the program and its widespread use.

Because of the small number of well-designed RR studies, the recently reported "multisite randomized evaluation" of the program (May et al., 2015) is of particular significance. The key result from this "scale-up" project, involving 866 students in 147 schools randomly selected from across the United States, was an effect size of +0.69 for mean total reading scores on the Iowa Test of Basic Skills, in favor of children who received RR compared to those in the control group. This large effect size is important in the context of a program that aims to expand its availability across the United States. Any large scale program that focuses on intervention for overcoming reading disabilities warrants close examination.

At first glance, the magnitude of the positive effect size appears considerable. A special WWC report on this study (U.S. Department of Education, 2014) endorsed the research design, stating that it is "a well-executed randomized controlled trial" (p. 2). We disagree. In this article, we address the following four issues with the May et al. (2015) study: (a) the selection of students; (b) the control group comparison process; (c) the results of participation in the program; and (d) claims regarding the effectiveness

of RR and the sustainability of gains arising from participation in the program. Prior to discussing these issues, we briefly overview the nature of the RR program.

The Reading Recovery Program

Reading Recovery is defined as a “preventative” early intervention program designed to accelerate to average levels of reading performance the progress of six-year-old children who show early signs of reading difficulty (Clay, 1987, 1993). Clay (1987) was confident about the effectiveness of RR. She claimed that it is a

programme which should clear out of the remedial education system all the children who do not learn to read for many event-produced reasons [i.e., environmental, cultural, or economic causes] and all the children who have organically based problems but who can be taught to achieve independent status in reading and writing despite this. (p. 169)

Largely on the basis of this claim, and despite significant design flaws with Clay’s own research on the effectiveness of RR (e.g., Center et al., 1995; Nicholson, 1989), the New Zealand Department (now Ministry) of Education funded the adoption of RR in schools throughout the country during the 1980s (Soler & Openshaw, 2006).

Children selected for RR receive 30 minutes of daily individual instruction over 12 to 20 weeks by specially trained RR teachers. Instructional emphasis is placed on developing within the child a self-extending system of reading strategies that involves the use of multiple cues (syntactic, semantic, visual, graphophonic) to detect and correct errors while constructing meaning from text (Clay, 1991). Specific activities have been detailed by Clay in several publications (e.g., Clay, 1993, 2005a, 2005b).

The RR program is in addition to the regular classroom reading program, and in New Zealand, RR complements the whole language, constructivist approach to beginning literacy instruction (Clay, 1993, 2005a, 2005b). However, in line with the view that RR should be exported to other countries (Soler & Openshaw, 2006), Clay (1993) maintained that RR is compatible with all types of classroom programs, and that, “in a brief period of

help. . .brings the hardest-to-teach children” (p. i) to at least average levels.

The RR i3 Scale-Up in the United States

May et al. (2015) reported that the i3 Scale-Up study was funded by a grant of \$45 million from the U.S. Department of Education as part of the 2010 economic stimulus package. An additional \$10.1 million was provided from “private sources,” although these sources were not disclosed. The purpose of the five-year grant is to increase RR availability in more than 1,400 schools, with targeted assistance for over 88,000 students.

In designing their follow-up study, the authors of the i3 Scale-Up study intended to follow the criteria used by WWC. May et al. provided a brief summary of previous WWC reports on RR. They noted the WWC conclusion that RR has “potentially positive effects with a small extent of evidence” (May et al., 2015, p. 7). Their interpretation of “small extent of evidence” relates to the fact that the sample sizes of the four studies that met the WWC criteria were small (i.e. less than 100). From this comment, May et al. concluded that the WWC findings on RR are “on the whole very positive” (p. 7) but that a large-scale and “highly rigorous examination of the impacts of Reading Recovery” (p. 7) was warranted. Despite attempts to adhere to the WWC standards, there are major shortcomings with the May et al. study.

Selection of Students

May et al. explained the two-tiered selection process for placement into the RR program, noting that the Reading Recovery Council of North America (RRCNA) *Standards and Guidelines* (2009) require schools to select “the lowest achieving first-grade students” and to serve “the lowest scorers first” (May et al., 2015, p. 24). In a significant departure from the RRCNA *Standards and Guidelines*, this requirement was not consistently followed in the scale-up study. Instead, schools initially identified a pool of “low-achieving students” (p. 24), and then selected the students who received the RR intervention. Despite RRCNA (2009) guidelines requiring schools to use only tests from the Clay’s observation survey (2002) to select the lowest achieving students, May et al. reported

that schools used various assessment measures, with many schools adopting various team approaches to identify suitable children for placement in RR. May et al. noted, “in a number of schools, students were nominated to the pool based on classroom teachers’ general observations or ‘gut’ impressions of their needs” (p. 24). The authors remarked that one important area of variation in the selection of students for RR was the “extent to which schools excluded particular groups of students from receiving Reading Recovery” (p. 25). Schools that excluded students reported preferring to “reserve slots for students they regarded as more likely to benefit from the intervention” (p. 25). Such practices call into question the confidence that schools have in the effectiveness of RR in meeting the needs of students who experience the most significant difficulties in learning to read, that is, the *lowest achieving* six-year-old literacy learners.

Clay was opposed to such practices, but she conceded, “Schools have wanted to select children for the intervention, who in their judgment, would be ‘able to profit from the intervention’ and have been willing to exclude some of the lowest-achievers from selection” (Clay, 2005a, p. 22). Nonetheless, the practice appears to be widespread.

In New Zealand, those six-year-olds with the greatest need for literacy supports are often not placed in the program (Chapman et al., 2007; Church, 2005; Clay, 2005; McDowall, Boyd & Hodgden, 2005) because of doubts about the effectiveness of RR for these students. In their New Zealand Council for Educational Research study of RR, McDowall et al. (2005) made reference to information from interviews of teachers that supported the widely known practice that many children with the most challenging literacy supports are either not placed in RR (in schools that offer the program) or withdrawn if progress is too slow.

Belgrave (2009) found evidence in support of this view. She reported that RR teachers interviewed in her study said they were “manipulating which children they took on, so as to have a degree of success with their students” (p.51). Teachers also told Belgrave that it was a “waste of time and money” to take the “very bottom students” (p. 51).

In an Australian study, Serry, Rose, and Liamputtong (2014) also reported that RR teachers questioned the effectiveness of the program for children with the greatest needs for early literacy

supports. Most participants in the Serry et al. study accepted that RR “intervention alone could not be relied on to return a child to the *average* [original emphasis] reading level of his/her classroom” (p. 68). These authors also found that RR teachers were cautious in their view that the program could “adequately meet the needs of children” with specific reading disability or dyslexia (Serry et al., 2014, p. 68). In these cases, RR teachers reported feeling very “powerless” (Serry et al., 2014, p. 69) in catering for such children.

In the United Kingdom, an anecdotal report is consistent with the findings about exclusions in Australia, New Zealand, and the United States. McLachlain (2014) wrote, “the RR teacher in the UK does not select children for the scheme if they have any sort of learning difficulties and those who make no progress are discontinued from the scheme early and eliminated from the stats as if they never took part” (p. 39). As in New Zealand (Chapman et al., 2015), McLachlain (2014) commented, “Most children who receive RR go on to need more intervention” (p. 39).

The selection practices carried out in the May et al. (2015) study are clearly not consistent with Clay’s (2005a) views about the need to cater for the “hardest to teach” students nor are they consistent with the RRCNA guidelines. However, data from the National Monitoring Reports of RR in New Zealand indicate that the program is of little or no benefit for low-progress readers (Chapman et al., 2015). Australian researchers have also reported such evidence. As Reynolds and Wheldall (2007) and Reynolds, Wheldall, and Madelaine (2009) have noted, RR is not an ideal choice for children with specific reading disabilities because the program does not explicitly target phonological processing skills in a structured manner, despite the abundance of evidence that such targeting is necessary for many children who struggle the most with reading.

The key issue regarding the selection of students in the May et al. (2015) study is less about the screening of students prior to selection into the RR program and more about the fact that the lowest achieving students were often not selected for this program because schools lacked confidence that the program would be beneficial for such students. There are good theoretical reasons why the program is inadequate for many students with specific reading disabilities (see Chapman et al., 2015). However, it

behooves RR to be very clear about the literacy learning profiles of students who are selected for placement in the program. Students who are successfully discontinued from RR are not usually those who are on the low end of the developmental continuum for learning to read (Chapman et al., 2015).

May et al.'s (2015) opening sentence in their i3 scale-up article stated, "Instructional initiatives that are able to identify students at *greatest risk* [emphasis added] and help them achieve grade-level proficiency are essential" (p. 547). We agree. But RR is not the program that includes such instructional initiatives (Chapman & Tunmer, 2011; Chapman et al., 2007, 2015). Given the rigidity of RR, the only way school systems can make it work is to exclude the lowest performing children, the ones for whom more systematic instruction in phonological skills is needed to enable them to achieve progress. However, the multiple cues underpinnings of the program (Greaney, 2011; Tunmer & Nicholson, 2011) mean that such instruction will not be available unless the program is modified.

Control Group Comparison Process

May et al. detailed the procedures for obtaining a randomized group of RR intervention and control students. Eight low-performing students eligible for RR were identified at each participating school. These eight students were matched into pairs, using pre-test text reading levels and English language learner status. One student in each pair was randomly assigned to the RR group, and the other student in the pair was assigned to the control group. Control students were given the option for non-RR intervention, if available, and also the opportunity to enter the RR program after their intervention counterpart had completed the program and if the student still required RR at that time.

Further information about the control group is presented in Footnote 4 (May et al., 2015, pp. 576–577). The authors stated that data were not collected on the participation of control group students in other interventions, though "preliminary analyses" (p. 577) the following year suggested that around 25% of the control students received no supplementary intervention, while the others received other interventions, "typically in small group settings" (p. 577). This statement is at odds with the WWC (U.S.

Department of Education, 2014) statement that the comparison students in this study “received no supplemental instruction during the intervention period” (p. 3).

We find it surprising that May et al. did not provide more detailed information about what the remaining 75% of the control students actually received by way of supplementary instruction. Yet, the WWC report stated that the May et al. research “meets WWC group design standards *without reservation*” (U.S. Department of Education, 2014, p. 2, emphasis added). Although the randomization process may have been appropriate, the RR intervention was compared with a mixture of unspecified small-group instruction for many students and no supplementary instruction for around a quarter of the students. Clearly, one-to-one instruction in general is better than no instruction or small group instruction. A more important question is whether RR is superior to other forms of one-to-one intervention, both in terms of cost and outcomes.

Results of Participation in the Program

Only 54.2% of the intervention students successfully completed the RR program. A relatively large number, 22.4%, were “referred on” for further remedial assistance because of inadequate progress in RR. That is, more than one fifth of the struggling readers did not benefit much from RR, which is consistent with what others have reported (e.g., Elbaum et al., 2000). This percentage may be an underestimate, as another 20% of the children in the treatment sample did not complete RR for a variety of reasons that lack specific details. Accordingly, the results reported by May et al. are for those students who had complete data sets. These students accounted for 69% of the students from the matched pairs in the schools that carried out the random assignments. May et al. noted that the missing student data resulted mainly from student mobility “or other factors that led to the inability to administer the ITBS tests to both treatment and control students in the pair” (p. 559). These “other factors” were not described; presentation of pre-test scores on those students who did not have complete data would have provided potentially important information about their pre-intervention reading achievement status compared to those for whom completed data were collected. Similarly, data on those

students who were referred on for further remedial assistance would have provided important information regarding the extent to which they performed more poorly than those students who were successful and whether or not the referred-on students performed any better than the control students.

The RR program outcome findings in the May et al. scale-up study are considerably lower than those typically reported each year in the RR monitoring data for New Zealand. The latest available monitoring report (for 2013) showed that 58.7% were successfully discontinued, 25% were “carried over” to the following year, 9.7% were referred on for specialist or long-term support, 3.9% left the school before completing the program, and 2.2% were “unable to continue” in the program (Ministry of Education, 2015, p. 15). The May et al. (2015) study suggests that RR is even less effective in the United States for those struggling readers who need the most help than in New Zealand.

A recent meta-analysis by Slavin et al. (2011) showed that the effectiveness of different reading interventions and the magnitude of effect sizes were positively related to the amount of explicit instruction in phonics that was included in the intervention. Not surprisingly, RR, which includes very limited explicit, systematic instruction in alphabetic coding skills, came out last. Related to this point is research indicating that RR is more effective for children from code-orientated classrooms than from whole language classrooms (Center et al., 2001). Slavin et al. also reported that the effect size for RR (0.24) was similar to that for volunteer and largely untrained tutors (0.25). As the May et al. (2015) study has showed, fidelity of implementation does not ensure that RR is an effective program for those students who struggle with learning to read. Neither do highly trained RR teachers necessarily provide beneficial instruction for those children who struggle the most. There are good theoretical reasons for this view.

We have argued elsewhere (Chapman et al., 2015) that the effectiveness of RR interacts with where children are located on the developmental progression from pre-reader to skilled reader. Because of limited knowledge of print at the outset of learning to read, and/or developmental delay in acquiring the phonological awareness skills that are essential for learning to read successfully (e.g., Pressley, 2006; Snow & Juel, 2005; Tunmer, Greaney & Prochnow, 2015), a large proportion of young struggling readers

operate at low developmental phases of word learning, which Ehri (2005) described as pre-alphabetic and partial-alphabetic phases. Delayed readers who are still in these phases, typically those students who struggle the most with learning to read, will not be able to grasp the alphabetic principle and discover spelling-to-sound relationships on their own or in a program that emphasizes text rather than word-level instructional approaches. These students will require more intensive and systematic instruction in phonemic awareness and phonemically based decoding skills than what is provided in typical RR lessons.

Two studies support this claim. Iversen and Tunmer (1993) found that the effectiveness of RR could be improved considerably by incorporating into the program more intensive and explicit instruction in phonological awareness and the use of letter-sound relationships, in combination with strategy training on how and when to use this knowledge to identify unknown words in text. Chapman et al. (2001) reported that students who were successfully discontinued from the program but who failed to achieve significant progress or maintain gains made in the program (most of the RR children in the study) had limited or severely limited phonemic awareness and phonemically based decoding skills at the beginning of RR, as well as during the year preceding entry into the program (Year 1) and during the year following participation in the program (Year 3). The relatively small number of children who received some modest benefit from RR was more advanced in phonological skills at the beginning of the program than children who derived little or no benefit from RR. Progress in learning to read following successful completion of RR was strongly related to phonological skills at discontinuation from the program.

The May et al. (2015) study could have provided more information about the reading skills of those students who did not successfully complete the program. These authors seemed unaware that the RR national monitoring reports in New Zealand have for many years provided pre-RR and post-RR scores for successful and "referred-on" students. The referred-on students consistently have lower scores prior to entry into the RR program than the successful students for all key observation survey measures: word knowledge, instructional book level, and writing vocabulary. Given the fairly consistent numbers of students who are not successful in

the RR program (around 15% in New Zealand; 22% in the May et al. study; up to 30% in other studies, e.g., Elbaum et al., 2000), it seems clear that the multiple cues approach to literacy instruction (Greaney, 2011; Tunmer & Nicholson, 2011), with associated inadequate attention to alphabetic coding skills, consistently fails to meet the needs of a significant number of students who operate at lower developmental phases of word learning (Tunmer & Chapman, 2015). Our research indicates that RR is not effective for those children at the lower end of the developmental continuum because they need much more explicit instruction in phonemic awareness and alphabetic coding skills than what is typically provided in the standard RR lesson (Chapman et al., 2001, 2015).

Sustainability of Gains Made in the RR Program

May et al. (2015) stated, “Reading Recovery is an intensive intervention targeting the lowest achieving 15% to 20% first-grade readers” (p. 549). They further stated that RR can disrupt the “trajectory of low literacy achievement, produce accelerated gains, and enable students to catch up to their peers and *sustain achievement at grade level into the future*” [emphasis added] (p. 549). Both statements are without foundation.

As May et al. (2015) reported, only 52.4% of the children in the RR intervention successfully completed the program; a large minority of students did not benefit from the program. Perhaps more importantly, the assertion that accelerated gains resulting from RR are sustained “into the future” (p. 549) is simply untrue. There are no data presented by May et al. to support the claim. Instead, May et al. appear to repeat the assertion made by Clay (1987) that the RR program would “clear out of the remedial education system all children who do not learn to read” (p. 169) and that the “hardest-to-teach” children could be brought to a level in reading that is equivalent to their same age peers (Clay, 1993, p. i).

After three decades of RR, there is virtually no empirical evidence to indicate that successful completions in RR result in sustained literacy achievement gains. Nicholas and Parkhill (2013) examined New Zealand data for 95 *successfully discontinued* RR students when they were in Years 4 to 6 (ages 8 to 10 years).

These authors reported scores on a New Zealand standardized test of reading comprehension (Progressive Achievement Tests; Darr, McDowall, Ferral, Twist, & Watson, 2008); 49% of the children were in the stanine range of 4 to 6; 6% were in the range of 7 to 10, and almost 45% were well below average in the stanine range of 1 to 3. Nicholas and Parkhill noted that the initial gains made by many students as a result of RR “are not sustained for almost half of the students” (p. 9).

In a similar study, Jesson and Limbrick (2014) also reported data for New Zealand students who had been successfully discontinued from RR during their second year in school, and who were in Years 4 to 6 at the time of the follow-up study. Data on standardized measures of reading comprehension were available for 479 students. They found that around 65% of the students were performing below the average level of stanine 5; nearly 40% were performing in the range of stanines 1 to 3. Jesson and Limbrick concluded that the lack of sustainability of RR gains is seen in the “large numbers of students who are achieving neither at national expectations on standardised reading and writing assessment tools nor at the average levels achieved in reading by age cohorts in their schools” (p. 112).

Considered together, the studies by Nicholas and Parkhill (2013) and Jesson and Limbrick (2014) indicate that gains made in RR by students who were successfully discontinued from the program do not last for more 50% of them in terms of maintaining at least average literacy learning performance outcomes. To this percentage should be added the 12–13% who each year do not benefit from RR and are referred on for further literacy supports.

Further evidence that challenges the claims made by May et al. that RR gains are sustained into the future have been reported by Chapman et al. (2015) in relation to New Zealand results on the most recent Progress in International Reading Literacy Study (PIRLS; Mullis, Martin, Foy, & Drucker, 2012). A New Zealand nation-specific question was included in the home-based “Learning to Read Survey.” Just over 60% of parents or caregivers ($n = 3,400$) completed this survey (Chapman et al., 2015). Parents/caregivers were asked if their child had been placed in a remedial program. Approximately 15% of parents/caregivers reported that their child had received remedial reading. Most

respondents (69%; $n = 600$) reported that RR was the program their child had received.

Our analyses of overall mean reading scores revealed a very large difference between scores for children who had received RR ($M = 493.10$; $SD = 79.58$) compared to those who did not receive any remedial support for reading ($M = 568.05$; $SD = 79.96$). The 75-point difference is equal to almost one standard deviation unit. In other words, students who had received RR three years earlier when they were six years old were performing very poorly compared to their same-aged peers who did not require placement in a remedial program.

To summarize, May et al. (2015) presented no evidence in support of their claim that gains made in RR are sustained over time. On the contrary, three New Zealand studies indicate that RR simply has not achieved its primary goals, at least in New Zealand, where the program was developed and first implemented. Clay's avowal that RR would "clear out of the remedial system all children who do not learn to read" (Clay, 1987, p. 169) and the RR New Zealand's website claim that RR operates as an "effective prevention strategy against later literacy difficulties" and, therefore, "may be characterised as an insurance against low literacy levels" (www.readingrecovery.ac.nz/reading_recovery) are false.

Similarly, in Australia, Reynolds and Wheldall (2007) and Reynolds et al. (2009) have noted that RR is not an ideal choice for children with specific reading difficulties because the program does not explicitly target phonological processing skills in a structured way. The RR program "has not demonstrated that it works for students who are most at risk of failing to read" (Reynolds & Wheldall, 2007, p. 213), and "the success of the program appears to be inversely related to the severity of the reading problem (Reynolds & Wheldall, 2007, p. 219).

Conclusion

The May et al. (2015) RR scale-up study cannot be taken as evidence that the RR program in the United States is effective. Although the WWC report on the research is very positive (Department of Education, 2014), many important features of the study

were not taken into account by the WWC. We have identified four key problems with the May et al. study:

1. The i3 scale-up study did not serve the lowest reading achievers but rather a selected group of low achievers, identified by procedures that were contrary to Clay's views and to the *Standards and Guidelines* of the Reading Recovery National Council of America (2009). Had the lowest achievers in each school been included in the intervention, the results would have been considerably lower.
2. The control group received different experiences during the intervention. Nearly 25% received no supplementary instruction, and the remainder received a range of unspecified group interventions. Such a comparison involving highly trained RR teachers in comparison with doing nothing, or having small group instruction of unknown quality and content, is an unsatisfactory method of assessing the effectiveness of an intervention program. It is surprising to us that the WWC would conclude that the scale-up study group design standards were met "without reservation" (U.S. Department of Education, 2014, p.2), and calls into question the seemingly very narrow requirements of the WWC which overlook crucial factors associated with the nature of the control/comparison processes.
3. The 54.2% rate of successful completion of RR is very modest, and the referred-on rate for those who were unsuccessful in the program of 22.4% is relatively high. The headline effect size of 0.69 over represents the effectiveness of the program considering that many of the poorest readers were not included in the study, and the comparison process involving a mixture of alternative control treatments is associated with confounding factors.
4. The claim by May et al. that RR is a program that leads to sustained literacy learning gains is without foundation. There are no data in the scale-up study that address this issue. Data elsewhere, especially in New Zealand, show that gains achieved by large numbers of RR students are not sustained between 2 and 4 years following successful completion of the program.

RR may be of some benefit to some children. But there is clear evidence in the May et al. (2015) study, and in other

reports (Chapman et al., 2015), that the program is of limited or no benefit to those children who most need early literacy supports. There are serious shortcomings and much-needed improvements in several aspects of RR, including the theoretical underpinnings of the program, the assessment battery that fails to include measures of phonological processing skills, and the specific instructional strategies emphasized in the program (e.g., multiple cues approach to word identification). Fundamental changes in all of those areas would very likely improve the effectiveness of the program, both in terms of outcomes and cost (Church, 2005; Reynolds & Wheldall, 2007; Tunmer & Chapman, 2003).

It is surprising that those responsible for the oversight of the RR program appear to be resistant to making changes to the theoretical framework, materials, and teaching approach, based on a wealth of contemporary scientific research, monitoring data, and follow-up studies, all of which identify significant shortcomings with this program. Instead of spending substantial sums of money comparing the effectiveness of this intensive one-to-one intervention program with either doing nothing, or with unspecified small-group instruction, we urge researchers and practitioners to consider the following question: Holding the basic parameters of the RR program constant (namely, that it involves one-to-one instruction for 30 to 40 minutes per day for 12 to 20 weeks by a specially trained teacher and that it supplements regular classroom reading instruction), are the specific procedures and instructional strategies of RR more effective than any other one-to-one tutorial (or small group tutoring) programs for struggling readers? This question seems especially worth pursuing given the serious concerns that have been raised about the theoretical underpinnings of the program.

References

- Belgrave, J. W. (2009). *The characteristics of children who are referred on from Reading Recovery*. (Unpublished master's thesis). Hamilton, New Zealand: University of Waikato.
- Center, Y., Freeman, L., & Robertson, G. (2001). The relative effect of a code-oriented and a meaning-oriented early literacy program on regular and low progress Australian students in Year 1 classrooms which implement Reading Recovery. *International Journal of Disability, Development, and Education*, 48, 207–232.

- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly, 30*, 240–263.
- Chapman, J. W., Greaney, K. T., & Tunmer, W. E. (2007). How well is Reading Recovery really working in New Zealand? *New Zealand Journal of Educational Studies, 42*, 17–29.
- Chapman, J. W., Greaney, K. T., & Tunmer, W. E. (2015). Is Reading Recovery an Effective Early Literacy Intervention Programme for Children Who Most Need Literacy Supports? In W. E. Tunmer & J. W. Chapman (Eds.), *Excellence and equity in literacy education: The case of New Zealand* (pp. 41–70). London: Palgrave Macmillan.
- Chapman, J. W., & Tunmer, W. E. (2011). Reading Recovery: Does it work? *Perspectives on Language and Literacy, Fall*, 21–24.
- Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2001). Does success in the Reading Recovery program depend on developing proficiency in phonological-processing skills? *Scientific Studies of Reading, 5*, 141–176.
- Church, J. (2005, December). *Accelerating reading development in low achieving children: A review of research*. Paper presented at the annual meeting of the New Zealand Association for Research in Education, Dunedin, New Zealand.
- Clay, M. M. (1985). *The early detection of reading difficulties, 3rd ed.* Auckland, New Zealand: Heinemann.
- Clay, M. M. (1987). Learning to be learning disabled. *New Zealand Journal of Educational Studies, 22*, 155–173.
- Clay, M. M. (1991). *Becoming literate: The construction of inner control, 2nd ed.* Auckland, New Zealand: Heinemann.
- Clay, M. M. (1993). *Reading Recovery: A guidebook for teachers in training*. Auckland, New Zealand: Heinemann.
- Clay, M. M. (2002). *An observation survey of early literacy achievement, 2nd ed.* Auckland, New Zealand: Heinemann.
- Clay, M. M. (2005a). *Literacy lessons designed for individuals. Part one: Why? When? and How?*. Auckland, New Zealand: Heinemann.
- Clay, M. M. (2005b). *Literacy lessons designed for individuals. Part Two: Teaching procedures*. Auckland, New Zealand: Heinemann.
- Darr, C., McDowall, S., Ferral, H., Twist, J., & Watson, V. (2008). *Teacher manual: Progressive Achievement Test: Reading*. Wellington, New Zealand: New Zealand Council for Educational Research Press.
- Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 135–154). Oxford, UK: Blackwell.
- Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92*, 605–619.
- Greaney, K. T. (2011). The multiple cues or “searchlights” word reading theory: Implications for Reading Recovery. *Perspectives on Language and Literacy, Fall*, 15–19.

- Iversen, S. A., & Tunmer, W. E. (1993). Phonological processing skill and the Reading Recovery program. *Journal of Educational Psychology, 85*, 112–125.
- Jesson, R., & Limbrick, L. (2014). Can gains from early literacy interventions be sustained? The case of Reading Recovery. *Journal of Research in Reading, 37*, 102–117.
- May, H., Gray, A., Sirinides, P., Goldsworthy, H., Armijo, M., Sam, C., . . . Tognatta, N. (2015). Year one results from the multisite randomized evaluation of the i3 Scale-Up of Reading Recovery. *American Educational Research Journal, 52*, 547–581.
- McDowall, S., Boyd, S., & Hodgen, E. (2005). *Reading Recovery in New Zealand: Uptake, implementation, and outcomes, especially in relation to Māori and Pasifika students*. Wellington, New Zealand: New Zealand Council for Educational Research.
- McLachlain, K. (2014). Further thoughts on Reading Recovery. *Learning Disabilities Australia Bulletin, 46*, 39.
- Ministry of Education. (2015). *Annual monitoring of Reading Recovery: The data for 2013*. Wellington, New Zealand: Author.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 International results in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timss.bc.edu/pirls2011/international-results-pirls.html>
- Nicholas, K., & Parkhill, F. (2013). Is Reading Recovery sustainable two to four years after discontinuation? *Education 3–13: International Journal of Primary, Elementary and Early Years Education 42*(6), 663–672. doi:10.1080/03004279.2012.759605.
- Nicholson, T. (1989). A comment on Reading Recovery. *New Zealand Journal of Educational Studies, 24*, 95–97.
- Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching*. New York, NY: Guilford Press.
- Reading Recovery Council of North America. (2009). *Standards and guidelines of Reading Recovery in the United States* (5th ed.). Columbus, OH: Author. Retrieved from http://www.readingrecovery.org/pdf/implementation/Standards_Guidelines-09.Full.Version.pdf
- Reynolds, M., & Wheldall, K. (2007). Reading Recovery 20 years down the track: Looking forward, looking back. *International Journal of Disability, Development, and Education, 54*, 199–223.
- Reynolds, M., Wheldall, K., & Madelaine, A. (2009). The devil is in the detail regarding the efficacy of Reading Recovery: A rejoinder to Schwartz, Hobsbaum, Briggs, and Scull. *International Journal of Disability, Development and Education, 56*, 17–35.
- Serry, T., Rose, M., & Liamputtong, P. (2014). Reading Recovery teachers discuss Reading Recovery: A qualitative investigation. *Australian Journal of Learning Difficulties, 19*, 61–73.
- Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best evidence synthesis. *Educational Research Review, 6*, 1–6.

- Snow, C. E., & Juel, C. (2005). Teaching children to read: What do we know about how to do it? In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 501–520). Oxford: Blackwell.
- Soler, J. and Openshaw, R. (2006). *Literacy crises and reading policies*. Routledge: London.
- Stockard, J. (2010). An analysis of the fidelity implementation policies of the What Works Clearinghouse. *Current Issues in Education*, 13(4), 1–24.
- Tunmer, W. E., & Chapman, J. W. (2003). The Reading Recovery approach to preventive early intervention: As good as it gets? *Reading Psychology*, 24, 337–360.
- Tunmer, W. E., & Chapman, J. W. (Eds.) (2015). *Excellence and equity in literacy education: The case of New Zealand*. London: Palgrave Macmillan.
- Tunmer, W. E., & Nicholson, T. (2011). The development and teaching of word recognition skill. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. Afflerbach (Eds.), *Handbook of Reading Research* (Vol. 4, pp. 405–431). New York: Routledge.
- Tunmer, W. E., Greaney, K. T., & Prochnow, J. E. (2015). Pedagogical constructivism in New Zealand literacy education: A flawed approach to teaching reading. In W. E. Tunmer & J. W. Chapman (Eds.), *Excellence and equity in literacy education: The case of New Zealand* (pp. 121–144). London: Palgrave Macmillan.
- U.S. Department of Education. (2013). *What Works Clearinghouse. Beginning reading: Reading Recovery*. Washington DC: Author. Retrieved from <http://whatworks.ed.gov>
- U.S. Department of Education. (2014). *WWC review of the report: Evaluation of the i3 scale-up of Reading Recovery year one report, 2011–12*. Washington, DC: Author. Retrieved from <http://whatworks.ed.gov>
- Wood, T. (2014). *Examining the inaccuracies and mystifying policies and standards of the What Works Clearinghouse: Findings from a Freedom of Information Act request*. Eugene, OR: National Institute for Direct Instruction, Technical Report 2014-5. Retrieved from <http://www.nifdi.org/research/recent-research/technical-reports>